2017 SESAR Innovation Days

Selected scientific papers on air traffic management
Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*):
00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

About SESAR

As the technological pillar of the Single European Sky initiative, SESAR aims to modernise and harmonise air traffic management in Europe. The SESAR Joint Undertaking (SESAR JU) was established in 2007 as a public-private partnership to support this endeavour. It does so by pooling the knowledge and resources of the entire ATM community in order to define, research, develop and validate innovative technological and operational solutions. The SESAR JU is also responsible for the execution of the European ATM Master Plan which defines the EU priorities for R&D and implementation. Founded by the European Union and Eurocontrol, the SESAR JU has 19 members, who together with their partners and affiliate associations represent over 100 companies working in Europe and beyond. The SESAR JU also works closely with staff associations, regulators, airport operators, airspace users, the military and the scientific community.

Learn more about SESAR:
www.sesar.eu
www.sesarju.eu
# Table of contents

About SESAR ................................................................. 3  
Forewords ................................................................. 6  
2017 SESAR Innovation Days – Programme Committee ................. 8  
Introduction ................................................................ 9  

**Airports**

Cooperation under Guidelines ........................................ 12  
DMAN-SMAN-AMAN Optimisation at Milano Linate Airport .................. 25  
Stakeholder Cooperation for Improved Predictability and Lower Cost Remote Services...... 39  

**Airspace management**

Assessing the Viability of an Occupancy Count Prediction Model.................. 56  
Probabilistic Traffic Models for Occupancy Counting................................. 69  
Towards Simplified Optimal Sector Splitting ........................................ 84  

**Big data and liability**

Autonomous Systems in Aviation: Between Product Liability and Innovation.......... 98  
New Data Sources to Study Airport Competition........................................ 110  
Understanding Door-to-Door Travel Times from Opportunistically Collected Mobile Phone Records .... 123  

**Conflict detection and resolution**

Identification of Spatiotemporal Interdependencies and Complexity Evolution in a Multiple Aircraft Environment ............................................. 138  
Probabilistic Aircraft Conflict Detection and Resolution Considering Wind Uncertainty..... 150  
Simulated Annealing for Strategic Traffic Deconfliction by Subliminal Speed Control under Wind Uncertainties ........................................... 165  

**Demand and capacity building**

Coordinated Capacity and Demand Management in a Redesigned ATM Value Chain ........ 182  
Optimal Delay Allocation under High Flexibility Conditions during Demand-Capacity Imbalance ............................................................. 195  

**Economics**

Building a Holistic ATM Model for Future KPI Trade-Offs .......................... 210  
An Agent-Based Auction Model for the Analysis of the Introduction of Competition in ATM ........................................................................... 224  
Ownership Form & Air Navigation Service Provider Performance .................. 238
Flow management

A Trajectory Clustering Framework to Analyse Air Traffic Flows................................. 252
Combining Visual Analytics and Machine Learning for Route Choice Prediction .......... 266

Human factor

Analysis of a Workload Model Learned from Past Sector Operations ....................... 282
Development of Neurometrics for Selective Attention Evaluation in ATM ................. 297
Machine Learning of Controller Command Prediction Models from Recorded Radar Data and Controller Speech Utterances ......................................................... 310

Meteo and the environment

Fuel Saving by Gradual Climb Procedure ................................................................. 326
Ground-based Wind Field Construction from Mode-S and ADS-B Data with a Novel Gas Particle Model ................................................................. 339
Improving Snow Nowcasts for Airports ............................................................... 357
Multi-criteria Environmental Impact Assessment and Optimisation of Aircraft Trajectories ................................................................. 370
Probability of Snow Nowcasting for Airports ...................................................... 382

Modelling and simulation

A Novel Framework to Assess the Wake Vortex Hazards Risk Supported by Aircraft in En-Route Operations ................................................................. 394
Assessing ATM Performance with Simulation and Optimisation Tools .................. 403
Simulating the Risk of Bird Strikes ...................................................................... 416

Performance

Initial Implementation of Reference Trajectories for Performance Review .................. 430
User-centric Cost-based Flight Efficiency and Equity Indicators ............................. 444

Traffic prediction

DART: A Machine-Learning Approach to Trajectory Prediction and Demand-Capacity Balancing ................................................................. 458
Effects of Reducing Wind-Induced Trajectory Uncertainty on Sector Demand ........ 472
Short-term 4D Trajectory Prediction Using Machine Learning Methods ................. 487
Forewords

Delivering smarter, safer and cleaner transport systems

Europe has the talent, skill and drive to lead the world in delivering smarter, safer and cleaner transport systems.

The SESAR Innovation Days is a testament to Europe’s capacity and foresight in this regard. It is also a great opportunity to showcase the value of the SESAR model for collaboration, which seeks to break down silos of knowledge within and outside Europe.

Overcoming fragmentation is critical to unlocking innovation and to boosting growth and jobs, as we have outlined in the Aviation Strategy for Europe.

We must invest in the “now” but prepare for the future. The SESAR Innovation Days help nurture the aviation talent of the tomorrow and develop new ideas to ensure greater mobility and connectivity through air travel in Europe.
Exploratory research: A catalyst for excellence in air traffic management in Europe

Encouraging new ideas and fresh thinking is critical for innovation in aviation in order to respond to the growing demand for air travel and the increasing number of air vehicles, such as drones, taking to the skies. That is why at the SESAR Joint Undertaking we support long-term research and have created an innovation pipeline in our research programme that transforms innovative ideas into solutions to increase the performance of air traffic management (ATM). Harnessing innovation to tackle the opportunities and challenges that lie ahead is also very much in line with the vision set out in FlightPath 2050, the 2015 Aviation Strategy and the European ATM Master Plan.

It is also the philosophy behind the SESAR Innovation Days, Europe’s largest ATM research-focused event – which gathers the world of academia and industry to discuss breakthrough research findings and opportunities. This publication provides an overview of the papers that were presented during the 2017 edition of the event, clearly providing new ideas on a wide range of topics related to air traffic management modernisation.

With these research outcomes, it is our intention to push the boundaries of ATM knowledge thereby developing and delivering solutions for smart, seamless and safe air travel for Europe and its citizens.
2017 SESAR Innovation Days – Programme Committee

General Chair: David Bowen – SESAR JU
Programme Chair: Dirk Schaefer – EUROCONTROL
Local Organisation Chair: Radosav Jovanovic – University of Belgrade

Programme Committee

Rosa Arnaldo – UPM
Tatjana Bolic – University of Trieste
Marc Bourgois – EUROCONTROL
Lorenzo Castelli – University of Trieste
Andrew Cook – University of Westminster
Francesca De Crescenzio - University of Bologna
Daniel Delahaye – ENAC
Peter Hecker – TU Braunschweig
Jacco Hoekstra – TU Delft
Gokhan Inalhan – ITU Istanbul
Christopher Johnson – University of Glasgow
Radosav Jovanovic – University of Belgrade
Dirk Kuegler – DLR
Guglielmo Lulli – Lancaster University
Alexander Martin – FAU Erlangen
Max Mulder – TU Delft
David Perez – Innaxis
Simone Pozzi – Deep Blue
Jose Javier Ramasco – IFISC
Damián Rivas – University of Seville
Sander Roosendaal – Honeywell Aerospace
Francisco Javier Saez – Cranfield University
Miguel Vilaplana – Boeing R&T Europe
Introduction

From 28 – 30 November 2017, the SESAR JU held its annual SESAR Innovation Days, which was kindly hosted by the University of Belgrade (Faculty of Transport and Traffic Engineering) in Serbia. With more than 200 participants from both Europe and further afield, the event featured a number of keynote presentations, panel and poster sessions highlighting some of today’s most exciting research taking place in the field of air traffic management (ATM).

Now its 7th year, the event is Europe’s largest ATM research-focused event, bringing together a very strong pool of scientific expertise to share research experiences, foster innovation and ultimately to accelerate the pace of change in ATM. The event is also the main vehicle for the SESAR JU to share progress and disseminate results of its exploratory research programme, funded by the European Commission’s Horizon 2020 initiative. Furthermore, the event also gives participants an opportunity to present research results from outside the programme, thereby ensuring a link between SESAR and the broader ATM and aviation research agenda.

Against this background, researchers from universities, research institutions, airlines, air traffic service providers and industry were invited to submit papers presenting the results of their long-term or innovative research. Papers were evaluated based on the innovative nature of the ideas, as well as the approach and methods applied. A total of 35 papers were selected by a Programme Committee (See page 8). All papers are listed in Elsevier’s Scopus database.

In addition to the papers published, researchers were given the opportunity to present their work via the poster sessions. A total of 21 posters were exhibited in parallel to the event, providing the opportunity for participants to learn about further interesting projects and to meet like-minded researchers.

The event was also the backdrop of the SESAR Young Scientist Award, which was awarded to Ramon Dalmau Codina for his thesis on the development and testing real-time algorithms for the optimal planning during continuous descent operations. The award was announced via video by European Commissioner for Transport, Violeta Bulc.

More information, visit: www.sesarju.eu/sesarinnovationdays
COOPERATION UNDER
GUIDELINES

How Guidelines for Cooperation Affect Interaction Behavior of Airport Experts

Anne Papenfuß
Institute of Flight Guidance German Aerospace Center (DLR)
Braunschweig, Germany
anne.papenfuss@dlr.de

Anna Biasotto
Institute of Psychology
Universität Heidelberg
Heidelberg, Germany

Abstract—Collaborative decision making is a key component for novel air traffic management concepts. They should enable proactive decision making that is flexible enough to take into account needs and priorities of several stakeholders. Experience from training of cockpit teams shows the benefit of models structuring the decision making process in terms of decision quality. Guidelines for cooperation were developed to enhance decision making in two exemplary airport management tasks. Qualitative interaction analysis was conducted to investigate the impact of these guidelines on interaction behavior of three teams of airport experts. Results show, beside strong inter-team differences, that guidelines have the potential to focus team decision making on more thorough situation assessment. The paper proposes metrics to analyze and quantify compliance of teams with a given process. Furthermore, ideas for advancement of the guidelines are derived.

I. Introduction

Collaborative Decision Making (CDM) is a key component for European air traffic management (ATM) concepts. In airport management, through collaborative decisions the goals and constraints of several stakeholders should be considered in decisions made at the day of operation. The expected benefit of CDM processes is that decisions found take into account more constraints thus serve better the circumstances of the situation and therefore are of higher quality. Consequently, CDM processes are an enabler for more performant airport operation.

Beside airport management, other operational areas could benefit from the positive effect of “higher-quality decisions” and greater flexibility through collaborative decision making. For example, within the Pilot Common Project of SESAR Joint Undertaking [1], CDM is named as component for enhancing airport throughput, system wide information management and integrating network management into airport management

In today’s ATM system, the working methods of operators are characterized by standard operating procedures. The rationale is that standard operating procedures guarantee safety by providing predictability and minimizing influences of personality.

In contrast to this, working methods like collaborative decision making are described as being more flexible in order to react upon specific situations and to take into account
needs and priorities of other stakeholders. Those features contradict to some extent the idea of rather predictable, inflexible standard procedures. At least, ideas have to be developed on how to integrate the two work design philosophies.

Summarizing these trends, we state a need for research looking into the working procedures required for those flexible and collaborative decision making concepts. In the domain of airport management, substantial experience was gathered. For instance, state of the art working processes at airports were analyzed via job-shadowing. Furthermore, initial ideas for guidelines for cooperation were presented to experts.

The adjustment of plans within an operations center between stakeholders with different goals was focus of the project "Collaboration within Control Centers (COCO)", financed and executed by the German Aerospace Center DLR. A simulation study was conducted to assess the influence of guidelines for decision making on perceived quality of the decision making process. This paper explores the theoretical foundation of these guidelines and the observable impact on the decision making process of airport experts. Furthermore, metrics to analyze and quantify compliance of teams with a given process are proposed.

II. Background

A. Collaborative Decision Making in Airport Performance Management

Airport operations have a large potential for optimization and thus influence performance of overall ATM. Airports are complex systems with multiple interconnections between numerous processes owned by a multitude of stakeholders [2] Each stakeholder at an airport plans its processes and actions according to individual goals and standards and corporate business plans. But most stakeholders miss information about intentions, goals and actions of other (cooperating or competing) parties at the same airport. Relevant information is not available, available but incorrect or available but too late (cf.[3]).

Hence, harmonizing plans between different stakeholders at an airport is rather time consuming and difficult, especially regarding partly conflicting goals of airport stakeholders and their unwillingness to share all information about their plans. Assessing the impact of other parties actions on one’s own plan is therefore difficult and an integrated view of total airport operations is missing, cf.[2].

Airport collaborative decision making (A-CDM) was developed [4], to foster a more proactive behavior and share a minimum set of relevant data, like the Target Off-Block Time, TOBT, between all stakeholders. Performance benefits of ACDM could be demonstrated, as more and more European airports adopt the concept [5].

This concept was further developed to a solution called Total Airport Management (TAM) [6]. TAM enhances the airside-focused A-CDM concept by integrating landside processes and developing ideas for highly collaborative decision making in Airport Operation Control Centers (APOC) (cf.[7]). Within the context of the Single European Sky ATM Research Program SESAR, processes and use cases for airport operations control centers (APOC) were developed and validated (cf.[8-10]).

Up to now most research focused on technical solutions for the socio-technical system APOC (e.g.[2, 11]). Ideas to foster pro-active, collaborative behavior mainly focused on the competitive roles of several airlines and involved the development of negotiation protocols and bonus-malus-systems [12]. Research on Performance Based Airport Management could show that airport experts in general see the potential for collaboration but they see the need for a mandatory framework or rules for collaboration and cooperation [3]. As one step towards feasible working procedures within an APOC, guidelines for cooperation were developed and evaluated in a high-fidelity simulation with airport experts [13]. The guidelines aimed at improving the flow of relevant information between stakeholders in order to enable each participant in the APOC to adapt his/her plans according to the traffic situation.
B. Models of Decision Making in Aviation

Decision Making is daily business in ATM. Good decisions are the basis for safe and performant air traffic. Laboratory and field studies have shown that human decision making behavior is not necessarily rational. Especially under time pressure, humans tend to make ad-hoc decisions guided by expectations and preferences, they tend to stick to sometimes inadequate goals and follow heuristics instead of analyzing a situation in depth.

Orasanu [14] identified for her descriptive model of aviation decision making the two major components “situation assessment” and “choosing a course of action”. Situation assessment includes the definition of the problem, assessing the risks associated with it and the time available for solving the problem.

The selection of actions distinguishes between the application of rules, the choice between several options as well as the creation of novel solutions. The course taken depends on the understanding if the situation, how much time is available as well as the availability of rules, all together called the situational constraints and affordances of the situation. With this model, observed decision making behaviors of pilots, e.g. during flight accidents, could be explained. For instance, if situation assessment was not sufficient pilots chose an action based on an ill-defined problem which might contribute to the fatal outcome of an accident.

To enhance decision making competencies of pilots, prescriptive models for crew training were developed. They should provide the crew with a structure for the decision making process which suits better to the complex socio-technical environment of the cockpit.

One model developed from Lufthansa and the German Aerospace Center, is the FOR-DEC model [15]. The model distinguishes the six phases Facts, Options, Risks and Benefits, Decision, Execution and Check. It was developed to structure the judgement and decision making processes within the cockpit and was used to train crew resource management (CRM). It was developed to take into account the complexity and dynamics of the cockpit environment, to be applicable to a wide range of situations, and to separate the phases of collecting information about the situation and evaluate possible solutions. The model can be run in several cycles until the desired goal is reached, furthermore sub-cycles and feedback loops are possible. The phases Facts, Options and Risks can be subsumed as ”situation assessment”, the Decision, Execution and Check-Phases can be mapped to the ”course of action” stage of the ADM model.

In team exercises, the authors found that the application of the FOR-DEC model helped teams to structure their communication and group interaction processes [15]. FORDEC is widely known in the pilot community but pilots report that in real life situations, especially under time pressure or where options are clear, following the model feels artificially [16].

C. Guidelines for Collaborative Decision Making

Based on the experience made with decision making in the cockpit, as well as the results from job-shadowing, the need for a structure for the collaborative decision making process in an APOC was postulated. Guidelines were developed to provide the team with a workflow that incorporates the findings of the CRM domain with respect to separate situation assessment and decision making phases. Additionally, for each team member expected activity within the phases was highlighted. The rule for this assignment was that stakeholders who inherit the relevant and reliable information for this phase must provide this information at that time to the other stakeholders. For example, the groundhandler has information about the maximum capacity for turn arounds and thus is able to analyze whether an event will cause an over-demand and thus delays.
On a more abstract level, the guidelines were designed to 1) provide the relevant information at the right time and 2) create transparency on dependencies between stakeholders’ individual planning. In contrast to the cockpit environment, time pressure and risk assessment are not a major issue in the APOC environment. But additional “affordances” exists, as the choosing of actions is influenced by conflicts between the stakeholders, especially conflicts of goals and conflicts of power.

Six phases were differentiated. They are depicted in Figure 2. The guidelines were first described in terms of information and decisions required within each phase. Afterwards, the required actions per stakeholder were broken down for these phases to generate instructions per phase. It must be highlighted, that the guidelines tested in this study were developed for two specific tasks and therefore do not claim to have universal validity. The two tasks are “prioritizing departures” and “stand- and gate allocation”.

First, information about the event that might disturb the airport operations is received (phase 1). Next, dependent on the task, either airport (AP) or groundhandler (GH) analyses the impact on his/her operations (phase 2) as their resources are the bottleneck. Phase one and two can be mapped directly to the situation assessment step of the ADM.

**FIGURE 1 — Phases of the guidelines for cooperation and activity of stakeholders**

Following, principal constraints for the final solution are elaborated by going through the phases 3) generation and distribution of a first, individual solution by either groundhandler or airport and phase 4) the evaluation of the impact of this solution on each stakeholders plan. Within this phase the Airlines (AL) are required to check the influence on their operations. Phase four might trigger a new task, so there is a feedback to phase 2. For instance, the airport might conclude that s/he should start a stand and gate allocation task because the impact of departure prioritization is too big. Phases five and six can be related to the course of action, as the solution is refined to get rid of the problems detected in phase 4. This might require that phase 5 is run several times. Finally, phase six finishes the process by decision of all stakeholders to agree on a final solution.

### III. Research questions

Experience with prescriptive models for aviation decision making in the cockpit showed the potential of these models to structure the teams decision making behavior. Thus, it was of interest, to what extent the proposed guidelines for an APOC team influenced the collaborative decision making process. Results from a simulation based evaluation of the guidelines show that operators rated efficiency of the decision making process when following the guidelines significantly better compared to an unstructured process [13]. The guidelines provided effective procedures to guide team functioning, provided the teams with clear agreements about how decisions were made so that the teams worked constructively on issues until they were resolved [13]. Beside these positive results, additional analysis was needed to understand whether teams actually behaved in conformance with these guidelines or came up with ad-hoc adaptations of the prescriptive decision process.
An exploratory and descriptive analysis of two research questions leads the following analysis:

We are interested to see, to what extent the proposed phases of the guidelines for cooperation manifested in the observable interaction behavior of the APOC teams. First, do teams follow the proposed transition of phases and the proposed activity patterns of stakeholders?

Second, are there distinctive features that can be observed when teams apply the guidelines for cooperation compared to an unstructured decision making process?

IV. Method

A. Data Collection

A high fidelity simulation in the Airport and Control Center Simulator (ACCES) facility of DLRs Institute of Flight Guidance was conducted. Experts from german speaking airports participated in this study. The set-up used in this study is described in detail in [13]. A generic airport simulation [17, 18], build upon Eurocontrol’s A-CDM implementation standard [4], was used to create three different scenarios where an event disturbs the scheduled Airport Operations Plan (AOP). In order to achieve best possible punctuality, two airlines, one groundhandler and the airport representative collaboratively decide on a new AOP. Within 45 minutes the TOBT of 34 flights in a two hour time window need to be adjusted so that punctuality is improved and the individual goals of all stakeholders are reached.

Four teams of four airport experts participated in this study, each running through three scenarios. Within the first two runs, participants were asked to make their decisions based on their experience and individual working methods. For the third run, participants were introduced to the guidelines, trained and asked to follow these guidelines in their decision making process. The material used in this study consisted of the second and third runs of three teams. Thus, the sample consists of one run with free structure (free) and one run with guidelines (guided) of each group.

All interactions between participants were present in the form of audio files. During the experiment, each participant was equipped with a microphone that recorded his interactions. Using the audio-editor Audacity, the four soundtracks were time synchronized and put into a shared project file.

B. Data preparation

The steps undertaken to prepare the raw audio data for data analysis are summarized in Figure 2.
Transcription. All utterances were transcribed completely and literally, including incomplete sentences and repetitions. Nevertheless, filler words and hecklings were omitted because the focus of the analysis was on the content. Dialect (pronunciation as well as choice of words) was adapted to standard German. Breaks and moments of silence were marked by a dash. Straying from the topic, typing, making jokes as well as questions raised to the experiment supervisor or the observers (e.g. regarding the interface) were not transcribed content wise. Instead, only the action was noted. Other peculiarities such as laughing, a joking or ironic tone, mumbling or loud thinking were mentioned in brackets and italic font after the content [19].

Event sampling. The course of conversation was divided into events whenever the speaker, the addressee or the topic changed, following approaches from discussion coding [20]. Therefore, consent or rejection of an utterance was classified as a separate event. Furthermore, the starting time in minutes and seconds was noted for each event. These data were arranged in the form of a table with one row per interaction and the columns time, content, speaker and addressee.

Prototype of Interaction process. As the guidelines were present in the form of individualized ‘checklists’ for each individual stakeholder, they had to be combined into one general process flow. The chronological order of the expected interactions within each phase was determined. This task was conducted for each of the two tasks “prioritizing departures” and “stand and gate allocation” and resulted in two prototypical flow charts of the interaction process.

Definition of categories. Initially, all six phases of the guidelines were established as categories. In order to evaluate this concept and determine further possible categories, one run was chosen and worked through. Thereby, the following final coding categories were established: phase 2 (conflict detection), phase 3 (generation of initial solution), phase 4 (identification of subsequent conflicts), phase 5 (optimization of solution), priority Flights (P), system (S) and other (R).

The categories ‘phase 2’, ‘phase 3’, ‘phase 4’ and ‘phase 5’ correspond to the phases of the guidelines. The original categories ‘phase 1’ (information about event) and ‘phase 6’ (implementation of final solution) were excluded, as no verbal interactions were found pertaining to these two categories. This finding was supported by the fact that the prototype of the interaction process does not intend any interaction in phases 1 and 6. In phase 1, the stakeholder with the bottleneck receives external information about the event via the system. Thus, no interaction with other stakeholders is required. Phase 6
starts, when the final solution has been agreed upon. In this phase, the final solution is documented in the system. Thus, no interaction between stakeholders is required.

The category ‘priority flights’ was introduced, because this was a topic of relevance in each of the six runs. As this topic was not in line with any of the phases, it was encoded as a separate category. Events, in which priority flights were identified, were assigned to this category. However, all events that addressed measures for priority flights were allocated to the respective phase of the guidelines.

**Coding.** In the next step, all events were assigned to a category. Each event was compared to the prototype of the interaction process and allocated to the phase, which resembled this event the most. In order to ensure internal consistency of the process of category allocation, certain rules were defined: First, an event could only be allocated to phase 3, if the stakeholder with the bottleneck was the one compiling an initial solution. This would be groundhandler or airport, depending on the specific subtask. No suggestions or requests by other stakeholders are allowed in this phase. Second, responses to the initial solution [phase 3] were characterized as phase 4, when the focus was on a general problem analysis. However, suggestions and requests pertaining to specific aircraft were assigned to phase 5 [optimization]. Third, when a conflict was induced by a decision or system input, this event was allocated to the category phase 4 [analysis of induced conflicts]. Fourth, after the interaction was in full swing, it could only go back to phase 2 [conflict detection] and phase 3 [initial solution], if the stakeholders switched to the other task. This is due to the fact that phases 2 and 3 are characterized as initial reactions to the problem at hand. Last, the content was prioritized over the speaker. For instance, if an airline took on a moderating role and suggested that the groundhandler should now work on an initial solution [phase 3], this event was allocated to phase 3, even though the Airline was not supposed to be active in this phase. The psychometric properties of the process of coding were examined by means of stability (intra-rater reliability) and reproducibility (inter-rater reliability). Krippendorff’s alpha ($\alpha$) amounted to .91 for the intra-rater reliability, which measured stability.

**C. Data Analysis**

The interaction process of the six runs was analyzed qualitatively. Furthermore, metrics were developed to derive quantitative results representing the conformance of the decision making process with the proposed guidelines for cooperation and to quantify features of the decision making process.

**Phase progression.** Each event was allocated to one of the guideline’s phases. Thus, the chronological sequence of a runs’ events depicts the phase progression of this run. For each run, the phase progression was displayed graphically. In addition, several sections of each run were further analyzed. Here, the main focus was on the initial part of the interaction, sections where the task switched and long phases of optimization. For these sections, it was examined whether the phases were completed in the linear order intended by the guidelines or whether the interaction oscillated between different phases.

**Compliance of transitions with guidelines.** The guidelines specify which phase transitions are allowed (guideline-consistent) and which are prohibited (non-consistent with guidelines). For each run, the absolute and relative amount of prohibited phase transitions was determined. For each group, a Pearson’s Chi-square test [21] was computed with IBM SPSS Statistics 24 to test if the type of run (free vs. guided) and transition compliance were significantly associated. All values exceeded the necessary case number of five observations.

**Focus of decision making process.** For each run, the number of events pertaining to phases 2, 3 and 4 were summed. Then, for each run, the portion of events pertaining to these three phases was compared to the portion of events pertaining to phase 5. For each group, a Pearson’s Chi-square test [21] was calculated. All observed values exceeded the critical value five.
V. Results

A. Qualitative Analysis of Phase Progression

Due to the limited space in this paper, a detailed description of the decision making process is given for team two only. It has to be mentioned that the detailed phase progression differed between all teams. Nevertheless, commonalities between all groups could be found that are reported subsequently.

Phase progression and communication activity of the free structured decision making process is depicted in Figure 3; Figure 4 shows the process with guidelines. The x-axis codes the time of events and the y-axis shows the phase coded for the events. Furthermore, grey dots visualize the speaker of the event. Topmost are events from the Airport (AP), followed by groundhandler (GH) and the two airlines (AL).

**FIGURE 3 — Phase progression and communication activity of run “free structure”**

![Figure 3](image)

**FIGURE 4 — Phase progression and communication activity of run “guidelines”**

![Figure 4](image)

First, it can be observed that in both conditions phase progression contains several loops from phase 3 to 5. In the freely structured decision making process (Figure 3), three sections are analyzed in detail. In section 1, after conflict detection (phase 2), both airlines rescheduled their flights. These actions were interpreted as optimization measures for specific flights and classified as phase 5. After the airlines concluded their actions, the groundhandler began to solve his conflicts (section 2). Then, however, it was decided to have the airport solve his induced gate conflicts first. Hence, the groundhandler was skipped and the task switched to planning gates and stands. Whilst developing the initial solution, the airport was interrupted several times by the airlines making offers to swap flights or by adjustments of the TOBT of a flight. These two examples of airline interference explain why the graph of the interaction detoured to phase 5 during section 2.

Following, the groundhandler got to resolve his conflicts (section 3). Thus, the task changed to departure prioritization and the interaction went back to phase 3. However, while the groundhandler was compiling an initial solution, the airlines made several inquiries, for instance suggesting a specific flight. In addition, the two airlines compared their number of conflicts (phase 4) and discussed specific
actions for optimization (phase 5). From minute 22 that decision making process was characterized by a long-lasting period of phase 5 which was interrupted three times by short periods of phase 4. This pattern of long optimization phases was found in all runs analyzed in this study.

With guidelines applied by the team, the decision making process of the stakeholders was structured as depicted in Figure 4. Three sections are highlighted. In section 1, the airport detected conflicts (phase 2) and compiled an initial solution (phase 3). Here, phases 2 and 3 alternated, because the airport took an iterative approach: He detected a conflict, solved it and communicated the solution and then checked for further conflicts. Whilst the airport was working on the initial solution, the airlines already disclosed their conflicts. For instance, one airline communicated an induced conflict for the priority flight, an action typical for phase 4. While working on the initial solution, the airport asked the airlines for their approval to change a gate. This was interpreted as an optimization procedure, as it targeted a specific flight (phase 5). These two examples explain why the graph of the phase progression (see Figure 4) oscillated between initial solution (phase 3), analysis of induced conflicts (phase 4) and optimization (phase 5). Nevertheless, it should be mentioned that the airport insisted several times on going back to concluding the initial solution (phase 3).

In the second phase, the groundhandler worked on an initial solution for the task departure prioritization. Thus, the phase progression went back to phase 3. Here, the moderating role of one airline became apparent, who directed this approach. Whilst the groundhandler was compiling the initial solution, the airlines discussed measures for optimization, but decided not to interfere. This was identified as a discussion about strategy, as it broached the issue of how to proceed.

During the optimization, one of the groundhandler’s changes [min 25:08] entailed severe conflicts for the other stakeholders (section 3). Instead of optimizing this situation, the airport was asked to reinstate his initial solution [min 27:19], leading to a repetition of phase 3. In this situation, the moderating role of one airline could be observed, who made two propositions regarding strategy.

Summarizing, in the run with free structure, the first actions were taken by the two airlines that changed their TOBTs. In contrast, in the run with guidelines, the first actions were taken by the airport, who had the best overview of the capacity bottleneck. In the run with free structure, there was initial confusion on how to proceed. Stakeholders were uncertain, whether airport or groundhandler should start with an initial solution. In contrast, in the run with guidelines, all actions were implemented consecutively and in the order intended by the guideline.

B. Quantitative Metrics

1. Compliance with proposed phases of the guidelines

No global difference was found between the two types of runs (free vs. guidelines) in terms of the number of guideline compliant phase transitions. Descriptive data of number of events and percentage of compliant transitions can be found in Table 1 together with further descriptive data of the teams’ interaction process.

<table>
<thead>
<tr>
<th></th>
<th>team 1</th>
<th>team 2</th>
<th>team 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>free</td>
<td>guide</td>
<td>free</td>
</tr>
<tr>
<td>n events</td>
<td>229</td>
<td>265</td>
<td>226</td>
</tr>
<tr>
<td>n evt. phase 2-4</td>
<td>65</td>
<td>74</td>
<td>70</td>
</tr>
<tr>
<td>n evt. phase 5s</td>
<td>164</td>
<td>191</td>
<td>156</td>
</tr>
<tr>
<td>% of compliant trans.</td>
<td>90.8</td>
<td>96.1</td>
<td>93.3</td>
</tr>
<tr>
<td>evt. in phase 5 (%)</td>
<td>71.6</td>
<td>72.1</td>
<td>73.0</td>
</tr>
</tbody>
</table>

TABLE 1 — Overview on descriptive data of decision making process
In all runs, more than 88% of all transitions were compliant with the guidelines, even in the unstructured decision making processes. The average percentage of compliant transitions is virtually identical, with 91.71% in the unstructured runs and 91.95% in the runs with guidelines.

### TABLE II — Chi-Test statistics for compliance of transitions with guidelines

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2$</th>
<th>p</th>
<th>odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>team 1</td>
<td>6.09</td>
<td>.011</td>
<td>2.58</td>
</tr>
<tr>
<td>team 2</td>
<td>3.27</td>
<td>.049</td>
<td>0.55</td>
</tr>
<tr>
<td>team 3</td>
<td>0.004</td>
<td>.543</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Introducing the guidelines had a different effect in all three groups. Chi-Square tests were calculated, as well as odds ratios, representing the likelihood that for each team in runs with guidelines transitions are compliant with the guidelines. In groups one and two, there was a significant association between the type of run and whether or not a phase transition was guideline conforming. However, the nature of the relationship was opposite. In group one, the odds of a phase transition being guideline-consistent was higher when guidelines were present. In group two, the odds of a phase transition being guideline-consistent was lower when guidelines were present. In group three, no significant association between type of run and whether or not a phase transition was guideline-consistent, was found. Consequently, accumulated over all phase transitions of the three groups, no association between the type of run and whether or not a phase transition was guideline-consistent was found $\chi^2(1) = .093, p = .419$.

2. **Focus of decision making process – Long optimization phase versus Situation Assessment**

Each of the six runs concluded with a long-lasting period of phase 5 with correction loops. No general difference regarding the length of this section was found. While in groups two and three this section started later when guidelines were present (group two: min 22:03 vs. 25:08; group three: 12:02 vs. 19:05), the opposite was true for group one (min 15:48 vs. 13:56).

Regarding the distribution of interaction events between the different phases, in all runs the majority of events were categorized as “optimization of solution” (phase 5), ranging from 51% to 77%. The absolute and relative numbers can be found in Table 1. Chi-square tests were calculated to assess if guidelines had an effect on the amount of interaction events in the optimization phase itself. The number of events in phase 5 was compared to summarized number of events in phases 2, 3 and 4. Parameters of the test for each group are summarized in Table 2. Odds ratio refer to the likelihood that an event did not belong to phase 5 in decision making process with guidelines compared to unstructured decision making process. No significant effect was found for team 1 where the length of phases was not affected by the guidelines. A significant effect was found for teams 2 and 3. When applying the guidelines, in both teams more interaction events where categorized as belonging to phases 2, 3 and 4 of the guidelines for cooperation.

### TABLE III — Chi-Test statistics for amount of events in Phase 5 “Optimization of Solution”

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2$</th>
<th>p</th>
<th>odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>team 1</td>
<td>0.13</td>
<td>.494</td>
<td>1.02</td>
</tr>
<tr>
<td>team 2</td>
<td>15.47</td>
<td>&lt;.001</td>
<td>2.14</td>
</tr>
<tr>
<td>team 3</td>
<td>6.27</td>
<td>.008</td>
<td>1.73</td>
</tr>
</tbody>
</table>

### VI. Discussion

This study examined to what extent airport experts implemented a proposed structure (guidelines for cooperation) in their decision making process. An approach for the
collection and analysis of interaction behavior was presented. Qualitative and quantitative analysis was conducted to get insight into the decision making processes.

A. Influence of guidelines on interaction behavior

The qualitative analysis of the phase progressions revealed that teams in an unstructured decision making process discussed on prioritization of tasks and oscillated between overall adjustment of plans (phase 3) to detailed refinement of single flights (phase 5). When applying the guidelines, they did not follow the process consequently but content analysis showed they were aware of the general structure and priority of the tasks to be conducted in phase 3. Nevertheless, with regard to the quantitative results derived, the impact of guidelines was not as visible as expected.

First, some features of the expected interaction process under guidelines could also be observed in freely structured runs. Secondly, teams did not strictly follow the proposed process of the guidelines. Third, interaction behavior analysis revealed strong team differences. Regarding the first argument, the phases of the proposed guidelines could also be identified in the runs with free structure. The steps of decision making process, especially phases 3, 4 and 5, seemed to be intuitive and were also followed when guidelines were not present. This is positive, as the activities proposed by the guidelines are not completely artificial to the expert.

The guidelines mainly influenced the sequence, as they were designed for. Nevertheless, the quantitative metric of conformant transitions was not sensitive to these changes.

When experts were not guided in their decision making process, refinements (phase 5) were conducted very early or even where the starting point for decision making. Mainly, Airlines pushed this phase by providing information or adaptions for their priority flights. In contrast to this, in runs with guidelines the decision making process started in conformance with the guidelines. The observed behavior in the “unguided” interaction process shows that decision making in natural environment tends to focus on heuristics and come up with quick initial solutions. This finding is in line with experiences made in crew resource management.

Additionally the perspective on the problem was dominated by the airlines; which reflects the situation at airports where airlines are the major customer and thus have a lot of power. The guidelines do not propose to change this situation, but the analysis of the situation should also include the perspective of the other stakeholders in order to find a solution in the APOC which increases overall airport performance. In two of three teams, differences in the decision making process in terms of increasing the focus on situation assessment was found. This is result is quite promising, as the teams tested the guidelines for the first time.

B. Metrics to analyze and quantify compliance of teams with a given process

Phase progression of the decision making process was used to analyze the teams’ interaction process. Content analysis revealed causes why teams left the proposed sequence of the phases. The insights gained by this analysis can be used to advance the guidelines in a next step. Furthermore, quantitative metrics were calculated, like percentage of compliant transitions and odds ratio which describes the likelihood of interaction within a certain phase. These metrics can be used to quantify the conformance of team behavior with proposed work procedures. With regard to teamwork and team process, research so far did not produce widely accepted methods and metrics to assess these processes. We demonstrated how team process metrics can be derived and used for statistical analysis by analyzing the interaction process.

There are some limitations with regards to metrics and quantitative analysis. First, the sample size of this study is a limit. Second, the duration of the runs was fixed to
45 minutes. It can be assumed that guidelines lead to shorter time needed to come to a decision every stakeholder agrees with. In this study we could not analyze such an effect. Third, the analysis revealed general strong differences in the interaction behavior between teams, both with regards to their freely structured and guided decision making process. Literature suggests a broad range of factors influencing teamwork process. In the setting of this study, it is likely that even all participants where in their role, their specific experience varied because of the different airports they came from. Furthermore, the teams consisted of different personalities. Extroverted persons tend to talk more, so even if guidelines foresee no activity for these persons their personality might lead to an active interaction behavior.

C. Advancing the guidelines

The detailed analysis conducted in this paper provided insights into decision making process in complex tasks with conflicting goals, hidden information profiles, influenced by different personalities. The limitations of a simulation study are discussed above. Nevertheless, the results provided evidence how to advance the guidelines.

First, results showed that airlines communicated their constraints, represented by priority flights, early and during the situation assessment phases. This finding matches with experiences made with the FOR-DEC model. There, it was observed that pilots thought the process to be too lengthy in case a suitable solution was “obvious”. The guidelines should foresee a phase where constraints that affect situation assessment are shared.

Second, following the approach of the FOR-DEC design, the structure of decision making process should be easy to remember. More emphasis should be on the necessity of a proper situation assessment. The challenge in the collaborative environment is that no single stakeholder has access to all information relevant for situation assessment. In the set-up used for this study, traffic situations were pre-defined and consequences on operations included in the instructions of the simulation run. Nevertheless, explicit assessments like “how many flights are affected” versus “what capacity is available” are valuable to help the team to understand the “affordance” of the situation.

Third, it is of interest to develop guidelines and to understand collaborative decision making in situations with higher risks and higher time pressure as in the APOC environment. Also, decision making processes with several teams contributing to decisions, thus inter-team-collaboration, should be addressed as future air traffic management concepts foresee those processes. For instance, inter-team collaboration is required in remote tower centers when airport control is distributed dynamically, dependent on situational factors. In flight-centric air traffic control, teamwork and task distribution between teams is regarded as one factor to achieve more efficient air traffic control [22]. This paper proposes an approach how to analyze and evaluate collaborative and team processes. Whilst these analysis methods are labor intensive when conducted manually, evolving methodologies like speech recognition, natural language processing and process mining bear the potential for more automated analysis.

REFERENCES


Abstract—This paper presents the design and validation of an optimization algorithm having the purpose of implementing an integrated Departure Manager - Surface Manager - Arrival Manager at Milano Linate airport. The work, based on Single European Sky ATM Research (SESAR) Solutions, has been tested on two actual case-study days, considering the airport stakeholders' objectives and constraints, and taking operative information from the Airport Collaborative Decision Making platform. Obtained results show that the proposed algorithm could increase average timeliness, reduce taxi time and fuel consumption of aircraft operating at Linate, thus contributing to reach a more sustainable and efficient air transport.

I. Introduction

Air transport generates, on a daily basis, thousands of flights that are managed in an safe and efficient way. According to forecasts, however, in the EU there will be 14% more flights in 2023 and 40% more in 2035 with respect to nowadays values ([1], [2]). Since with the currently available infrastructures and services it will be impossible to organise and manage such an increased number of flights, suitable and effective corrective measures must be envisaged and implemented from now. It is worth to be emphasised that objectives for such measures must include an enhancement not only in air traffic capacity and safety, but also in its environmental and economical sustainability.

The presented work fits in such context, and tackles Air Traffic Management (ATM) improvement by defining an optimization algorithm for computing the best solution to
the problem of integrated departures, surface and arrivals management. In order to maintain the strongest links with the real world, the work has been developed using as a reference Milano Linate airport, located in northern Italy, and has been tested with actual data coming from Linate’s Airport Collaborative Decision Making (ACDM).

II. Problem description

ATM has the objective to ensure safe and efficient movement of aircraft along all phases of operations, both on ground and airborne [3]. Considering the airport area, many stakeholders that operate around aircraft can be identified, each taking care of its specific tasks: Airport Operator, Ground Handlers, Air Traffic Controller Operators (ATCOs), Aircraft Operators, etc. Every single decision taken by any of the stakeholders has inevitably consequences on the other stakeholders’ decisions, hence affecting the global efficiency of the whole air transport process. Therefore, from ACDM logic point of view, every single decision should not be taken for optimizing the particular task, but rather for maximising the global efficiency of the airport system.

Fostered by ENAC (Ente Nazionale per l’Aviazione Civile) and ENAV (Ente Nazionale per l’Assistenza al Volo), many efforts have been undertaken in Italy to reach the objectives set at EU level, especially for the main airports, starting from Roma Fiumicino and Milano Malpensa. Some of these efforts have been directed to the study and development of an Extendend - Arrival MANager (E-AMAN), leaving aside its integration with Departure MANager (DMAN), Surface MANager (SMAN) and ACDM [4]. These, however, are essential enabling tools to reach important objectives (such as the reduction of both queues at the runway threshold and of quantity of fuel burned during taxi time), and for the exploitation of the maximum airport traffic potential.

This paper briefly presents the work developed in [5], where it has been decided to approach and solve the aforementioned problem with a vision of departures and arrivals management integrated with the ground handling, in close connection with ACDM. The work has been contextualised at Linate, Milan city airport, which, in 2016 Italy’s ranking, is [6]:

- 3rd for aircraft movements (118,535);
- 4th for passenger movements (9.7 Mi);
- 8th for cargo movements (15 ktons).

Among the other reasons, a full ACDM platform has been active for several years at Linate.

Linate (figure 1) has one main Runway (RWY) which is normally used for departures and arrivals (RWY 36-18), and a second one, parallel, that can be [but rarely is] used for general aviation (RWY 35-17). Parallel to the main runway, the main taxiway runs from the north apron to RWY 36 holding point. Save for particular circumstances, RWY 36 is normally in use. In order to reach the holding point of RWY 36, general and business aviation aircraft, parked at the west apron, must travel along the taxiway running north of the main runway, and then go through the main taxiway, which is also used by commercial flights. Hence, the single main taxiway can constitute a bottleneck that introduces ground traffic congestions and delays that can be avoided by means of a properly designed optimization algorithm for defining the optimal aircraft ground sequence. In addition, the necessity to use the single runway in mixed mode (concurrently for both departures and arrivals) constitutes a challenge for an algorithm that has the objective to define the overall optimal flights schedule.
III. Proposed solution approach

As previously mentioned, to yield the best results on the overall efficiency of the airport system, any stakeholder decision should be thought of as global rather than local. However, because of the high complexity of the problem, manually finding global solutions is simply not viable. A properly defined optimization algorithm can therefore represent a valuable support to help operators take decisions and exercise control on the overall process. Following EU guidelines ([8], [1] with SESAR Essential Operational Changes and [9]), such algorithm should consider Arrival MANager (AMAN), DMAN, SMAN and ACDM concurrently to obtain a global solution and provide ATCOs with the optimal Target Start up Approval Time (TSAT) and Target Take-Off Time (TTOT) for departures, and Target LanDing Time (TLDT) for arrivals.

In order to obtain a solution for the integrated DMANSMAN-AMAN problem, the presented study followed the works of Kjenstad et al. ([10], [11]), which were applied to German Hamburg airport (where there are two runways) and Swedish Arlanda airport (where there are three runways), and have been considered as the baseline formulation. Their approach consisted in an heuristic decomposition of the integrated problem in three sub-problems (ground routing problem, runway scheduling problem and ground scheduling problem), all modelled as Mixed Integer Linear Programming (MILP). Although this approach may not give the optimal solution, it allows to dramatically reduce the computational effort, giving the solution almost in real time. It is therefore suitable for dynamically following the unavoidable and unpredictable changes present in real-world scenarios (e.g. traffic or meteorological variations, closing of a runway, etc.), providing ATCOs (and potentially other stakeholders) with up-to-date information and cues.
Some modifications and additions have been applied to the cited baseline formulation, in order to improve it on one hand and to better fit it to the context of Linate on the other.

**Ground routing problem.** This is the first step considered by the algorithm (SMAN). The aim is to compute, for each aircraft, a feasible route from its parking stand to the RWY and vice-versa, minimizing taxi time and exploiting all airport resources. Developed Linate airport topology is represented in figure 2 with an oriented line graph. Green colour is assigned to parking positions, while double arrow arcs symbolise parking stands with push-back. The runway is depicted in blue, while red nodes represent holding points: as in the real airport, they are useful to the algorithm to let aircraft wait and avoid conflicts, and for this reason they are used in step 3 to obtain an optimal (feasible) solution to the ground schedule problem. Nodes indicated as $Q_i$ represent release points for push-backs. Defining $u^a_f$ a binary variable which considers if an arc $a$ of the airport graph is assigned or not to flight $f$, and $l^a_f$ the running time for $f$ through $a$, the objective function can be written as:

$$
\min \sum_{f \in F} \sum_{a \in A} u^a_f \left( l^a_f + \frac{0.1}{\text{card}(F)} \sum_{f' \in F} u^{a'}_{f'} \right). \tag{1}
$$

Ground routing problem is a shortest path problem, and has been modelled as a modified maximum flow model, in which the units to send from the source to the sink are flights $F$ that have to be routed. The running time $l^a_f$, which represents a cost associated to each arc, has been extracted from ACDM platform.

It can be noted that the second term within parenthesis increases the time cost $l^a_f$ of arc $a$ proportionally to the usage of that arc by all considered flights, and is pre-multiplied by the term $\frac{0.1}{\text{card}(F)}$ in order to assign a lower weight to resources utilization with respect to the choice of the shortest path. This term was not present in Kjenstad et Al.’s works, and has been conceived to allow the algorithm to utilise every single resource of the airport and optimise the traffic flow; without this, in fact, a traffic congestion could happen if too many aircraft are assigned the same path (although, as it will be shown, step 3 tries to cancel traffic delays). Additionally, if the cost-time of one particular arc is lower, even only slightly lower than that of another, the algorithm would always assign the former in the calculated shortest path for aircraft, reducing, in practice, the exploitation of the full airport capacity (this can be the case, for example, of multiple instances of almost time-equal taxiways running towards/from parallel runways or de-icing zones, or in general in the airport layout). The downside of this approach is that the model becomes non linear (Non Linear Programming [NLP]) in the variable describing the usage of arcs of the graph $u^a_f$, so the problem is NP-Hard, but it has been verified that the impact of this consequence on computational time is minimum.

The constraints that have been considered regard entry point (parking position for departures and runway for arrivals), exit point (runway for departures and parking position for arrivals), balance from an arc to another, the fact that cycles are prohibited, and the impossibility to run a specific taxiway if that particular taxiway is unusable for the considered aircraft (i.e. a liner can’t pass through the west apron). Arrival and departure gates are assumed assigned (by the Airport Operator in ACDM), and cannot be changed.

**Runway scheduling problem.** This is step 2 of the integrated problem decomposition, whose goal is to find an optimal scheduling for arrivals and departures at the RWY (DMAN+AMAN). Desired take-off and landing times are defined. Because of Eurocontrol-related necessities, Calculated Take-Off Time (CTOT) can be assigned to a departing flight, therefore it must depart at that particular time. If CTOT is not assigned, then Expected Take-Off Time (ETOT), computed as Estimated Off-Block Time (EOBT) + Estimated taxi-Out Time (EXOT) is used as desired take-off. Differently from the baseline formulation, flights with assigned CTOT must always take-off within their Slot Tolerance Window (STW), while others could be dropped by the algorithm. In this case a new ETOT and, consequently, a new Departure Tolerance Window (DTW), will be assigned to that flight. The same strategy is applied to arriving aircraft, for which the desired landing
time is Estimated Landing Time (ELDT), around which Arrival Tolerance Window (ATW) is defined. Hereafter it will be indicated with $\delta$, the time at which a particular departure $d$ is expected to take-off, i.e. either ETOT or CTOT, and with $\lambda$, the ELDT associated with a particular arrival $l$.

CTOT, ETOT and ELDT have been extracted from ACDM, while tolerance windows for departing aircraft STW and DTW have been defined as per Eurocontrol [3]. ATW, instead, has been determined taking in consideration the amount of time every aircraft spends to move from the holding point to the runway, and the fact that the approach phase is quite critical, so its time variation should be limited by ATCOs. Tolerance windows are then defined as:

- **DTW**: by default 15 min. before and 15 min. after ETOT;
- **STW**: by default 5 min. before and 10 min. after CTOT;
- **ATW**: fixed at 15 min. before and 5 min. after ELDT.

Following Kjenstad et Al.’s formulation, let $\alpha_d$ and $\beta_l$ be the lowest times associated with the tolerance window of a departing flight ($H_d$, which can either be DTW or STW) or of an arriving flight ($H_l$, equal to ATW), and $\alpha_l$ and $\beta_l$ the highest values. So $H_d = \{\alpha_d, \beta_d\}$ and $H_l = \{\alpha_l, \beta_l\}$ and $\delta_d \in H_d$ $\delta_l \in H_l$. Finally, the time horizon $H$ is the time window between the lowest $\alpha$ and the highest $\beta$ among all the flights that the algorithm has to schedule, for which $H_d \subseteq H$ and $H_l \subseteq H$ (figure 3).

For each departure (arrival) $f \in F$ and each time period $t \in H_f$ a binary variable $x_{ft}$ is introduced which is 1 if and only if $f$ takes-off (lands) at time $t$. Taking-off or landing at time $t$ has a cost $c_{ft}$. For departure $d$ (arrival $l$) such cost increases with $|t - \delta_d|$ ($|t - \lambda_l|$). For each departure $d$ without a CTOT, a binary variable $y_d$ is introduced which is equal to 1 if and only if $d$ is dropped. Dropping a departure $d \in D$ has large cost $w_d$ (fixed, for computational reasons, to the speculative value of 100).

Basically, the algorithm attempts to assign, within each specific tolerance window $H_f$, a departure time or an arrival time ($x_{ft} = 1$): the former is the optimal TTOT and the latter is the optimal TLDT for the integrated problem DMAN+AMAN. As an addition to the baseline formulation, in the presented approach if take-off time cannot be assigned to a particular departure ($y_d = 1$), that flight will be iteratively postponed until time fits the global schedule.

The objective function can be formulated as the minimization of the cost of dropped flights plus overall deviation from the desired arrival and departure times:

$$\min_{d \in D} w_d \cdot y_d + \min_{f \in F, t | x_{ft} = 1} c_{ft} \cdot x_{ft}$$

Some constraints have been introduced in the model for the purpose of taking into consideration operative procedures, like the assumption that an arriving aircraft will always land (i.e. go-around and/or emergency procedures are not considered), that a departing aircraft with CTOT assigned must take-off while others can be dropped (at high cost), and that an aircraft cannot take-off before it has reached the runway i.e. not earlier than Target Off-Block Time (TOBT) + EXOT. Moreover, time separation between arrivals and departures has been modelled, in order to consider wake vortex turbulences and standard arrival/departure procedures.

Since the model has a linear objective function and constraints, but integer variables, it belongs to the class of Integer Linear Programming (ILP) problems.

**Ground scheduling problem.** As in Kjenstad et Al.’s work, this is step 3 of the integrated problem decomposition, whose goal is to establish the time $t$ (continuous variable) at which a flight $f \in F$ should enter every node and arc of its route $r_f = \{v_0, a_1, v_1, a_2, ..., a_n, v_f\}$, for obtaining a completely conflict-free schedule, and for guaranteeing smooth traffic flow
through taxiways. It is necessary to associate a schedule vector \( t_i = (t_i^v, t_i^r, t_i^p, t_i^q, \ldots, t_i^v, t_i^v) \) with the route of each flight (SMAN). The overall schedule \( t \) must:

- assign a schedule time [input time] to arcs and nodes of shortest paths computed at step 1;
- satisfy the order of arrivals and departures on the runway established at step 2;
- obey to all precedence and separation constraints;
- minimise overall taxi time, that is the time that aircraft spend between the parking position and the runway with engines on, and vice-versa.

**FIGURE 2 —** Graph of Linate airport.

**FIGURE 3 —** Times and tolerance windows for the runway scheduling problem.

With \( t_{i_{arr}}^{g(i)} \) the time an arrival aircraft is scheduled to arrive to its gate is denoted, while \( t_{i_{depart}}^{g(i)} \) indicates the entry time in the arc following the node representing the gate, that is the time a departing aircraft leaves its gate. These times, from the algorithm perspective, correspond to Target In-Block Time [TIBT]-the former-and to TSAT-the latter. Entry and exit points at the RWY, computed at step 2, are indicated with \( t_{RWY}^{in} \) [TLDT] and \( t_{RWY}^{out} \) [TTOT]. The objective function can thence be formulated as:

\[
\min \sum_{i\in\mathcal{L}} (t_{i_{arr}}^{g(i)} - t_{i_{RWY}}^{RWY}) \mathcal{P} \sum_{e\in\mathcal{D}} (t_{i_{RWY}}^{RWY} - t_{i_{arr}}^{g(i)}) \tag{3}
\]
Ground scheduling problem can be seen as a job-shop scheduling problem, in which aircraft represent jobs to be processed by machines (airport resources like nodes and arcs of the airport graph).

The schedule must then satisfy simple constraints, such as the observance of the optimal runway schedule found at step 2, the compliance with the route sequence found at step 1, the fact that an aircraft cannot stop on arcs (but only at parking positions and holding points) and cannot leave its stand before the last updated TOBT derived from ACDM platform. Moreover, disjunctive pairs of constraints must be modelled (using binary variables), because two aircraft cannot occupy the same node at the same time and must be separated in time either for safety reasons or for operative procedures (like at parking positions or release points). Holding points, where aircraft can queue for holding, constitute an exception to the latter constraint.

Since the model has a linear objective function and constraints, but both integer and continuous variables, it belongs to MILP problems.

Integrated vision of the algorithm. The presented algorithm has been applied to two case-study days for which ACDM data has been made available to authors. This means that since the algorithm has been run off-line, the progress of time has been simulated with a fictitious parameter. With this parameter, it has been possible to implement some new logic blocks with respect to the baseline formulation, and local procedures (airport regulations, ATCO procedures, etc.) have also been applied.

The algorithm flowchart is presented in figure 4, where the new logic blocks are highlighted with red hexagons:

- if at step 2 take-off time can’t be assigned to a departure (with no CTOT), the flight is dropped and its ETOT is postponed until an optimal TTOT is found (as it has been pointed out above);

- if a departure is scheduled within 15 minutes from current time, that flight is scheduled so its TSAT, TTOT and path cannot be modified any more, in order to give ATCO the final optimal values;

- similarly, if an arrival is scheduled within 15 minutes from current time, that flight is on final so its TLDT, TIBT and path cannot be modified, in order to give ATCO the final optimal values.

From the flowchart it can be understood that, at each iteration, the algorithm has to concurrently schedule new and old flights, taking into consideration the fixed optimal times and paths already computed for scheduled and on final flights (which are not yet take-off or on-blocks), and re-optimising aircraft that do not have fixed times or paths (among which dropped flights are).
IV. Results

Presented NLP, ILP and MILP problems have been implemented in AMPL modelling language [12], and solved by CPLEX solver version 12.6.3.0 on a PC with Intel i7 CPU, 4 cores (running at 1.6 GHz) and 4 GB RAM. The algorithm has been applied to all flights of the two case-study days: November, 8th 2016 and February 15th 2017 (two days without particular traffic congestion problems). The average run-time for the scheduling simulation of all flights (almost 300 per day) was 25 seconds, with less than 0.1 seconds for solving each step, confirming that heuristic decomposition is effectively useful for obtaining very low computational time. Results have then been compared with what actually happened on those days.

As a means to better describe the presented algorithm's *modus operandi*, let us introduce two definitions:

- **target values**: optimal values computed by the algorithm (TTOT, TSAT, TLDT and TIBT), and values estimated by ACDM platform, which are not derived from an optimization routine (TSAT, TTOT).
actual values: actual times at which flights operate on the airport following a First Come First Served (FCFS) procedure, which are recorded in ACDM platform (Actual Off-Block Time [AOBT], Actual Take-Off Time [ATOT], Actual Start up Approval Time [ASAT], Actual In-Block Time [AIBT], Actual Landing Time [ALDT]).

For the comparison between FCFS and optimal procedures three different problems arose:

- it was not possible to directly compare target values, because they were not available for arriving aircraft since AMAN is not implemented at Linate so far;
- it was not possible to directly compare actual values, because the algorithm was to be run off-line, so optimal ones do not exist;
- it was not possible to compare actual values with optimal target values, because this would have delivered too optimistic results.

For these reasons, it has been decided to compute an estimation of actual values also for the optimal case, and compare real FCFS values with these fictitious optimal ones. This has been done deriving from ACDM delay information for flights following FCFS procedure, and considering that part of such delay was due to airport operations, especially at parking positions (differences between actual and authorization start up and push back times, handling procedures, and others implemented by the airport stakeholders). This delay, then, has been added to optimal target values in order to simulate real operations following the optimization algorithm.

In order to define a set of parameters suitable to evaluate the quality of results, a preliminary consideration has been done. It is quite common, in fact, to use delay, defined as the difference between the actual time of occurrence of a particular event and its target time, as a judging parameter. The evaluation logic for delay is obviously the less, the better, but given that it’s a signed quantity, since actual time of occurrence can anticipate target time, this logic leads to the consequence that negative values are highly desirable. However, for a number of events considered in the present study, both delay and advance have a negative impact on optimal ATM. It has therefore been introduced, and will be used to evaluate some of the results, a different quantity, defined as the absolute value of delay: time deviation.

To judge the quality of results, three parameters have been considered: average time deviation at the runway, mean taxi time and mean fuel consumption. These are useful to understand whether the proposed algorithm is able to meet SESAR objectives, like time deviation at runway threshold, reduction of fuel consumption, increase of traffic fluidity, enhancement of safety along taxiways and stakeholders consciousness. Other values, like time deviation at parking position (which is commonly used for estimating airports performances and quality levels), would not have given the same match grade with European objectives. Moreover, the proposed algorithm acts on departing time of aircraft from parking positions in order to have less traffic on taxiways and to have less delay during the overall flight, so aircraft could, in theory, wait some additional minutes at parking in order to optimise global efficiency.

Time deviation at the runway has been computed from the absolute value of the mean difference between actual and desired values\(^1\), taxi time from the mean difference between actual off-block (in-block) and take-off (landing) times, and fuel consumption starting from the mean difference between actual start up (shut down) and take-off (landing) times. The computation of fuel consumption followed ICAO directives [13], which state that the fuel used throughout taxi run can be estimated, to a first approximation, taking the fuel flow data from the Engine Emissions Data Bank, and knowing the number of the aeroplane’s engines.

\(^1\) For departures, operative delays at the runway are already computed by the algorithm, so actual optimal times are considered equal to computed target times. This way, desired values for the optimal case are CTOT and ETOT, while for FCFS procedure TTOT computed by ACDM has been utilised. Similarly for arrivals, for which the desired time is ELDT, both for optimal and FCFS procedures.
Results obtained from the analysis of November 11\textsuperscript{th} are reported in tables I, II and III, where it can be seen that the algorithm can optimise the flights scheduling, with the exception of time deviation at the runway, in which the algorithm obtained the same results of the controllers following FCFS procedure. This derives from the fact that November 11\textsuperscript{th} day was not a critical day in terms of traffic congestion, and shows that the algorithm performance isn’t worse than ATCOs’.


<table>
<thead>
<tr>
<th></th>
<th>LDTD</th>
<th>Taxi time</th>
<th>Fuel usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.19 min (1.20)</td>
<td>4.13 min (2.11)</td>
<td>7.8 ton</td>
</tr>
<tr>
<td>FCFS</td>
<td>1.61 min (1.37)</td>
<td>4.30 min (0.94)</td>
<td>8.1 ton</td>
</tr>
<tr>
<td>Opt vs. FCFS</td>
<td>-26%</td>
<td>-4%</td>
<td>-4%</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>TOTD</th>
<th>Taxi time</th>
<th>Fuel usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>2.38 min (3.73)</td>
<td>10.18 min (3.91)</td>
<td>22.1 ton</td>
</tr>
<tr>
<td>FCFS</td>
<td>2.38 min (2.89)</td>
<td>11.31 min (4.43)</td>
<td>23.7 ton</td>
</tr>
<tr>
<td>Opt vs. FCFS</td>
<td>-10%</td>
<td>-10%</td>
<td>-7%</td>
</tr>
</tbody>
</table>

TABLE III — All flights results of 8/11/2016.

<table>
<thead>
<tr>
<th></th>
<th>Time Deviation</th>
<th>Taxi time</th>
<th>Fuel usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>3.55 min</td>
<td>14.31 min</td>
<td>29.9 ton</td>
</tr>
<tr>
<td>FCFS</td>
<td>3.99 min</td>
<td>15.61 min</td>
<td>31.9 ton</td>
</tr>
<tr>
<td>Opt vs. FCFS</td>
<td>-11%</td>
<td>-8%</td>
<td>-6%</td>
</tr>
</tbody>
</table>

Results obtained from the analysis for February 15\textsuperscript{th} are reported in tables IV, V and VI, where where better results can be noted with respect to both FCFS and case-study day #1.

TABLE IV — Arrivals results of 15/2/2017.

<table>
<thead>
<tr>
<th></th>
<th>LDTD</th>
<th>Taxi time</th>
<th>Fuel usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.08 min (1.38)</td>
<td>3.65 min (5.29)</td>
<td>5.7 ton</td>
</tr>
<tr>
<td>FCFS</td>
<td>1.71 min (5.03)</td>
<td>4.01 min (0.30)</td>
<td>7.5 ton</td>
</tr>
<tr>
<td>Opt vs. FCFS</td>
<td>-37%</td>
<td>-9%</td>
<td>-23%</td>
</tr>
</tbody>
</table>

TABLE V — Departures results of 15/2/2017.

<table>
<thead>
<tr>
<th></th>
<th>TOTD</th>
<th>Taxi time</th>
<th>Fuel usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>2.36 min (3.15)</td>
<td>9.59 min (3.88)</td>
<td>20.6 ton</td>
</tr>
<tr>
<td>FCFS</td>
<td>2.72 min (2.54)</td>
<td>11.70 min (4.39)</td>
<td>24.5 ton</td>
</tr>
<tr>
<td>Opt vs. FCFS</td>
<td>-13%</td>
<td>-18%</td>
<td>-16%</td>
</tr>
</tbody>
</table>

TABLE VI — All flights results of 15/2/2017.

<table>
<thead>
<tr>
<th></th>
<th>Time Deviation</th>
<th>Taxi time</th>
<th>Fuel usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>3.44 min</td>
<td>13.24 min</td>
<td>26.3 ton</td>
</tr>
<tr>
<td>FCFS</td>
<td>4.43 min</td>
<td>15.71 min</td>
<td>31.9 ton</td>
</tr>
<tr>
<td>Opt vs. FCFS</td>
<td>-23%</td>
<td>-16%</td>
<td>-18%</td>
</tr>
</tbody>
</table>

In all tables, values of the standard deviation of Take-Off Time Deviation (TDTD), LanDing Time Deviation (LDTD) and taxi time are presented in brackets. It can be noted that for arrivals LDTD standard deviation is lower for the optimum solution than for FCFS, while for departures TOTD it is greater: this derives from the fact that at Linate arrival times are at present not taken into particular consideration, while departing flights are already
managed following some kind of optimization process. Moreover, dropped flights could compromise this value, so some limitations on re-iterations should be considered with airport stakeholders for the purpose to obtain both timeliness and low deviation from the mean value. On the contrary, the optimal standard deviation of taxi time for departing aircraft, which at Linate is by far more important than for arrivals, is greater for optimal than for FCFS, meaning that having a general view of the optimization process works better than merely concentrating on time deviation values.

TOTD is presented in figure 5, in which it can be noted that the algorithm is well capable to accomplish the desired time of departure. Particular attention should be given to the zero time deviation columns: compared to FCFS performance, with the optimal solution 159 vs. 96 flights (63 more, +65%) depart at their desired time, meaning that they are neither late nor in advance, and this is a good result for the general management of airport resources. Similar considerations can be drawn for landing aircraft, whose results are represented in figure 6 in terms of LDTD.

FIGURE 5 — TOTD for both case-study days.

Outbound Taxi Time Difference [OTTD], defined as the difference between optimal and FCFS taxi time, is presented in figure 7. It can be noted that most aircraft have negative values, meaning that the optimal scheduling is capable to save taxi time with respect to FCFS. Note that in this figure inbound taxi time is not considered, since at Linate this is not of particular interest because parking positions are very close to the runway exit points.
In general, if controllers had been able to follow the optimal scheduling, they would have saved a few-but valuable-minutes in terms of time deviation at the runway and taxi time. This corresponds to have less airport noise, to increase safety (since there would have been less aeroplanes simultaneously running on taxiways), and also to save a considerable amount of fuel during the taxi. Therefore, as a last analysis $\text{CO}_2$ potentially saved by the algorithm has been calculated. Following ICAO directives [13], with a conversion ratio of $3.16 \text{kg}_\text{CO}_2/\text{kg}_\text{fuel}$ it can be computed that, on the first day 6.9 ton of $\text{CO}_2$ would not have been emitted, while on the second day the saving would have been 17.8 ton, respectively almost equivalent to an average 20 kg and 56 kg of $\text{CO}_2$ saved by each aircraft in the two days. In addition, taking the mean value of the price of Jet A-1 fuel for November 2016 and February 2017 [14], monetary saving potentially available for airlines thanks to the algorithm effectiveness could be evaluated: Alitalia, the major airline company operating at Linate, would have saved an average of about 1,900 Euro each day.

V. Conclusions and future works

ATM improvement is a fundamental objective for the EU, and through SESAR programme important results have been achieved. Looking at the Essential Operational Changes defined within the European ATM Master Plan, the work presented in this paper tried to understand if ATM could be improved at Linate airport. The objective was to design an algorithm capable to help airport stakeholders, in particular ATCOs, to take decisions on aircraft start up time, take-off time and landing time, optimizing the global efficiency of the airport system. By exploiting specific tools of Operational Research, the DMAN-SMAN-AMAN integrated problem has been heuristically decomposed in three sub-problems and adapted to the local context. The comparison between algorithm results and what actually happened on two case-study days shows potential benefits in reduction of average values of flight untimeliness, taxi time and fuel consumption, yielding a lower noise impact, an increased safety, and a considerable save on $\text{CO}_2$ emissions and money every day.

The presented algorithm permits to obtain the described good results with very low computational time, substantially improving ATM at Linate airport. Additional analyses should be conducted taking into consideration different operative conditions, possibly running the algorithm in real-time, in order to compare the actual optimal values with the target ones.

Future developments may include dynamic computation of delay along taxiways, in order to achieve a complete Variable Taxi Time (VTT) calculation and a full implementation of
SMAN. In this work, in fact, running times on taxiways have been taken from airport ACDM data, in which they are considered fixed-and therefore independent from meteorological conditions. For future works, however, a computation of the actual running time in every single time of the day and day of the year—for every relevant meteorological condition—would be crucial for a further improvement of the algorithm output reliability.

Additionally, de-icing shall be modelled, but only if and when confidence on times of that particular phase will be sufficiently high: for the optimization algorithm efficiency sake, in fact, the availability of precise and up-to-date input data it is pivotal. Since de-icing procedures strongly depend on the type of ice, and type of aircraft and operator’s own procedures, in this work it has not been implemented.

Moreover, procedures for taking in consideration the characteristics of the recently-introduced electric taxi capability could also be developed and integrated, since it introduces remarkable differences in taxi times, push back procedures, runaway crossing and other taxi-related details.

Finally, an implementation in a larger airport, like for example Milano Malpensa, could be interesting, because it has two main runways and a complex taxiway network, a situation that can the presented algorithm advantages, delivering, as for Linate, a more sustainable and high-performing aviation.

Acknowledgment

Special thanks to SEA (Società per Azioni Esercizi Aeroportuali) personnel who gave us, with patience and kindness, all information and support needed to test the algorithm.

List of acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACDM</td>
<td>Airport Collaborative Decision Making</td>
</tr>
<tr>
<td>AIBT</td>
<td>Actual In-Block Time</td>
</tr>
<tr>
<td>ALDT</td>
<td>Actual Landing Time</td>
</tr>
<tr>
<td>AMAN</td>
<td>Arrival MANager</td>
</tr>
<tr>
<td>AOBT</td>
<td>Actual Off-Block Time</td>
</tr>
<tr>
<td>ASAT</td>
<td>Actual Start up Approval Time</td>
</tr>
<tr>
<td>ATC</td>
<td>Air Traffic Control</td>
</tr>
<tr>
<td>ATCO</td>
<td>Air Traffic Controller Operator</td>
</tr>
<tr>
<td>ATM</td>
<td>Air Traffic Management</td>
</tr>
<tr>
<td>ATOT</td>
<td>Actual Take-Off Time</td>
</tr>
<tr>
<td>ATW</td>
<td>Arrival Tolerance Window</td>
</tr>
<tr>
<td>CTOT</td>
<td>Calculated Take-Off Time</td>
</tr>
<tr>
<td>DMAN</td>
<td>Departure MANager</td>
</tr>
<tr>
<td>DTW</td>
<td>Departure Tolerance Window</td>
</tr>
<tr>
<td>E-AMAN</td>
<td>Extendend - Arrival MANager</td>
</tr>
<tr>
<td>ELDT</td>
<td>Estimated Landing Time</td>
</tr>
<tr>
<td>EOBT</td>
<td>Estimated Off-Block Time</td>
</tr>
<tr>
<td>ETOT</td>
<td>Expected Take-Off Time</td>
</tr>
<tr>
<td>EXOT</td>
<td>Estimated Taxi-Out Time</td>
</tr>
<tr>
<td>FCFS</td>
<td>First Come First Served</td>
</tr>
<tr>
<td>ICAO</td>
<td>International Civil Aviation Organization</td>
</tr>
<tr>
<td>ILP</td>
<td>Integer Linear Programming</td>
</tr>
<tr>
<td>LDTO</td>
<td>Landing Time Deviation</td>
</tr>
</tbody>
</table>
REFERENCES

STAKEHOLDER COOPERATION FOR IMPROVED PREDICTABILITY AND LOWER COST REMOTE SERVICES

Joen Dahlberg, Tatiana Polishchuk, Valentin Polishchuk, Christiane Schmidt
Communications and Transport Systems, Linköping University
firstname.lastname@liu.se

Abstract—We consider the problem of adjusting flight schedules (arrival/departure slots) in order to optimize staffing in a Remote Tower Center. We explore tradeoffs between the number of affected flights, the times by which the movement slots are shifted in comparison with the original schedules, the number of airports controlled from one Remote Tower Module, and the number of modules necessary to provide air navigation services to the five Swedish airports with Remote Towers. We consider different variants of the problem (allowing different times for shifts, different numbers of airports that can be assigned to a module etc.), and prove one variant NP-complete, give polynomial algorithms for others, and formulate a general version as an integer program (IP). Our results show that cooperation between airlines, airport owners and ANSPs may help in reduction of Remote Tower Center operation costs by requiring fewer controller positions handling traffic at the airports.

I. Introduction

Predictability and cost efficiency are among the most sought-after key performance indicators (KPIs) which both Air Navigation Service Providers (ANSPs) and airspace users (airlines) have strived to improve via a variety of actions. It has been generally accepted that efforts of a single actor (e.g., ANSP alone, or a single airline) might not have as large a potential for further improvements as a cooperative act of several stakeholder—cases in point are CDM (Collaborative Decision Making), UDPP (User-Driven Prioritization Process), and related initiatives. This paper explores opportunities for collaboration between airlines, airport operators and an ANSP: Is it possible to play with flight schedules in order to smooth operations and save staffing costs in a remote air navigation service center?

Remote Towers Services (RTS) are one of several technological and operational solutions that the SESAR Joint Undertaking delivers to the ATM community for deployment. Today, many small aerodromes struggle with financial difficulties, and a large cost is air traffic control. The RTS concept implementation splits the cost of Air Traffic Services (ATS) provision and staff management between several airports, providing significant cost savings for small airports (30-120 movements a day). The difference in terms of investment is also significant when comparing installation of sensors to the construction of a new tower. Maximizing the efficiency of human resources (HR) is of particular importance because labour accounts for up to 85% of air traffic service (ATS) costs [23].
Motivation

Remotely Operated Towers (ROT) are a novel approach to air navigation service provision via digitization and integration of airport functions. Within the Remote Tower concept it becomes possible to control multiple aerodromes from a single remote location. In particular, a single air traffic controller (ATCO) works with more than one airport from a single position in the Remote Tower Module (RTM). In such a setting, it is of utmost importance to ensure that movements never occur simultaneously in the jointly controlled airports, as one controller may not be able to cope with multiple movements that occur at two airports simultaneously: professional ATCOs, participating in a study on their ability to operate two airports remotely within a single RTM, stated that situations with simultaneous departures and landings are critical for safety and sometimes “impossible to handle” [14]. That is, if two airports simultaneously have movements, they cannot be controlled from the same RTM; in the extreme case, if e.g. simultaneous movements occur in all 5 Swedish airports with remote towers, then 5 RTMs will be needed, undermining the whole idea of operational cost sharing. This motivates investigation of possibilities to perturb flight schedules in order to provide smoother operations of the Remote Tower Center (RTC).

Indeed, airlines create the flight schedules looking only at the individual airports (flights origins and destinations) and their constraints, disregarding the possibility that an airport may be controlled remotely together with another one. It may thus be the case that airlines would be willing to move around their slots: airlines do not adhere to their schedules precisely anyway, so few slot shifts should not be a deal breaker. Moreover, in remote airports slotting is not an issue and hence the parties may hope that the airlines will actually arrive within the allocated slots. Last but not least, the ATCOs may anyway have to put one of the aircraft on hold (even if the aircraft are in different airports) should simultaneous movements occur.

Roadmap

In the remainder of this section we review related work. The next section gives formal definitions related to our optimization problem and settles its computational complexity. In Section III we formulate our problem as an Integer Program. In Section IV we feed the model with real flight data for five Swedish airports planned for remote operation and study how the simultaneous movements in the airports can be handled. Section V concludes the paper and outlines future work directions.

Related work

RTC aims at providing ATS for multiple airports by air traffic controllers located remotely as defined in [15]. Researches studied various aspects of the RTS concept. Möhlenbrink et al. [12] and Papenfuss et al. [18] considered usability aspects within the novel remote control environment. Wittbrodt et al. [28] stress the role of radio communication in the context of a remote airport traffic control center. In a safety assessment of the Remotely Operated Tower (ROT) concept, Meyer et al. [11] suggest functional hazard analyses and pinpoint the issue of getting reliable probability values for the models. Oehme and Schulz-Rueckert [16] propose a sensor-based solution for aerodrome control that removes the dependency on visibility conditions and tower location. In [5], [14], [13], [10] and [17] various aspects of work organization and human performance issues related to the remote operation are considered. The authors propose several methods to control two airports from a single center. Using simulations they studied how the monitoring performance may influence the system design and behavioral strategies, and suggested several ideas on the design of novel RTC workplaces.

Distributing the total traffic load between controller positions is the subject of sectorization research—a well studied area in ATM; see, e.g., the survey [4] and references therein. While for sectorizations of the Terminal Maneuvering Area (TMA) of an airport the goal is to distribute the monitoring task from a single airport to several air traffic controllers.
(ATCOs), for an RTC we want to merge the monitoring task of several airports into a single Remote Tower Module (RTM) (then staffed by a single ATCO).

Assigning airport traffic to Remote Tower Modules was initially considered in [1]. That model did not take into account the possibility to switch assignments during the day (that is, an airport was assigned to the same module throughout a complete day) or load balancing between the different modules. Josefsson et al. [9] create an optimization framework with multiple objectives and additional constraints based on the model proposed in [1], and demonstrate how it enables personnel planning at RTCs on real data. In [8] Josefsson et al. take a more detailed look at the problem, and schedule the actual ATCO shifts in the RTC taking into account factors such as maximum time “in position”, minimum rest times, etc. Also, for this framework, several objectives are considered, and the authors present how the framework facilitates RTC controller shift planning on real data. In particular, a limitation was placed, preventing the airports from being assigned to the same controller for the whole hour during which the potential conflict occurs. During the conflict hours more controllers are obviously needed because the airports with the potential conflicts are to be controlled by separate controllers. The resulting statistics show a noticeable increase in the lower bound on the total number of controllers for the day with the highest traffic load during the year 2016 (from eight necessary controllers without conflict avoidance to 10 necessary controllers with conflict avoidance). In [8], this constraint is ensured by postprocessing: if airports that are in conflict during a period are scheduled for the same module, local adjustments to the assignment are made to break this. The results show that there is a need to adjust initial airport schedules to optimize their scheduling within RTC, which will provide immediate HR savings. That is, one way to circumnavigate the need for more staff at the RTC could be to ask airlines for slight adjustments to their schedules, by which they would contribute to cost savings via a decreased RTC staff demand.

In general, slot (re)allocation and trading is a recurring topic in ATM [21], [26], [27], [3], [19], [20], [2], [7].

II. Preliminaries

We start from describing the airports and defining the considered optimization problems; we also comment on the problems computational complexity.

Input: Airports and Conflicts

Over the last years, Swedish ANSP Luftfartsverket (LFV) has been working on the deployment of the RTS concept as an alternative to traditional ATS. In 2015 and 2016 LFV and a Swedish airports operator conducted a joint feasibility study to analyze the impact of the transition from traditional tower ATS to RTS for even more airports in Sweden. The study confirmed that RTS is technically and operationally feasible, the level of risk is manageable, and that it is deemed financially advantageous to use RTS for these airports.

LFV provides remote air navigation services for Örnsköldsvik Airport since April 2015. Two additional airports, Sundsvall-Midlanda and Linköping SAAB, will be connected during 2017. Eventually, it is planned to control five additional airports from a single RTC. The general properties of the airports in consideration may be summarized as follows:

- **Airport 1 (AP1)**. Small airport with low traffic, few scheduled flights per hour, non-regular helicopter traffic, sometimes special testing activities.

- **Airport 2 (AP2)**. Low to medium-sized airport, multiple scheduled flights per hour, regular special traffic flights (usually open 24/7, with exceptions).
- **Airport 3 (AP3)**. Small regional airport with regular scheduled flights (usually open 24/7, with exceptions).

- **Airport 4 (AP4)**. Small airport with significant seasonal variations.

- **Airport 5 (AP5)**. Small airport with low scheduled traffic, non-regular helicopter flights.

Our input are aircraft movements at each airport, which we received from the Demand Data Repository (DDR) hosted by EUROCONTROL. We split the time into 5-min intervals, called slots, and put every flight into its slot (e.g., if the arrival or departure time is 0853, the movement is put into the slot 0850-0855, etc.). Formally, our input is a matrix $F$ with 5 rows (a row per airport) and a column per each slot; the entry $F_{as}$ in row $a$ and column $s$ is equal to 1 if a movement happens at airport $a$ at time slot $s$, and is equal to 0 otherwise (see an example input matrix in Figure 1).

**Figure 1** — An example of the input matrix $F_{as}$, where 1's in the matrix cells represent movements.

<table>
<thead>
<tr>
<th>Slot</th>
<th>00</th>
<th>05</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AP2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AP3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AP4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AP5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

At any single airport, at most 1 movement occurs during any slot (therefore the entries in $F$ are only 0's and 1's). However, it often happens that two movements occur during the same slot in different airports; we define this as a **conflict** (in terms of $F$, a conflict is two 1s in the same column). The main constraint in our airports-to-RTMs assignment problem is that conflicting airports should never be assigned to the same RTM.

Figure 2 shows the number of conflicts in the schedules for all airport pairs for the year 2016, while Figure 3 illustrates the number of days during which these conflicts occur. It can be seen that the number of conflicts is very high, and they occur almost every day for most airport pairs.

**Figure 2** — The number of potential conflicts in schedules for each airport pair during the year 2016.

<table>
<thead>
<tr>
<th>Conflict count</th>
<th>AP1</th>
<th>AP2</th>
<th>AP3</th>
<th>AP4</th>
<th>AP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP1</td>
<td>1058</td>
<td>6473</td>
<td>3400</td>
<td>3021</td>
<td>366</td>
</tr>
<tr>
<td>AP2</td>
<td>1058</td>
<td>6473</td>
<td>3400</td>
<td>3021</td>
<td>366</td>
</tr>
<tr>
<td>AP3</td>
<td>621</td>
<td>6473</td>
<td>2603</td>
<td>2517</td>
<td>1449</td>
</tr>
<tr>
<td>AP4</td>
<td>366</td>
<td>3400</td>
<td>2603</td>
<td>2517</td>
<td>1449</td>
</tr>
<tr>
<td>AP5</td>
<td>339</td>
<td>3021</td>
<td>2517</td>
<td>1449</td>
<td>1449</td>
</tr>
</tbody>
</table>

**Figure 3** — The number of days when potential conflicts in schedules occur.

<table>
<thead>
<tr>
<th>Conflict days</th>
<th>AP1</th>
<th>AP2</th>
<th>AP3</th>
<th>AP4</th>
<th>AP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP1</td>
<td>278</td>
<td>285</td>
<td>365</td>
<td>359</td>
<td>365</td>
</tr>
<tr>
<td>AP2</td>
<td>341</td>
<td>366</td>
<td>363</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>AP3</td>
<td>316</td>
<td>366</td>
<td>362</td>
<td>362</td>
<td>362</td>
</tr>
<tr>
<td>AP4</td>
<td>278</td>
<td>363</td>
<td>362</td>
<td>359</td>
<td>359</td>
</tr>
<tr>
<td>AP5</td>
<td>285</td>
<td>365</td>
<td>362</td>
<td>359</td>
<td>359</td>
</tr>
</tbody>
</table>

**Output: Perturbed flights and Airport-to-RTM assignment**

Our goal is to introduce “small” shifts to the flight schedules, leading to decreased number of required RTMs. The extent of a shift may be measured by the maximum slot shift (how far a flight is moved) and by the number of shifted flights. We denote the maximum shift by $\Delta$ and the number of shifts by $S$ ($\Delta$ is measured in minutes and is necessarily a multiple of 5, since we shift only by whole slots).
The last piece of the problem statement is the maximum number of airports per module. In all earlier optimization frameworks for RTC staff scheduling [1], [9], [8] the number of airports that may be consolidated in a single module is a parameter (let MAP denote this bound); for all studies in Sweden MAP is set to 2 (that is, at most 2 airports can be monitored from a single RTM), while in other parts of Europe larger numbers for MAP are considered.

Problem Formulation
We are now ready to formulate the most generic version of our problem:

**Flights Rescheduling and Airport-to-Module Assignment (FRAMA)**

**Given:**
- flight slots in a set of airports (the matrix $F$)
- the maximum allowable shift of a flight, $\Delta$
- the maximum total number of allowable shifts, $S$
- the maximum number of airports per RTM, MAP
- the total number of modules, $M$

**Find:** New slots for the flights and an assignment of airports to RTMs such that
- at most $S$ flights are moved
- each flight is moved by at most $\Delta$
- no conflicting airports are assigned to the same RTM
- at most MAP airports are assigned per module
- at most $M$ modules are used

Note that the above formulation is a decision (or feasibility) problem: there is nothing to optimize in it, i.e., there is no objective function. As with any feasibility problem, it can be turned into an optimization problem by moving one of the constraints into the objective function. When considering the optimization version of FRAMA, our primary objective will be to minimize the number $M$ of the used RTMs, while respecting the bounds $\Delta$, $S$ and MAP.

Complexity and Heuristics for FRAMA

The next section presents our integer program (IP) for FRAMA; even though solving IPs is NP-hard in general, we demonstrate in Section IV that smaller instances of the problem can be solved using commercial off-the-shelf optimization software. In the remainder of this section we discuss the computational complexity of the problem and possible polynomial-time solution approaches.

**Theorem 1.** FRAMA is NP-complete, even if $\Delta = 0$ and $\text{MAP}=3$.

**Proof.** The proof is by reduction from Partition into Triangles (PIT), which was shown to be NP-complete for graphs of maximum degree four by van Rooij et al. [24]. An instance of PIT is given by a graph $G = (V, E)$ (of maximum degree four), and the question is whether $V$ can be partitioned into triples $V_i, V_j, V_k$ such that each $V_i$ forms a triangle in $G$ (that is, such that for each triple of vertices $V_i$ each vertex in $V_i$ is connected to both other vertices in $V_i$).
Given an instance of PIT, that is, a graph \( G = (V, E) \) with maximum degree four, we construct the matrix \( F \), the input of FRAMA, as follows: per vertex we have an airport, that is, \( F \) has \(|V| \) rows. Per non existing edge of \( G \) (that is, for each edge in \( G \)'s complement) we have a time slot. Let \( G' = (V, E') \) be the complete graph on the vertex set \( V \), then we have \(|E' \setminus E| \) time slots, one per edge \( e' \in E' \setminus E \). For the time slot corresponding to \( e' = \{v,w\} \) we add two 1’s to the time slot column: to the airports of \( v \) and \( w \), all other entries in that column are 0’s. See Figure 4 for a graph \( G \) and its complement, and Table I for the resulting matrix \( F \).

**FIGURE 4 — Left: graph \( G \) (instance of PIT), right: the complement of \( G \).**

![Graph G and its complement](image)

**TABLE I — The resulting matrix \( F \) for the PIT instance given by the graph from Fig. 4. grouping the airports into the two triples 1,2,3 and 4,5,6 is equivalent to the triangles of the same vertices in \( G \).**

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Any solution to FRAMA with \( \Delta = 0 \) and MAP=3 then groups the airports, the vertices, into triples, such that there are no conflicts between any of the three airports in a triple, that is, such that there is an edge between any of the three vertices in the triple. Thus, we would obtain a solution to PIT.

Given a solution to FRAMA with \( \Delta = 0 \) (and, thus, \( S = 0 \)) and MAP = 3, it can obviously be verified in polynomial time.

On the other hand, if \( \Delta = 0 \) (no rescheduling) and MAP=2, then minimizing the number of modules is equivalent to finding a maximum matching in the “airport conflict graph” that has a vertex for every airport and an edge between two airports if they can be put into the same module (i.e., if they have no conflicts), see Figure 5 for an example. A maximum matching may be found in polynomial time (see e.g. [22]), implying an efficient algorithm for this (restricted) version of FRAMA (\( \Delta = S = 0 \), MAP=2).

**FIGURE 5 — (a) An example for an airport conflict graph: there is no edge between airports 1 and 2, because they are in a conflict. An edge, for example, between AP1 and AP3, indicates that these two airports are not in conflict. (b) A maximum matching in the airport conflict graph: we match AP1 with AP5 and AP3 with AP4, vertices that are not matched, AP2 in this case, constitute a single module, thus, the maximum matching results in 3 modules in this case.**

![Airport conflict graph and maximum matching](image)

For a general and more interesting case of \( \Delta > 0 \) (i.e., with rescheduling allowed) and MAP=2, we do not know the complexity of FRAMA. One possible approach to the problem is to first remove all the conflicts and then assign the airports to RTMs (i.e., solve the
rescheduling and the assignment separately). The assignment problem is trivial in the absence of conflicts (the airports are arbitrarily packed into the RTMs, with MAP airports per module), so we now discuss how to optimally deconflict the flight schedules: The deconfliction problem is reduced to matching by forming a bipartite graph with all flights in one part and all slots in the other part; a flight \( f \) is connected to all slots within distance \( \Delta/5 \) from its original slot (i.e., to all slots to which \( f \) may be rescheduled). All edges have weight 1, except for the ones between \( f \) and its original slot—this edge has weight 0; see Figure 6 for an example. We now find the minimum-weight matching in the graph that matches all flights (this can be done e.g., with flow techniques, see e.g. [22]). The matching minimizes the total number \( S \) of the shifted flights. If no such matching exists, \( \Delta \) must be increased. (The algorithm may be extended to minimize also the total amount of shifted minutes—just set the weight of each edge equal to the length of the shift.)

**FIGURE 6** — Figure 6. (a) A bipartite graph for the deconfliction problem: vertices representing flights are to the left, vertices representing slots to the right. Black edges indicate the original slot of the flight, gray edges connect a flight to slots within \( \Delta = 10 \), that is, each flight has two \( (\Delta/5 = 2) \) edges to earlier slots and two edges to later slots. Gray edges have weight 1, black edges have weight 0. (b) A minimum-weight matching in the graph, shown in bold, of cost 2.

For a small number of airports (and the given five airports are feasible for this approach), we can also enumerate all pairs of airports, and completely eliminate all conflicts for the given pairs (matching as described above) with a given \( \Delta > 0 \), and check whether any combination of these airports leads to the minimum possible number of modules etc.

While the above heuristic of solving the rescheduling and the assignment separately runs in polynomial time, it may find suboptimal solutions to FRAMA (it is not necessary to remove all the conflicts). In the next section we give a unified approach to rescheduling and airports assignment.

### III. The Integer program

We formulate FRAMA as an Integer Program (IP). We use decision variables \( x_{am} \) to indicate whether airport \( a \) is assigned to module \( m \), \( z_m \) to indicate whether module \( m \) is used, and \( y_{atf} \) to indicate whether flight \( f \) arrives/departs at/from airport \( a \) in time slot \( t \). Moreover, the variable \( w_{ab} \) indicates when there is a conflict between airport \( a \) and airport \( b \) (that is, when \( F_{a,t} = F_{b,t} = 1 \) for some slot \( t \)).

We let \( A, M, T, \) and \( V_a \) denote the set of airports, modules, time slots, and flights at airport \( a \), respectively. The cost to move flight \( f \) at airport \( a \) to time slot \( t \) is \( p_{atf} \), and the scheduled time for flight \( f \) at airport \( a \) is \( s_{atf} \). We let \( \delta \) denote the maximum shift distance for scheduled aircraft in terms of time slots, that is, \( \delta = \Delta/5 \).

\[
\min \sum_{m \in M} z_m + \sum_{a \in A} \sum_{t \in T} \sum_{f \in V_a} p_{atf} y_{atf} \tag{1}
\]

\[
\sum_{t \in T} y_{atf} \leq 1 \quad \forall a \in A, f \in V_a
\]

\[
\sum_{f \in V_a} y_{atf} \leq 1 \quad \forall a \in A, t \in T
\]

\[
\sum_{a \in A} y_{atf} \leq 1 \quad \forall f \in V_a, t \in T
\]

\[
z_m \leq \sum_{a \in A} y_{atf} \quad \forall m \in M, t \in T, f \in V_a
\]

\[
y_{atf} \leq \sum_{m \in M} z_m \quad \forall a \in A, t \in T, f \in V_a
\]

\[
0 \leq y_{atf} \leq 1 \quad \forall a \in A, t \in T, f \in V_a
\]

\[
z_m \in \{0, 1\} \quad \forall m \in M
\]

\[
x_{am} \in \{0, 1\} \quad \forall a \in A, m \in M
\]

\[
y_{atf} \in \{0, 1\} \quad \forall a \in A, t \in T, f \in V_a
\]
The objective function (1) minimizes the number of modules used and the sum of shifts, where \( c_1 \) and \( c_2 \) are weights assigned to the two components of the objective function, number of modules and the sum of shifts. If the number of shifts is minimized, i.e., \( \min \ S \), then \( p_{af} = 1 \) if \( t = s_{af} \) and \( p_{af} = 0 \) if \( t = s_{af} \). If the total amount of shifts is minimized, i.e., we minimize the total amount of shifted minutes, then \( p_{af} = |t - s_{af}| \). The constraint set (2) states that if an airport is assigned to a module, then said module is used. The constraint set (3) states that each airport must be assigned to exactly one module. The constraint set (4) states that no more than one aircraft may arrive/depart at/from each airport and time slot. The constraint set (5) states that each aircraft must arrive/depart in a time slot \( \pm \delta \) from its scheduled time. The constraint set (6) states that if two aircraft arrive/depart in the same time slot at airport \( i \) and airport \( j \) respectively, then there is a conflict. The constraint set (7) states that if there is a conflict, then the two airports may not be assigned to the same module. The constraint set (8) states that at most MAP airports can be assigned to each module.

Our IP formulation of FRAMA optimizes a linear combination \( c_1 M + c_2 S \) of \( M \) and \( S \) (alternatively, we could move one of the terms into constraints, giving an upper bound on it, and optimize the other term). We choose \( c_1 \) and \( c_2 \) such that minimizing the modules is the primary objective. Moreover, the IP computes new slots for flights and assigns airports to RTMs, such that each flight is moved by at most \( D \) (given by \( \delta = \Delta /5 \)), no conflicting airports are assigned to the same RTM (constraint (7)), and at most MAP airports are assigned per module (constraint (8)). Thus, our IP does solve FRAMA.

IV. Experimental study

We use traffic data from the five airports, described in Section II, on October 19, 2016. This is the day with highest traffic in 2016 and, thus, can be considered to be the most difficult day for 2016; using one day of traffic uses the underlying assumption that we
do not want to move flights by more than a day. Altogether 286 flight movements were
scheduled on this day for the five airports. Because of self-induced conflicts, i.e., more
than one flight movement in a given slot at a single airport, we use only 233 movements
for the first set of experiments.

All optimization problems are solved using the optimization solver Gurobi [6]. Python [25]
scripts are used for implementing the optimization problems, as well as handling data
and results. There is one optimization problem for each pair (Δ, MAP) of maximum shifts
and maximum number of airports per module. The objective function weights, c₁ and c₂,
are chosen such that minimizing the number of modules is the primary objective. This
is done by choosing c₁ >> c₂.

Original traffic

For the experiments we vary MAP (and then look at the results for different δ for a fixed
MAP in each case). We start with MAP = 5, see Table II for the results. If no rescheduling
is allowed, we need 5 modules. For δ = 1, that is, if we allow rescheduling of at most ± 5
minutes, it is sufficient to use two modules. To reduce the number of modules to one, we
need δ = 7, that is, we allow rescheduling of ± 35 minutes. Note that we have 12 × 24 = 288
slots for flight movements, that is, with sufficiently large shifts we can accommodate
the 233 flight movements in a single module. In Figure 7 we present the number of
shifts as a function of the maximum shifts (in minutes). This shows the tradeoffs from
allowing more shifts, larger shifts (more minutes) and more APs/module (higher MAP).
The results for MAP = 4, MAP = 3, and MAP = 2 are given in Tables III left, III right and IV,
respectively.

<table>
<thead>
<tr>
<th>δ</th>
<th># of modules</th>
<th># of shifts = S</th>
<th>maximum shift (in mins) = Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>26</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>26</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>26</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>26</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>118</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>108</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>99</td>
<td>45</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>91</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>85</td>
<td>55</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>83</td>
<td>60</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>81</td>
<td>65</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>79</td>
<td>70</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>78</td>
<td>75</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>75</td>
<td>90</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>74</td>
<td>95</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>74</td>
<td>100</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>73</td>
<td>105</td>
</tr>
</tbody>
</table>
FIGURE 7 — Number of shifts as function of the maximum shift (in minutes). Results for a single module (RTM) are shown in blue, results for 2 modules are shown in red. Note that when allowing 4 airports per ATM (MAP = 4) still 2 modules are needed.

In the second set of experiments, we actually use all 286 flight movements scheduled for operation on October 19, 2016: in case of a self-induced conflict, the model shifts either of them, that is, we start with possible more than one flight movement per time slot and airport, and obtain a feasible assignment if there is at most one flight movement per time slot and airport. Thus, for δ = 0 this is infeasible by definition. For the results for MAP = 5 see Table V left. For the 233 movements 2 modules were sufficient with δ = 1, for 286 movements, these are sufficient (and the problem at all feasible) only for δ = 2. Similarly, for 233 movements 1 module was sufficient with δ = 7, for 286 movements 1 module is sufficient only for δ = 37. The results for MAP = 3, 2 are given in Tables V right, VI left and right, respectively. The results for MAP = 5 and MAP = 4 yield the same number of shifts for δ = 0, ..., 4: the constraint related to MAP is not binding for δ < 37, i.e. it has no impact.

TABLE III — Given δ, resulting M and S. Left: MAP = 4, right: MAP = 3.

<table>
<thead>
<tr>
<th>δ</th>
<th>M</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>26</td>
</tr>
</tbody>
</table>

TABLE IV — Given δ, resulting M and S for MAP = 2.

<table>
<thead>
<tr>
<th>δ</th>
<th># of modules</th>
<th># of shifts</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

TABLE V — Given δ, resulting M and S with 286 movements. Left: MAP = 5, right: MAP = 4.

<table>
<thead>
<tr>
<th>δ</th>
<th>M</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>infeasible</td>
<td>infeasible</td>
</tr>
<tr>
<td>1</td>
<td>infeasible</td>
<td>infeasible</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>103</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>79</td>
</tr>
<tr>
<td>36</td>
<td>2</td>
<td>79</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
<td>158</td>
</tr>
<tr>
<td>38</td>
<td>1</td>
<td>154</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>δ</th>
<th>M</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>infeasible</td>
<td>infeasible</td>
</tr>
<tr>
<td>1</td>
<td>infeasible</td>
<td>infeasible</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>103</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>79</td>
</tr>
<tr>
<td>288</td>
<td>2</td>
<td>79</td>
</tr>
</tbody>
</table>
We also evaluate the computation times for the case of solving the instances in two steps: we solve two optimization with $c_2 = 0$ and $c_1 = 0$ respectively and fix the $\sum_{k=1}^n z_k$ to be equal to the optimal number of modules used when solving the second optimization problem. The computation times lie between 1,264 and 146,488 seconds for MAP= 5, see Table VII. The computation times for MAP= 4, MATP= 3, and MAP= 2 are given in Tables VIII, IX, and X, respectively.

**TABLE VI** — Given $\delta$, resulting $M$ and $S$ with 286 movements. Left: MAP= 3, right: MAP= 2.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$M$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>infeasible</td>
<td>infeasible</td>
</tr>
<tr>
<td>1</td>
<td>infeasible</td>
<td>infeasible</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>103</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>79</td>
</tr>
<tr>
<td>288</td>
<td>2</td>
<td>79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$M$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>infeasible</td>
<td>infeasible</td>
</tr>
<tr>
<td>1</td>
<td>infeasible</td>
<td>infeasible</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>61</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>60</td>
</tr>
<tr>
<td>288</td>
<td>3</td>
<td>60</td>
</tr>
</tbody>
</table>

**TABLE VII** — Given $\delta$, resulting $M$, $S$, and computation time for MAP= 5 with 286 movements (solved in two steps).

<table>
<thead>
<tr>
<th>$\delta$</th>
<th># of modules</th>
<th># of shifts $= S$</th>
<th>computation time in sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>infeasible</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>infeasible</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>103</td>
<td>1,40</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>80</td>
<td>1,26</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>79</td>
<td>1.79</td>
</tr>
<tr>
<td>36</td>
<td>2</td>
<td>79</td>
<td>7.97</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
<td>158</td>
<td>8.42</td>
</tr>
<tr>
<td>38</td>
<td>1</td>
<td>154</td>
<td>9.34</td>
</tr>
<tr>
<td>39</td>
<td>1</td>
<td>151</td>
<td>40.84</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td>149</td>
<td>46.61</td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td>147</td>
<td>45.12</td>
</tr>
<tr>
<td>42</td>
<td>1</td>
<td>144</td>
<td>38.10</td>
</tr>
<tr>
<td>43</td>
<td>1</td>
<td>141</td>
<td>40.20</td>
</tr>
<tr>
<td>44</td>
<td>1</td>
<td>139</td>
<td>43.57</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td>137</td>
<td>9.24</td>
</tr>
<tr>
<td>46</td>
<td>1</td>
<td>136</td>
<td>106.31</td>
</tr>
<tr>
<td>47</td>
<td>1</td>
<td>135</td>
<td>148.79</td>
</tr>
<tr>
<td>48</td>
<td>1</td>
<td>134</td>
<td>100.03</td>
</tr>
<tr>
<td>49</td>
<td>1</td>
<td>133</td>
<td>94.08</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>132</td>
<td>479.12</td>
</tr>
<tr>
<td>51</td>
<td>1</td>
<td>130</td>
<td>433.79</td>
</tr>
<tr>
<td>52</td>
<td>1</td>
<td>128</td>
<td>348.83</td>
</tr>
<tr>
<td>53</td>
<td>1</td>
<td>126</td>
<td>11.65</td>
</tr>
<tr>
<td>288</td>
<td>1</td>
<td>126</td>
<td>46.49</td>
</tr>
</tbody>
</table>

**TABLE VIII** — Given $\delta$, resulting $M$, $S$, and computation time for MAP= 4 with 286 movements (solved in two steps).

<table>
<thead>
<tr>
<th>$\delta$</th>
<th># of modules</th>
<th># of shifts $= S$</th>
<th>computation time in sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>infeasible</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>infeasible</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>103</td>
<td>1.31</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>80</td>
<td>1.06</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>79</td>
<td>1.22</td>
</tr>
<tr>
<td>288</td>
<td>2</td>
<td>79</td>
<td>60.92</td>
</tr>
</tbody>
</table>
Increased traffic volume

To evaluate the behavior for the case of more traffic, we considered “2x”-traffic for October 19, 2016. That is, each of the original flight movements was duplicated and shifted randomly by plus/minus one hour, and then shifted again, randomly, by plus/minus 15 minutes. If two flight movements end up in the same slot, one of the movements is deleted. Moreover, the “2x” data was created from all data of the year 2016, that is, shifted duplicates of flights from October 18, 2016 and October 20, 2016 may now happen on October 19, 2016. Consequently, we do not end up with exactly twice the number of movements, for October 19, this data set has 416 flight movements (after deleting double movements in time slots) out of 575 flight movements (all of the movements from 2016 that the duplication and shifting process mapped to October 19, 2016). For the results see Table XI and Figure 8. For MAP= 2, we obtain the optimal number of modules of 3 for $\delta = 1$, that is, at most 5 minutes shifts, and only 33 shifts.

### TABLE IX — Given $\delta$, resulting $M$, $S$, and computation time for MAP= 3 with 286 movements (solved in two steps).

<table>
<thead>
<tr>
<th>$\delta$</th>
<th># of modules</th>
<th># of shifts = $S$</th>
<th>computation time in sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>infeasible</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>infeasible</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>103</td>
<td>1.36</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>80</td>
<td>1.28</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>79</td>
<td>1.09</td>
</tr>
<tr>
<td>288</td>
<td>2</td>
<td>79</td>
<td>51.79</td>
</tr>
</tbody>
</table>

### TABLE X — Given $\delta$, resulting $M$, $S$, and computation time for MAP= 2 with 286 movements (solved in two steps).

<table>
<thead>
<tr>
<th>$\delta$</th>
<th># of modules</th>
<th># of shifts = $S$</th>
<th>computation time in sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>infeasible</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>infeasible</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>61</td>
<td>0.55</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>61</td>
<td>1.09</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>60</td>
<td>0.98</td>
</tr>
<tr>
<td>288</td>
<td>3</td>
<td>60</td>
<td>100.30</td>
</tr>
</tbody>
</table>

### TABLE XI — For 2x traffic.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th># of modules</th>
<th>$S$</th>
<th>$\Delta$</th>
<th>$S$ for 3RTMs (1-3AP/RTM)</th>
<th>$S$ for 3RTMs (1-2AP/RTM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>30</td>
<td>5</td>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>24</td>
<td>10</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>23</td>
<td>15</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>23</td>
<td>20</td>
<td>-</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>111</td>
<td>25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>101</td>
<td>30</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>96</td>
<td>35</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>92</td>
<td>40</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>88</td>
<td>45</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>87</td>
<td>50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>84</td>
<td>55</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>81</td>
<td>60</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>81</td>
<td>65</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>81</td>
<td>70</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>81</td>
<td>75</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>80</td>
<td>80</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
In comparison, for the original ("1x") traffic, only 7 shifts were necessary. Figure 8 highlights the aforementioned tradeoffs from allowing more shifts, larger shifts [more minutes] and more APs/module [higher MAP] even more clearly.

FIGURE 8 — Number of shifts as function of the maximum shift (in minutes). Results for two modules are shown in blue, results for 3 modules with 1-3 airports per module are shown in red, results for 3 modules with 1-2 airports per module are shown in yellow.

V. Conclusion and Future Work

In this work we considered an optimization problem for remote towers that shift flights to other, nearby, slots in order to minimize the total number of modules in the remote tower center. We discussed the computational complexity of the problem, and suggested several approaches. In particular, we present an integer program and experiments for the five Swedish airports planned for remote operation. These experiments underline the applicability of our approach and show that allowing for shifts of only a few minutes can significantly reduce the number of modules needed for operation.

Our results show that cooperation between airlines, airport owners and ANSPs may help in reduction of Remote Tower Center operation costs by requiring fewer controller positions handling traffic at the airports: already minor shifts of the movement slots may significantly reduce the number of modules necessary for operation. In the original set of flight movements, even without self-induced conflicts, 5 modules are necessary, while already 32 shifts of 5 minutes lead to only 2 necessary modules. We can observe the same trend for the increased traffic ("2x" traffic): without any shifts 5 modules are necessary, 30 flight movements shifted by 5 minutes reduce this number to 3, and 111 flight movements shifted by at most 25 minutes reduce it even to 2.

Our definition of a conflict may be too conservative and too precautionary. The discussions with operational experts on this topic will continue. One question of interest is distinguishing between arrivals and departures; another is taking uncertainty into account. It is clear that the potential conflicts can not be disregarded, and will definitely be reflected in the resulting staff planning solutions. Still, in practice, some controllers may be able to handle conflicts; elaborating other metrics of workload/complexity to quantify benefits of allowing marginal conflicts is the subject of our future research.

From the theoretical perspective, the computational complexity of FRAMA with $\Delta > 0$ and even $\text{MAP} = 2$ is open.

In our current problem formulation, we do not care which airlines are affected by the reassignment of slots. That is, even if several airlines use the given airports, we might—in the worst case—reschedule flights of only a single airline. Thus, in future studies, we may take equity into account, as for example considered by Jacquillat and Vaze [7] for scheduling interventions in case of air traffic congestion without the option to
increase airport capacity. For example, if we have three airlines, with airline 1 operating 150 flights, airline 2 operating 75, and airline 3 operating 25 flights, and we need to reassign a slot for 60 flights, we might aim for 36 new slots for airline 1, 18 new slots for airline 2, and 6 new slots for airline 3. Analogously, we could distribute a total amount of shift minutes to the airlines.

VI. Acknowledgment

We thank the anonymous reviewers for the helpful comments. This research is part of the KODIC project supported by the Swedish Transport Administration (Trafikverket) and Swedish ANSP Luftfartsverket (LFV).

The question of flights rescheduling to optimize RTC operations was asked after presentation of [1] at ICRAT 2016.

REFERENCES


ASSESSING THE VIABILITY OF AN OCCUPANCY COUNT PREDICTION MODEL

Nicolas Suarez, Iciar Garcia-Ovies, Danlin Zheng
CRIDA
Madrid, Spain

Jean Boucquey
EUROCONTROL ATM/RDS/ATS
Brussels, Belgium

Abstract—This paper describes the initial validation of the approach developed by COPTRA to improve planning accuracy using an occupancy count prediction. It is based on the use of uncertainty to improve trajectory and sector occupancy estimations. The operations described in the paper take place in a Trajectory Based Operations (TBO) environment in which it is assumed an advanced Demand and Capacity Balancing (a-DCB) system is in place.

The paper focuses on the first two validation exercises performed by COPTRA to establish the current performance model baseline and to determine the improvement achieved when applying the COPTRA occupancy prediction model.

I. Introduction

To understand the importance of uncertainty in planning and thus understand how the integration of uncertainty into trajectory and sector occupancy estimation improves the efficiency of the planning process, we need first to understand the main notions behind TBO and a-DCB.

A. Trajectory Based Operations

Trajectory Based Operations (TBO) is expected to be one of the major improvements of the future ATM system. TBO will provide flexibility to airspace users and increase predictability of the ATM network. This will lead to more efficient management of ANSP’s (Air Navigation Service Providers) and airport resources while applying the required safety standards.

TBO proposes a paradigm shift from radar operations (in which the current and planned a/c position are known), to Trajectory operations (in which the current and planned a/c position are known and shared). TBO is ultimately based on the ability of the cockpit automation to fly the aircraft more precisely and predictably, resulting on a reduction of the routine controller tasks through a reduction of the associated uncertainty [1]. TBO differs from “Radar Based Operations” in which the “controller” attempts to predict the flight path of the aircraft using a cognitive process that uses a combining its general intent (flight plan) and current (and recent past) sensed position from surveillance [2].
TBO, or more specifically 4D Trajectory Management (TM) facilitates the shift from flight management through tactical intervention, towards a more strategic focus on planning and to intervention by exception [3]. The availability of precise, four-dimensional flight intent allows de-confliction of aircraft through pairwise “separation” and group deconfliction “Flow Control”.

TBO provides a strategic focus on planning and intervention [3]. It binds ATM components during tactical planning and flight operations by synchronizing the view of the trajectory between different actors. It also ensures consistency between the trajectory and/or generic constraints that originate from the various ATM components and the various regions that shape this trajectory.

The trajectory monitoring with respect to its 4D target windows tolerances, is performed on the ground, preferably through automated means. Advanced DCB (a-DCB) processes will ensure that Flow and Capacity Management operations are conducted on a holistic, seamless, continuous, and fully collaborative basis. This establishes an optimised and stable Network Operations Plan (NOP), enabling all partners concerned to fine-tune the planning of their resources per the latest known information [1].

TBO enables the effective dynamic adjustment of airspace characteristics meeting predicted demand, whilst aiming to keep any distortions to the Planned Trajectories to the absolute minimum. It also provides sufficient flexibility for optimization purposes. In a nutshell, TBO creates an environment where air and ground stakeholders share a common view of the aircraft’s trajectory enabling flight management to follow as closely as possible the Airspace User’s (AU) ideal profile, whilst optimising the flow of air traffic. TBO acts as the glue between the ATM components by synchronizing the trajectory prediction and ensuring consistency between the trajectory and/or generic constraints that originate from the various ATM components and the various regions that shape this trajectory.

The introduction of TBO requires the development of advanced ATM tools and methods to allow the effective management of individual trajectories, both in isolation and in the context of a flow. By synchronizing the trajectory, its constraints (with tolerance levels) and its generic constraints [9], ATM stakeholders increase their awareness enabling them to better anticipate on the events that may impact them.

TBO brings measurable improvements in ATC planning [1] that will lead to improvements in the prediction of sector occupancy counts. Ultimately, TBO enables the implementation of a-DCB to switch capacity management from the current global hour-based traffic limitations to minute based streamlined actions at sector level.

B. Advanced DCB

DCB is carried out through a layered planning process applied at the regional level, in close cooperation both with Sub-Regional and Local levels. It starts with the long-term planning phase, several years in advance, and finishes during the flight execution phase, through the medium and short-term planning phases. It is Airspace User oriented meaning that the new process shall offer as much as required en-route capacity so that Airspace Users can meet their business objective.

SESAR proposes to enhance DCB to manage flights after departure, filling the gap between ATFCM and ATC. In addition, the User Driven Prioritization Process is triggered in case of severe capacity drop so that Airspace Users can prioritize the flights of high marginal cost.

Advanced DCB evolves the existing DCB process and concept to a distributed network management function. This function takes full advantage from the SESAR layered collaborative planning and the Trajectory Management principles, as well as the SWIM technology to improve the effectiveness of ATM resource planning and the network
performance of the ATM system in Europe. SESAR 2020 addresses the development of a-DCB from three perspectives:

- Improving the local network intelligence\(^1\) and closing the gap between DCB and ATC
- Improving the collaborative network functions and establishing a compound network intelligence
- Improving the Shared Situation Awareness and encouraging collaborative network solutions

a-DCB measures rely on improved predictability to enable ANSPs adoption and improvement of the tactical capacity management procedures to optimise traffic throughput (with the STAM -Short Term Air traffic flow and capacity Measures). These measures are supported by automated tools for hotspot detection, and for the promulgation and implementation of STAM including CDM (Collaborative Decision Making). These tools are envisaged to be at local and regional network management function level for information sharing and CDM. a-DCB measures are built on the basis of STAM deployment (hotspot, coordination tool, occupancy traffic monitoring values (OTMV)). The enhancements foreseen focus on improved predictability of operations, including iSBT/iRBT (initial Shared Business Trajectory / initial Reference Business Trajectory) supported traffic and complexity prediction, weather, airport operations (departure sequences, ground handling, gate management, runway usage, etc.), What-if function and network view capabilities.

To achieve a-DCB some aspects of Air Traffic Management (ATM) need to be improved (amongst others):

- Prediction of traffic and workload in the 1 to 4 hour range before the flight enters the airspace region of interest.
- Operator’s awareness of performance impacts and trade-offs
- Efficiency of the daily operations? Reaction to short notice traffic situations
- Exploitation of non-standard capacity opportunities
- Integration of the use of local and network datasets

COPTRA addresses the prediction of traffic and workload in the 1 to 4 hour range, with the aim to help increase the efficiency of the daily operations.

COPTRA aims to improve the information of the estimation of demand through the modelling and understanding of the uncertainty associated to it. Thus, the contribution of COPTRA will have a direct impact on the a-DCB, enabling the achievement of the new ATM paradigm focus in a more strategic phase-based.

COPTRA provides a novel approach to the estimation, based on the development and use of a toolset based on:

- Integration of trajectory uncertainty, within the trajectory description
- Estimation of sector occupancy based on the use of enhanced trajectory description and on the computation of distribution of the probabilistic occupancy counts
- Quantification of departure time uncertainty using stochastic network models.

All three toolsets combine to produce an approach geared towards the provision of better sector occupancy predictions.

\(^1\) Network Intelligence refers to the “shared situational awareness” that will be obtained through the combination of common sets of values and rules, as well as the existence of highly interconnected local network management systems.
STAM consists in smoothing sector workloads by reducing traffic peaks through short-term application of minor ground delays, appropriate flight level capping and exiguous rerouting to a limited number of flights. These measures can reduce the traffic complexity for ATC with minimum curtailing for the airspace users. STAM is based on high-quality data for prediction and accurate traffic analysis and will be an important contribution to dynamic DCB. Improving the occupancy count estimation, will provide better application of short term air traffic flow and capacity measures to improve tactical capacity management.

The use of uncertainty in the estimation of occupancy count is a field of research that has not been fully addressed. Amongst the research performed in this area, we can cite [4], [5] who have developed a technique for analyzing airspace demand predictions based on observed data from a prototype TFM decision support system. [6]–[8] evaluate the potential impact of improved accuracy in flight timing predictions on reducing uncertainty in traffic demand predictions, hence leading to better identification of congestion. None of the work performed so far has addressed the use of uncertainty to improve occupancy estimations.

II. The COPTRA Approach

COPTRA’s approach attempts to improve planning accuracy in the tactical phase. A brief description of the approach is provided to understand the validation strategy and process (please refer to [9]–[12] for a full description).

The COPTRA approach has two steps:

► Obtaining the probability that a flight is in a sector.

► Computing the distribution of the probabilistic occupancy count from the individual probabilities of a flight being in a sector.

To compute the probability that a flight is in a sector, a set of probable trajectories and a description of temporal and spatial uncertainty for each flight is provided, as shown in Fig. 1. On it, the full probabilistic setting of a flight f between departure and destination is shown. It is characterized by many probable trajectories (here r1, r2, and r3) each one of them associated with a probability, as well as their uncertainty (not shown in the figure). Different probable trajectories can cross different sectors, e.g., r1 crosses sectors 1, 2, 4, 5. Each of these probable trajectories is defined as a three-dimensional probability density (PDF), used to compute the probability that a flight is in a sector. Although two sources of uncertainty are considered, it is assumed that only time-delay uncertainty is significant for the model.

**FIGURE 1 — Full probabilistic of a flight between departure and destination**

Airspace management 59
Then, the probabilistic occupancy count is calculated. The probabilistic occupancy count of a sector \( s \) at time \( t \) is represented as \( \Theta_{st} : \mathbb{N} \rightarrow [0, 1] \), which is a discrete PDF. For any number of flights, \( N \), the PDF \( \Theta_{st} \) will tell us the probability that \( N \) flights are in the sector \( s \) at time \( t \). Using recursive methods and dynamic programming techniques, it is possible to compute the convolutions among the probabilities to obtain the probability that \( N \) flights are in the sector \( s \) at time \( t \). The algorithm and calculation method is described in full in [11].

### III. Validation Experimental Method

COPTRA is at the TRL-1 maturity level and is expected to mature up to TRL-2 level. Based on this assumption, the validation performed within COPTRA will aim to provide an initial operational concept (including the algorithm description), the operational context of application, as well as an initial assessment of the potential performance benefits accrued by the application of the COPTRA algorithm. The result should be the formulation of the technology and an initial proof-of-concept of said technology. This should be completed with a description of the potential operational application of it.

Keeping these premises in mind, COPTRA validation strategy will be based on the use of experimental work built on the analysis of historical datasets (obtained from EUROCONTROL and CRIDA / ENAIRE)\(^2\) and the use of the COPTRA algorithm. The analysis will be based on the results obtained from five exercises as defined in the COPTRA validation plan:

- **Exercise 01**: Establish the quality of the current estimation process using an improved flight plan (imFPL) as the most probable trajectory given a flight plan.
- **Exercise 02**: Establish the initial viability of the proposed methods to estimate the occupancy in a set of sectors.
- **Exercise 03**: Determine the potential improvements brought by the COPTRA approach in terms of occupancy count prediction accuracy and uncertainty.
- **Exercise 04**: Assess the possible improvements of using occupancy count distributions in predicting hotspot and linking a probability to their occurrence.
- **Exercise 05**: Explore ways to display how uncertainty (as conveyed by occupancy count distributions) can be visualized through enhanced occupancy count graphs. These new visualizations will be presented to flow management experts and their feedback collected.

This paper focuses on the results generated in the first two exercises. For each exercise, this paper presents the dataset used to produce the results, the process followed and the validation goals.

#### A. Exercise 01

The first exercise compares occupancy counts obtained through flight plans (FPL) in different time horizons, with the occupancy counts obtained from the use of the improved flight plan, imFPL. The objective is to assess the quality of the predictions used currently to estimate the occupancy count of a sector and establish a baseline for further validation.

This exercise introduces the concept of the imFPL, which is a Flight Plan that has no uncertainty. The methodology to create the imFPL relies on two aspects: first, the original flight plan - which could be the initial or intermediate one - and second the radar tracks of the flown trajectories linked to the selected flight plan. The imFPL should

\(^2\) The information is delocalized and time and date shifted to account for potential privacy issues.
provide the most probable trajectory between a given city pair. To do so, the historical set of radar tracks associated to a FPL is analyzed to select the most probable one. This way, each FPL is associated to its most probable trajectory, which is called imFPL. In this initial validation exercise of COPTRA, only one radar track is linked to each FPL: its flown trajectory. This simplification of the methodology allows the identification of the potential benefits expected with the use of the imFPL in order to perform a further refinement on the methodology to obtain imFPL in later studies.

The applicability of the imFPL, understood as a "sort" of probabilistic trajectory, is mainly to improve the accuracy of the predicted traffic demand, with respect of using the original flight plan, since it will allow a better calculation of the entry and occupancy counts in a sector. The idea behind this is that FPLs managed by NM and ANSPs are not always as accurate as expected, since in many occasions AUs fill their FLP in a generic way that do not match reality. The introduction of imFPL leads to a better understanding of what is expectable given a FPL, and therefore NM and ANSPs will know that although the FPL states a certain trajectory, historical data concludes that this FPL always follow a different trajectory. Since the imFPL is closer to the trajectory linked to the selected flight plan, it is expected to have an hourly entry count prediction closer to the actual value. This represents the best possible outcome of today’s non probabilistic estimation process and constitutes the baseline for the COPTRA validation process.

The dataset used in this exercise was constructed from the historical databases from EUROCONTROL and CRIDA/ENAIRE. To elaborate the dataset the following information was selected for a specific timeframe:

- FPLs at three-time horizons (3 hours, 1 hour and last filed FPL in the system. In those cases where there was only one FPL, all three FPL are assumed to be the same.
- Radar tracks obtained from the controller working position recordings, used as the imFPL.

The selection of the specific validation scenario arises from the comparison of three different criteria (calculated between April 2016 to June 2016) that give raise to three different tables:

- Ranking of days with more controller issued vectors.
- Ranking of sectors with more controller issued vectors.
- Ranking of O/D (origin/destination) with more controller issued vectors.

The number of controller issued vectors is used as the critical criteria to select the most appropriate day and sectors because it is expected that the more controller issued vectors used to shorten the planned trajectory, the higher will be the deviation from the planned trajectory, and thus the higher deviation between the planned and the real occupancy count. The cross-reference and analysis of the three tables leads to the most suitable day and sectors of study considering the limitations of data acquisition [data must be available in both the EUROCONTROL and CRIDA / ENAIRE’s databases].

The result of this analysis leads to filter the dataset for those flights traversing through the Barcelona ACC in four particular sectors: LECBLVL, LECBP1L, LECBP1U and LECBPP2 that occurred on the 12 of May of 2016.

The validation process consisted on the application of three steps (occupancy is calculated using the method included in [13]):

- Calculation of the occupancy count \([\text{OCC}_{FPL}]\) using the FPLs estimated at the three-time horizons.
- Calculation of the occupancy count using imFPL \([\text{OCC}_{imFPL}]\). This is taken as the best value that could have been estimated.
Calculation of the difference between the occupancy count variables \( \text{OCC}_{\text{FPL}} \) and \( \text{OCC}_{\text{imFPL}} \). The difference between the control group \( \text{OCC}_{\text{imFPL}} \) and the dependent group \( \text{OCC}_{\text{FPL}} \) is calculated using the effect size [14].

In this exercise, occupancy counts predictions are made in a variable timeframe corresponding to the three look-ahead times specified (3 hour, 1 hour and 0 hour). For the case of the imFPL, as radar tracks are used, no prediction is made since they are the real occupancy counts.

The validation objective is twofold:
- Determine the current occupancy estimation results (in the best possible scenario) and establish the occupancy count error.
- Establish the baseline for further validation experiments.

### B. Exercise 02

The second exercise compares the occupancy counts based on the imFPL as described in exercise 01, which is the real flown trajectory linked to each FPL, with the occupancy counts obtained using the COPTRA occupancy count algorithm.

The dataset used in this exercise was constructed from the historical databases located in EUROCONTROL and CRIDA. To elaborate the dataset the following information was selected for the same timeframe and geographical location as exercise 01:
- Measured actual occupancy count (directly obtained from the operational database).
- Planned occupancy count (directly obtained from the operational database).

The validation process consisted on the application of the following three steps:
- Estimation of the occupancy count using the the COPTRA algorithm: \( \text{OCC}_{\text{probabilistic}} \).
- Calculation of the occupancy count using imFPL (as obtained in Exercise 01): \( \text{OCC}_{\text{imFPL}} \).
- Calculation the difference between the occupancy counts variables. The difference is calculated using effect sizes. In this case the control group is the \( \text{OCC}_{\text{imFPL}} \).

The validation objective is twofold:
- Improve the prediction of hotspots through the provision of probabilistic occupancy counts.
- Understand the impact of the use of probabilistic occupancy counts on the surrounding environment (contiguous sectors).

In this exercise, occupancy counts predictions obtained with the COPTRA algorithm are made in the 3 hours lookahead time.

### IV. Results and Discussion

#### A. Exercise 01 Results

As discussed in the Section III, the dataset elaborated to perform Exercise 01 consisted of the set of occupancy estimations performed on the selected day (at three-time periods), coupled with the actual occupancy measurements, in four Spanish air traffic control sectors. Since the data is based on the observed data, it is set in 20 min intervals. Fig. 2
Airspace management shows a sample of the data obtained for sector LECBP1U. This specific sector has been chosen because there was a regulation in effect between 08:00 and 10:40. As it can be observed in the figure there is in general a difference between the estimated data and the real occupancy, even in the 3-hours-before time horizon.

**FIGURE 2 — Comparison between estimated occupancy and measured occupancy. Period ranging from 08:00 to 12:40**

As it could be expected, there is an improvement on the quality of estimations as the time of the flight approaches, with more accurate prediction in the 0 hour look ahead time than in the 3 hour look ahead time. The observations performed in Fig. 2 are corroborated by the effect size calculations shown in Table I. Since the standard deviations between the control group and the test groups differ significantly, the effect size has been calculated using Glass’ $\Delta$ [14].

Table I presents the occupancy calculations at three time horizons (3 hours before flight time, 1 hour before flight and zero-hour FPL). As the FPL approaches the time of flight, the quality of the predictions improves. This is reflected by a decrease in the standard deviation (SD), the mean square error (MSE) and on the Glass’ $\Delta$. However, even in the best possible situation the error always remains significant. This is further corroborated by the value of the $\Delta$ and of the associated t-test.

**TABLE I — Effect Size data for Exercise 01.**

<table>
<thead>
<tr>
<th></th>
<th>SD</th>
<th>MSE</th>
<th>Glass’ $\Delta$</th>
<th>Cl</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>LECBLVL</td>
<td>3h</td>
<td>2,7506</td>
<td>31,0000</td>
<td>1.5690</td>
<td>[0.5672;2.5708]</td>
</tr>
<tr>
<td></td>
<td>1h</td>
<td>2,5774</td>
<td>28,2857</td>
<td>1.2258</td>
<td>[0.3050;2.1465]</td>
</tr>
<tr>
<td></td>
<td>0h</td>
<td>2,4862</td>
<td>14,0000</td>
<td>0.5393</td>
<td>[-0.2674;1.3461]</td>
</tr>
<tr>
<td>LECBP1L</td>
<td>3h</td>
<td>2,4099</td>
<td>45,4286</td>
<td>1.5018</td>
<td>[0.5169;2.4869]</td>
</tr>
<tr>
<td></td>
<td>1h</td>
<td>3,1483</td>
<td>31,3571</td>
<td>1.1979</td>
<td>[0.2831;2.1126]</td>
</tr>
<tr>
<td></td>
<td>0h</td>
<td>3,3533</td>
<td>21,1429</td>
<td>0.9297</td>
<td>[0.0671;1.7923]</td>
</tr>
<tr>
<td>LECBP1U</td>
<td>3h</td>
<td>4,4308</td>
<td>68,1429</td>
<td>1.6671</td>
<td>[0.6398;2.6943]</td>
</tr>
<tr>
<td></td>
<td>1h</td>
<td>3,6132</td>
<td>54,9286</td>
<td>1.5480</td>
<td>[0.5515;2.5445]</td>
</tr>
<tr>
<td></td>
<td>0h</td>
<td>4,4733</td>
<td>34,2857</td>
<td>1.1227</td>
<td>[0.2235;2.0218]</td>
</tr>
<tr>
<td>LECPP2</td>
<td>3h</td>
<td>1,6723</td>
<td>31,9286</td>
<td>1.8668</td>
<td>[0.7851;2.9483]</td>
</tr>
<tr>
<td></td>
<td>1h</td>
<td>3,1796</td>
<td>11,1429</td>
<td>0.6649</td>
<td>[-0.1570;1.4867]</td>
</tr>
<tr>
<td></td>
<td>0h</td>
<td>2,6520</td>
<td>6,2857</td>
<td>0.3069</td>
<td>[-0.4798;1.0936]</td>
</tr>
</tbody>
</table>

Exercise 01 shows clearly that there is room for improvement in the prediction of the occupancy count for a specific set of sectors.
B. Exercise 02 Results

As discussed in Section III, the dataset elaborated to perform Exercise 02 consisted of the set of occupancy counts calculated using the COPTRA algorithms and on the real occupancy counts.

It must be pointed out that the estimation of the sector occupancy counts performed by the COPTRA model is performed in 1 minute intervals, whilst the sector occupancy counts estimations in Exercise 01 were performed in 20 minute intervals. Even though the aggregation of the data does not have any effect on the quality of the predictions, it does have an impact on the total sector occupancy counts (which are logically lower when aggregated in 1 minute intervals instead of 20). The estimations used in Exercise 01 were obtained from the operational logs (which are presented in 20 minute intervals) and cannot be changed. As for the estimations obtained through the COPTRA performance model, they were aggregated in 1-minute intervals to increase the granularity of the observations and thus obtain a more accurate representation of the evolution of the traffic.

Fig. 3 shows the comparison between the real occupancy and the occupancy using COPTRA prediction model for the same day as Exercise 01. This figure presents all the different occupancy estimations associated to their calculated probability. It must be pointed out that for the remaining of the section, all the data shown corresponds to the most likely occupancy count prediction.

Further examination of this figure shows that predictions are better adjusted than those performed currently, even though there is room for improvement. This observation is corroborated by the data shown in Table II and Table III, in which significant improvements in the calculated Glass’s $\Delta$ as compared with those in Table I can be observed. For example, the $\Delta$ in Table I for sector LECBP1U at time zero is 1.12, whilst the $\Delta$ in Table III for the same sector and time is 0.93.

Fig. 4 presents the predicted occupancy using the most likely (highest probability) sector occupancy count for the 24 hour period and sector LECBP1U. The figure shows potential saturation periods at the period ranging from 08:00 to 10:00, 13:00 to 14:00 and 18:00 to 20:00. This information has the potential to be used for the identification of hotspots, but it needs to be further treated to be completely useful.

Fig. 5 zooms into the data shown in Fig. 4 for the period ranging from 08:00 to 12:40 (which includes the regulation that was actually implemented on the test day). Observation of this figure indicates that the use of uncertainty produces a “smoothing” effect that reduces the estimated peaks. In the present time, occupancy is calculated based on an all-ornothing event – that is, the aircraft is in the sector or not- and therefore aircraft are fully counted at each sampling time. However, when using COPTRA algorithm and introducing uncertainty, the event of an aircraft being in a sector is spread in time and in fractional part, since there is a probability associated to that aircraft being in the sector at one time or another. This induces a “smoothing” effect that erodes peaks in occupancy count prediction. Given the specifications of the COPTRA algorithm this “smoothing” effect is proportional to the standard deviation of the entry and exit times. The reduction of this effect will be explored in later validation exercises.
Table II presents the effect size calculated for the dataset. We can observe an increase in the quality of the predictions as compared with the results of Exercise 01. This increase implies that the performance of the COPTRA prediction model is established.
However, the use of the COPTRA prediction model has a smaller dispersion and provides more consistent results than the current model.

**TABLE II — Effect Size data for Exercise 02.**

<table>
<thead>
<tr>
<th>Sector</th>
<th>SD</th>
<th>MSE</th>
<th>Glass’ Δ</th>
<th>CI</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>LECBLVL</td>
<td>1.3842</td>
<td>2.5104</td>
<td>0.6456</td>
<td>[0.5688;0.7224]</td>
<td>22.3809</td>
</tr>
<tr>
<td>LECBP1L</td>
<td>1.8319</td>
<td>2.1697</td>
<td>0.5061</td>
<td>[0.4307;0.5815]</td>
<td>17.6969</td>
</tr>
<tr>
<td>LECBP1U</td>
<td>2.3142</td>
<td>4.3931</td>
<td>0.5191</td>
<td>[0.4437;0.5946]</td>
<td>13.3277</td>
</tr>
<tr>
<td>LECBPP2</td>
<td>2.4153</td>
<td>5.8417</td>
<td>0.4630</td>
<td>[0.3880;0.5380]</td>
<td>14.8377</td>
</tr>
</tbody>
</table>

To further gain insight on the performance of the COPTRA prediction model, Table III shows the effect size calculations for the same dataset but for the period ranging from 08:00 – 10:40 (the time at which a regulation was active in LECBP1U). Table III shows a significant improvement of the occupancy estimation for sector LECBP1U performed with the results derived from the COPTRA algorithm.

**TABLE III — Effect Size data for Exercise 02. Period ranging from 8:00 to 12:40**

<table>
<thead>
<tr>
<th>Sector</th>
<th>SD</th>
<th>MSE</th>
<th>Glass’ Δ</th>
<th>CI</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>LECBLVL</td>
<td>1.4315</td>
<td>4.4413</td>
<td>1.0952</td>
<td>[0.9042;1.2842]</td>
<td>17.6434</td>
</tr>
<tr>
<td>LECBP1L</td>
<td>2.0090</td>
<td>5.6655</td>
<td>0.8744</td>
<td>[0.6935;1.0533]</td>
<td>13.9563</td>
</tr>
<tr>
<td>LECBP1U</td>
<td>2.5778</td>
<td>10.9181</td>
<td>0.9398</td>
<td>[0.7566;1.1229]</td>
<td>15.1275</td>
</tr>
<tr>
<td>LECBPP2</td>
<td>2.1673</td>
<td>13.0107</td>
<td>1.4133</td>
<td>[1.2102;1.6163]</td>
<td>22.6296</td>
</tr>
</tbody>
</table>

Altogether, the results obtained in Exercise 02 show that the COPTRA prediction model is more capable to estimate the occupancy than the current situation.

**C. Limitations of the results**

It must be pointed out that results are based only on archived data. This implies that the predictions are always made over data which might have been impacted by a previous traffic flow capacity measure. Full validation of the results would require the use and test of the algorithms on a real-time data.

Additionally, the dataset was constructed using a partial view of the European Air Traffic Management network limited to Spain. Given the nature of a-DCB, the study should be enlarged to include the complete ECAC area.

The identification of hotpots currently relies strongly on the value of the pre-established sector capacity. This fixed value has less meaning within a context in which an aircraft can be estimated to be in different sectors simultaneously with different probabilities due to the use of uncertainty. To account for this effect, a-DCB should propose the identification of hotpots using complexity calculations based on the occupancy sector probabilities.

Lastly, the validation performed in Exercise 01 and Exercise 02 is strictly based on mathematics. Its objective was to establish the viability of the proposed COPTRA algorithms. To complete the validation process, the validation exercise must focus on the application of the algorithm to an operational environment.

**D. Summary of the findings**

Exercise 01 has established the limitations of present day estimations. As seen in the analysis of the data, the performance of the current prediction model is strongly dependent on the quality of the available FPL. Furthermore, even if the FPL was optimal (in the sense that it would present little or no difference with the flight trajectory), the predictions are of limited quality. This situation is corroborated from the observation of current a-DCB operations.
Exercise 02 has proven the viability of the COPTRA algorithms. The performance of the prediction model is significantly improved over the baseline established in Exercise 01. When the prediction is focused on a specific time range, the prediction model performs significantly better than the baseline.

E. Practical implications of the results

Having more information related to the estimation of occupancy for a given sector and time will lead to fewer false positives (hotspots declared and not occurring) and negatives (hotspots not declared and occurring). This will occur once the performance of the prediction models is improved from today’s standards. COPTRA has already established the theoretical ground to proceed along this path. The initial results (as shown in this report and on the ongoing validation efforts) are promising and show an improvement from present conditions.

V. Conclusions

Based on the probabilistic model and algorithms presented in this paper and in [9]–[12] the paper has described the operational context of the use of uncertainty in a trajectory based operations environment, to address the advanced Demand and capacity balancing. The paper also described the validation approach that COPTRA is applying to ensure that the model and algorithms have operational use.

The paper focuses on the initial validation steps taken by COPTRA to establish a baseline and the technical viability of the algorithms used within an operational environment.

The results obtained in Exercises 01 and Exercise 02 show a clear improvement of the occupancy prediction model proposed by COPTRA vs. current operations. These results will be completed in further ongoing validation exercises that will establish the operational use and validity of the results.

The work presented here has opened further research questions amongst which we can highlight the need to redefine our understanding of sector capacity within an environment in which uncertainty is taken into account. Attention must also be paid to the smoothing effect observed when making predictions based on uncertainty.

REFERENCES

I. Introduction

Air Traffic Management incurs complex operations involving numerous actors and processes carrying many unknowns and uncertainties. Today, these uncertainties (e.g. exact takeoff time, route changes ...) are only taken into account in the system in very limited ways and expert experience and judgement are relied upon to cater for them, often leading to bigger margins or conservative capacity estimates.

One way to make uncertainty explicit is to use probability distributions. Early 2016, the SESAR 2020 Exploratory Research project COPTRA has been started. COPTRA’s main objective is to research ways to explicitly account for uncertainty in trajectory and traffic predictions using probabilistic trajectories and traffic situations.

Predicting occupancy counts is central to ATC planning and Demand-Capacity Balancing: the predicted values are used to chose the right airspace sectorisation or decide on necessary regulations. Today, however, the uncertainties on the inputs of counting process (like the take-off time) make the count predictions highly volatile. Following COPTRA’s objective to make these uncertainties explicit and manageable, this paper describes an approach to attach uncertainty figures to sector entry and crossing times using historical data. Probabilistic sector sequences built from these figures are then used to compute occupancy count distributions using the algorithm detailed in [1].

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 699274. This document is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 699274.
In Section II, traffic uncertainty is further discussed and detailed. Section III describes how COPTRA proposes to model probabilistic trajectories and to combine them to compute occupancy count distributions. The way this trajectory model is used to represent traffic demand uncertainty is presented in Section IV. In Section V, the approach is applied to MUAC’s EDYYB5KL sector. Results and on-going work are presented in Section VI.

II. Traffic uncertainty

Many sources of uncertainty exist in ATM leading to nonoptimal preventive actions (increased margins or buffers). In [2], Irvine details the Capacity buffer theory which states that sector capacity is set to control the probability of occupancy counts exceeding the peak acceptable level. The theory establishes a direct link between count uncertainty and sector capacity. This stresses for the need to better manage uncertainty and to find ways to make uncertainty explicit.

One major source of uncertainty affects the actual take-off (or more precisely, in the frame of this study, the off-block) time of the flight. Off-block and take-off can be delayed for numerous reasons. Delay data is extensively collected and documented (see e.g. EUROCONTROL’s Central Office for Delay Analysis [3]).

When predicting sector occupancy counts, uncertainty also comes from the differences between the flight planning information (used to predict sector occupancy) and the way the flights are eventually flown. Figure 1 and 2 show, for a full day of the MUAC Brussels airspace, respectively the planned traffic and the actual traffic. Sources of these differences include the actions taken by ATC to deconflict traffic, weather or turbulence avoidance, emergency or flight diversions...

**FIGURE 1 — Filed traffic (13/10/2016 — MUAC Brussels airspace)**
III. Combining Probable Trajectories

A. Probabilistic Trajectory Model

In order to cope with the uncertainties just outlined, COPTRA developed a probabilistic trajectory model able to take into account:

- Flight plan or route uncertainty: filed flight plans have to follow rules (like route network and route opening schemes) which are potentially relaxed or changed during execution (e.g. “[Fly] direct” instructions given by controllers).

- Flight execution uncertainty: even with the flight route fixed, many inputs to the trajectory prediction process are not fully known or remain uncertain (weather, aircraft performance...) and lead to uncertainty on the flown trajectory.

A flight is then represented by a series of routes each having a given probability. A probable route is made of a series of points whose attribute values (longitude, latitude, altitude, time...) are expressed as probability distributions.

For the problem at hand, i.e. computing occupancy count distributions, the probable routes have been simplified to probable sector sequences: The probable route becomes a list of sectors crossed by the flights along with a probability distribution of the entry and exit times. In this model, the entry and exit time distributions are approximated by Gaussian distributions.

A probabilistic trajectory $T_f$ for flight $f$ can then be described formally as follows:

$$ T_f = \left\{ \rho_{(i,j)}^f \left[ \epsilon_{(i,j)}, \mu_{(i,j)}^e, \sigma_{(i,j)}^e, \mu_{(i,j)}^l, \sigma_{(i,j)}^l \right] \right\} $$

with $i$ ranging from 1 to $n$, for the different sector sequences and $j$ ranging from 1 to $m_{(i)}$, on the sectors making the $i^{th}$ sequence. $f$ identifies the flight, $\epsilon_{(i,j)}$ is the identifier of the crossed sector, $\mu_{(i,j)}^e$ and $\sigma_{(i,j)}^e$ are the parameters of the entry time distribution while $\mu_{(i,j)}^l$ and $\sigma_{(i,j)}^l$ parametrize the exit time distribution.

B. Computing occupancy count distributions

In [1], Gonze et al. describe a polynomial time algorithm to compute occupancy count distributions from probabilistic sector sequences following the model just described. An outline of the algorithm is given here. The reader is referred to [1] for details.
Given a set of probabilistic trajectories, the algorithm proceeds in two steps:

- For each flight trajectory, sampling time and sector, the probability \( p_{(f,s,i)} \) of the flight being in the sector at the given time is computed.
- From these probabilities, the occupancy count distributions \( \Theta_{(s,t)} \) are computed for each sector and sampling time.

These two steps are now detailed:

1. **“In sector” probability:** The probability that flight \( f \) is in sector \( s \) at time \( t \), when following trajectory \( i \) is

   \[
   p_{(f,s,i)} = P_{s}^{t_{e}^{(f,s,i)}} \leq t - P_{s}^{t_{l}^{(f,s,i)}} \leq t
   \]

   where \( t_{e}^{(f,s,i)} \) is the flight entry time in sector \( s \) when following the \( i \)th trajectory. Similarly \( t_{l}^{(f,s,i)} \) is the exit time. In other words \( p_{(f,s,i)} \) is the probability that, at \( t \), \( f \) has entered the sector but not left it yet.

   \[
   p_{(f,s,i)} = \sum_{i=1}^{n_{f}} \frac{p_{(f,s,i)}^{i} p_{(f,s,i)}^{i-1}}{p_{(f,s,i)}^{i} p_{(f,s,i)}^{i-1}}
   \]

2. **Occupancy count distributions:** The occupancy count distribution,

   \[
   \Theta_{(s,t)} : \mathbb{N} \rightarrow [0,1] : k \rightarrow \Theta_{(s,t)}(k)
   \]

   gives the probability of having \( n \in \mathbb{N} \) flights in sector \( s \) at time \( t \). It is the convolution of binomial distributions giving the probability of having flight \( f \) in sector \( s \) at time \( t \) with probability \( p_{(f,s,i)} \). By standard methods computing \( \Theta_{(s,t)} \) has an exponential computational cost.

   However, by ordering the flights \( f_{j} \) with \( j \in 1...m \) and defining \( q_{(i,j)} \) as the probability that, amongst the \( j \) first flights, there are \( i \) flights in sector \( s \) at time \( t \), we have

   \[
   \Theta_{(s,t)}(k) = q_{(i,j)}
   \]

   \( q_{(i,j)} \) can then be defined using the following recurrence relation (for fixed \( s \) and \( t \))

   \[
   q_{(i,j)} = q_{(i-1,j-1)}(1 - p_{(f,s,i)}) + q_{(i-1,j-1)} p_{(f,s,i)}
   \]

   which says that the probability of having \( i \) flights amongst the \( j \) first ones is the sum of

   - The probability of having \( i \) flights in the sector amongst the \( j - 1 \) first flights and not having flight \( j \) in it, and,
   - The probability of having \( i - 1 \) flights in the sector and having flight \( j \) in it.

   Using this recurrence relation and applying dynamic programming techniques, the \( \Theta_{(s,t)} \) distribution is computed in polynomial time.

### C. Required inputs

To feed the algorithm computing the \( \Theta_{(s,t)} \) occupancy count distributions, it is required to determine for each subject flight \( f \):
The list of probable sector sequences and their respective probability: \( n \) and \( \rho_{(v)} \) for each sector sequence, the list of \( m_{(v)} \) sectors crossed \( \{ es_{(v)} \} \) and the entry and exit time distributions. As these distributions are assumed Gaussian, they are fully specified by their means and standard deviations: \( \mu_{(v)} \), \( \sigma_{(v)} \), \( \mu_{(v)}' \), \( \sigma_{(v)}' \).

Hereafter an approach is described that uses historical data to derive these distribution parameters.

Other ways to compute these inputs are possible and the COPTRA project also explores how to get them using probabilistic trajectory prediction.

IV. Probabilistic Traffic Model

The overall objective of the approach that will now be described is to compute the uncertainty attached to a given flight plan by deriving a series of probable sector sequences and the corresponding entry and exit time distributions. To that effect a probabilistic traffic model is first built from historical data. The model is then used to compute probabilistic sector sequences that are fed into the algorithm described in subsection III-B to get the occupancy count distributions \( \Theta_{(v,d)} \).

The full traffic model is composed of sector sequence, entry time and crossing time distributions. In this paper, we present models for entry time and crossing time distributions. Sector sequence modelling is currently work in progress.

A. Historical data

\textsc{Eurocontrol} maintains a repository, the Demand Data Repository or DDR2 [4], accessible to the ATM research community. Amongst others, the repository stores, in different formats, all the flights for the last five years.

The model described hereafter was built from the traffic crossing \textsc{Eurocontrol}’s MUAC Brussels airspace during three consecutive AIRAC cycles (1607, 1608 and 1609) as stored in so-called AllFT+ formatted files. However, as the airspace of interest and the input data are parameters external to the model building process, models could be built from any historical dataset for any airspace of interest.

The input dataset contains 244108 unique flights crossing MUAC’s Brussels airspace (EDYYB). For each flight in this dataset, the following features are extracted or computed:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Departure airport</td>
<td>ADEP</td>
</tr>
<tr>
<td>Destination airport</td>
<td>ADES</td>
</tr>
<tr>
<td>Callsign</td>
<td>CS</td>
</tr>
<tr>
<td>IFPL Identifier</td>
<td>IFPLID</td>
</tr>
<tr>
<td>Week of the year</td>
<td>WOY</td>
</tr>
<tr>
<td>Day of the week</td>
<td>DOW</td>
</tr>
<tr>
<td>Hour of the day</td>
<td>HOD</td>
</tr>
<tr>
<td>Entering EDYYB from</td>
<td>FROM</td>
</tr>
<tr>
<td>Leaving EDYYB to</td>
<td>TO</td>
</tr>
<tr>
<td>Sector</td>
<td>ES</td>
</tr>
<tr>
<td>Delta off-block time</td>
<td>DOBT</td>
</tr>
<tr>
<td>Delta entry time</td>
<td>DETI</td>
</tr>
<tr>
<td>Crossing time</td>
<td>XGTI</td>
</tr>
</tbody>
</table>

\* Significant revisions in Aeronautical Information Publication are made every 28 days. This period is called AIRAC (Aeronautical Information Regulation and Control) cycle.
This feature set is designed to capture most relevant traffic characteristics and to account for both temporal (yearly, weekly or daily) and spatial (routes or flows) trends. While the probabilistic trajectory model uses “absolute” entry and exit times, when building the traffic models, the sector entry time is computed as the delta from the actual off-block time. Similarly the time difference between the flight time of entry in the sector and the time of exit form the sector (sector crossing time) is used instead of the exit time. These two features allow to build models which are independent of the actual time at which the flight took place while permitting, for a subject flight, to compute entry and exit times distributions from the predicted or actual off-block time.

From the processed dataset, empirical entry time and crossing time conditional distributions can be derived for each sector. Figure 3 and Figure 4 show respectively typical entry time and crossing time distributions for a specific sector. The entry time distribution is conditioned on the departure airport. The crossing time distribution is unconditioned: it shows the crossing time frequencies for all the flights in the data sample.

**FIGURE 3 — Entry time distribution for traffic from EHAM (MUAC sector EDYYB37H)**

The distributions illustrated are not Gaussian, as it is the case for to the vast majority. Modelling entry times or crossing times with Gaussian distributions would lead to an important loss of information: For instance, the multi-modality of the crossing time distribution probably accounts for the different routes crossing the elementary sector. Conditioning the distributions (e.g. on the route) might be a way to “separate” these different distributions. However, it might quickly lead to very specific models. This is the classical “bias-variance” trade-off (see e.g. [5]). Moreover, as it will be seen, adding conditions also increases the size of the model exponentially.

It was then decided to look for models that could represent any of these distributions (conditioned and unconditioned) with sufficient accuracy. This would allow to explore...
models with a varying number of conditional parameters and start with simple models that could be enriched if necessary.

The use of Gaussian Mixture Models (GMM) is described hereafter. An additional and important advantage of GMM is that they can be used with the probabilistic trajectory model described in subsection III-A.

B. Gaussian Mixture Models

A Gaussian mixture model is a probabilistic model that assumes that all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters ([5], [6]).

\[
\text{GMM} = \sum_{i=1}^{n} w_i \mathcal{N}(\mu_i, \sigma_i)
\]

with the corresponding probability density function:

\[
\rho_{\text{GMM}}(x) = \sum_{i=1}^{n} w_i \mathcal{N}(x | \mu_i, \sigma_i)
\]

where \( w_i \) is the weight of the \( i \)th Gaussian of the mixture with \( \sum_{i=1}^{n} w_i = 1 \). Similarly \( \mu_i \) and \( \sigma_i \) are, respectively, the mean and the standard deviation of the \( i \)th Gaussian. The resulting distribution would “generate” elements distributed along the Gaussian \( \mathcal{N}(\mu_i, \sigma_i) \) with a probability of \( w_i \). In its general setting, GMM supports multivariate Gaussian distributions. Univariate distributions are sufficient here.

The process of fitting a GMM to given data is an unsupervised learning problem. To fit the GMM to the data, the unknown parameters \( n \), \( w_i \), \( \mu_i \) and \( \sigma_i \) for \( i \) ranging from 1 to \( n \) will be determined from the historical dataset.

For a given \( n \), these parameters can be determined using expectation-maximization or Bayesian techniques. Bayesian Information Criterion (BIC) can help to select the value of \( n \) [7].

In the frame of this work, expectation-maximization as implemented in the Python Scikit-learn toolbox [6] was used to fit the GMM. As BIC led to high values for \( n \), visual inspection was used to determine the quality of the fit while favouring models having a small number (2 or 3) of components. Figures 5 and 6 show the GMM fitted to the empirical distribution of figures 3 and 4 respectively.

**FIGURE 5 — EDYYB37EH entry times from EHAM fitted with a 3 components GMM**
Both the entry times from Amsterdam and the crossing times were fitted by expectation-maximization with 3 components GMM. The parameters for the delta entry time GMM are:

<table>
<thead>
<tr>
<th>$i$</th>
<th>$w_i$</th>
<th>$\mu_i$ (s)</th>
<th>$\sigma_i$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.68</td>
<td>1534.08</td>
<td>46.64</td>
</tr>
<tr>
<td>2</td>
<td>0.15</td>
<td>1844.98</td>
<td>202.37</td>
</tr>
<tr>
<td>3</td>
<td>0.17</td>
<td>1458.48</td>
<td>187.52</td>
</tr>
</tbody>
</table>

while the parameters for the crossing time GMM are:

<table>
<thead>
<tr>
<th>$i$</th>
<th>$w_i$</th>
<th>$\mu_i$ (s)</th>
<th>$\sigma_i$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.38</td>
<td>138.54</td>
<td>73.22</td>
</tr>
<tr>
<td>2</td>
<td>0.41</td>
<td>451.23</td>
<td>111.10</td>
</tr>
<tr>
<td>3</td>
<td>0.21</td>
<td>568.11</td>
<td>198.81</td>
</tr>
</tbody>
</table>

Once fitted, a GMM can work either as a classifier or a predictor: these two ways will be used hereafter. As a classifier, the model is given a time (planned or predicted) and returns the index of the most probable GMM component for this value:

\[
\text{GMM}_{\text{class}}(t) \to i_{\text{max}}
\]

When used as a predictor, the GMM gives, for an input value, the probabilities that it has been drawn from the different components:

\[
\text{GMM}_{\text{pred}}(t) \to \{\rho_k^{\text{GMM}}\}_{k=1}^3
\]

The next section describes the general approach to use Gaussian Mixture Models to build probabilistic traffic models.

C. General Model Structure

For each elementary sector, a series of GMM is fitted to model the entry time distributions and the crossing times. An additional model is fitted for each (major) airport to model take-off delays (difference) between the planned off-block time and the actual one.

When the models are conditioned on parameters that can take many values (e.g. the departure airport), the value set is divided in two: the set of values that concerns a majority of elements in the dataset and the rest. For example, amongst the 601 different departure airports appearing in the traffic dataset for elementary sector EDYYB37EH, 20 of them are the origin of more than 50% of the flights. For each value in this first, majority set, a separate GMM is fitted that will be used as a predictor. A single GMM is
fitted for the second set and used as classifier to determine the distribution parameters for the different values grouped by classes [e.g. ICAO region]. See Section V for the details.

A full traffic model is so defined by:

- A set of parameters: the features of the flight that will determine the GMM to use [e.g. the departure airport, the day of the week...]
- For each parameter: the list of GMM to apply. This list may contain either a single model like for the elementary sector crossing times or multiple models with different models corresponding to different values of the parameter, like for the airport of departure.

D. Compatibility with the probabilistic trajectory model

The previous subsections have shown how empirical entry and crossing time distributions can be approximated by Gaussian Mixture Models. It is described now how GMM-approximated distributions can be used in the probabilistic trajectory model described in Section III-A where the entry time and exit time distributions are approximated by Gaussian distributions.

When a GMM is used as a classifier, it returns which of the Gaussians is the most probable for a given value. On the other hand, if a GMM is used as a predictor, it returns the probabilities of its different Gaussian components. In the first case, the parameters of the selected Gaussian can be directly used in the probabilistic trajectory model. In the second case, a separate trajectory can be produced with the parameters of the different components, each with the probability of the corresponding component.

If more than one parameter are used, the joint probability table of the predicted probabilities is built [the parameters are assumed independent]. The size of the joint probability table is exponential in the number of parameters. This provides a strong justification to keep both the number of parameters and components in the models as small as possible.

Formally, let us assume a sector crossing (entry and exit times) for flight $f$ modelled by $q$ parameters. From the full model, the mixture models $\mathcal{M}_k$ for $k$ ranging from 1 to $q$ have been selected for the flight given the values of the different parameters. This set of GMM is made of two subsets: $q_1$ models $\mathcal{M}_e$ affecting the delta entry time and $q_2$ models $\mathcal{M}_{cl}$ affecting the crossing time. Each $\mathcal{M}_k$ has $n_k$ components of weights $w_k(l)$ [with $l$ from 1 to $n_k$]. A GMM used as a classifier has a single component of weight 1.

The sector crossing will be represented by $n = \prod_{k=1}^{q} n_k$ probabilistic sector sequences and the set of possible GMM component combinations is the cross product of the integer sequences from 1 to $n_k$:

$$I_f = \times_{k=1}^{q} \{1...n_k\}$$

$I_f$ is assumed to be ordered in some way [e.g. lexicographical order] so that $I_{(i)} = (l_{(i,1)},..., l_{(i,q)})$ selects the $i^{th}$ combination of GMM components. Then the probability of each of the $n$ sector sequences is

$$p_{(i)} = \prod_{k=1}^{q} w_k(l_{(i,k)})$$
and the entry and exit time distribution parameters are

\[
\begin{align*}
\mu^j_{(i,j)} &= \sum_{k=1}^{q_1} \mu_k \left( M_k^{(i,j)} \right) \\
\sigma^2_{(i,j)} &= \sqrt{\sum_{k=1}^{q_1} \sigma^2_k \left( M_k^{(i,j)} \right)} \\
\mu^j_{(i,j)} &= \mu^j_{(i,j)} + \sum_{k=1}^{q_2} \mu_k \left( M_k^{(i,j)} \right) \\
\sigma^2_{(i,j)} &= \sqrt{\left( \sigma^2_{(i,j)} \right)^2 + \sum_{k=1}^{q_2} \sigma^2_k \left( M_k^{(i,j)} \right)}
\end{align*}
\]

where \( j \) is the index of the elementary sector crossing in the \( i \)th probabilistic trajectory.

V. Application

To fix the ideas, practical details on how entry and exit times of a single elementary sector can be modelled using the framework just described are given now. The model uses the airport of departure and the flight state (airborne or not) as mixture selection parameters. It is built for MUAC elementary sector EDYYB5KL. The historical dataset is built from the traffic of 3 consecutive AIRAC cycles (1607, 1608 and 1609). It contains 91389 crossings of EDYYB5KL. The model is used to derive occupancy count distributions for traffic predicted on the 5th of May, 2017.

A. Model building

The model describes off-block delays, delta entry times and crossing times by three series of GMM. Note that the off-block delays are airport specific, the delta entry times depend on both the sector of interest and the departure airport while crossing times depend on the elementary sector only.

1. Off-block delay and delta entry time models

Off-block delay and delta entry time models use the departure airport as parameter. Both model types are built the same way.

Traffic flying through a given sector potentially originates from a wide number of airports (552 in this case). Fitting separate GMM for each of these airports would be a large undertaking. To limit the effort, only the 11 most frequent airports (representing 59% of the traffic) have their off-block delay and delta entry time empirical distributions modelled by a GMM. A residual model is built for the remaining 541 airports.

The following table lists the 11 airports representing close to 60% of the traffic and the number of components of the fitted GMM:

<table>
<thead>
<tr>
<th>Airport</th>
<th>%</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGGL</td>
<td>17.00</td>
<td>1</td>
</tr>
<tr>
<td>EGKK</td>
<td>12.19</td>
<td>3</td>
</tr>
<tr>
<td>EGSS</td>
<td>8.71</td>
<td>2</td>
</tr>
<tr>
<td>EGGW</td>
<td>4.50</td>
<td>2</td>
</tr>
<tr>
<td>LFPG</td>
<td>4.48</td>
<td>2</td>
</tr>
</tbody>
</table>
To build the *residual* model, a single GMM is fitted to the off-block delay or delta entry time distributions for all the remaining departure airports. In this case, the fitted GMM has 3 components. The remaining airports are then grouped by ICAO region (i.e. the two first letters of their ICAO code). The average off-block delay or delta entry time is then computed per region. This average is then input to the GMM used as a classifier to determine for each region the most probable component. The parameters [mean and standard deviation] of the selected component are used as the parameters of the offblock or delta entry time distribution for any flight originating from the region.

2. **Crossing time model**

A single 3 components GMM is fitted to model the sector crossing times. It has the following parameters:

<table>
<thead>
<tr>
<th>i</th>
<th>w_i</th>
<th>( \mu_i ) (s)</th>
<th>( \sigma_i ) (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.37</td>
<td>115.04</td>
<td>77.38</td>
</tr>
<tr>
<td>2</td>
<td>.31</td>
<td>330.67</td>
<td>159.85</td>
</tr>
<tr>
<td>3</td>
<td>.32</td>
<td>474.25</td>
<td>90.54</td>
</tr>
</tbody>
</table>

3. **The complete set of models**

In total the traffic model for sector EDYYB5KL is composed of 25 GMM fitted to historical data:

- 12 GMM for the off-block delay models: 11 predictor GMM for the most frequent airports and one additional classifier GMM used for all the remaining airports.

- 12 GMM for the delta entry time models: 11 predictor GMM for the most frequent airports and one additional classifier GMM used for the remaining airports.

- 1 predictor GMM to model sector crossing times.

B. **Model use**

The probabilistic traffic model is applied to the list of flights predicted for given target and look-ahead times. The list of flights known at a given time as well as their predicted trajectories is extracted from the Enhanced Traffic Flow Management System (ETFMS) logs for the day of interest.

The model selection parameters are extracted for each flight (in case of the model example given above the airport of departure and the state — airborne or not — of the flight). These parameters are used to select the models (set of GMM) that will be applied to compute the distribution parameters. If the off-block time is predicted (the flight is not yet airborne), the off-block delay model is used to compute the parameters of entry time distribution.

The currently predicted delta entry and crossing times of the flight are then input to the selected models and the returned probabilities are used to compute the parameters of the entry and exit time distributions for all the combinations of mixture components. So, for each flight, the application of the model results in a set of probabilistic trajectories.
The sets of probabilistic trajectories computed for all flights in the list are fed in the algorithm described in the section III-B to compute the occupancy count distributions $\Theta_{s,t}^\ell$.

VI. Results

Figure 7 and Figure 8 show the occupancy count distributions, actual and predicted counts around respectively 11:00 and 14:00 on the 5th of May 2017 for MUAC elementary sector EDYYB5KL using the approach described in this paper. The predictions were done 30 minutes and 1 hour in advance. The red curve represents the means of the occupancy count distributions computed every minutes. The (blue and red) dashed lines give the 90% interval computed from the occupancy count distributions (respectively the 5% and 95% quantiles). The blue curve gives the sector instantaneous occupancy counts as computed from the actual flight profiles extracted from the AllFT+ archive of the day. The grey curve is the occupancy counts as computed from the flights known at the time of prediction (baseline).

**FIGURE 7 — Occupancy count distributions (red and dashed) along actual (blue) and baseline (grey) occupancies as predicted at 10:30 for 11:00**

![Figure 7](image)

**FIGURE 8 — Occupancy count distributions (red and dashed) along actual (blue) and baseline (grey) occupancies as predicted at 13:00 for 14:00**

![Figure 8](image)

To further assess the model, probabilistic predictions (Probabilistic counts) and current occupancy count predictions (Baseline counts) were compared to the occupancy counts computed from the final profiles available in the corresponding AllFt+ archive (Actuals counts). The comparison was done for 37 target times falling every half an hour from 05:00 to 23:00 on the 5th of May, 2017. For each target time $t$, the predictions were compared for 11 look-ahead times (l) ranging, every half an hour, from $t - 5h$ to $t$. 
The comparisons were done using the Ranked Probability Score (RPS) ([8], [9]).

If \( F_{n}^{\theta}(n) = \sum_{i=0}^{n} \Theta_{i}(l) \) is the cumulative distribution function of the occupancy count distribution \( \Theta_{i} \) for sector \( s \) at time \( t \) predicted with look-ahead time \( l \):

\[
F_{n}^{\theta}(n) = \sum_{i=0}^{n} \Theta_{i}(l)
\]

and \( H[n] \) is the discrete Heaviside step function:

\[
H[n] = \begin{cases} 0, & n < 0, \\ 1, & n \geq 0 \end{cases}
\]

then, for the probabilistic count prediction \( \Theta_{i} \) and the actual count \( o \), the RPS is computed as follows:

\[
\text{RPS}(\theta^{o}(n), \theta(n)) = \sum_{n=0}^{\infty} \left( F_{n}^{\theta}(n) - H[n - o] \right)^{2}
\]

The RPS has the useful property of being able to handle both probabilistic predictions (the probabilistic counts) and deterministic predictions (the baseline counts). Deterministic predictions are considered as distributions with single value of probability 1. In this case the RPS is equal to the absolute error.

Figure 9 and Table I show the means and standard deviations of the scores computed between the actual counts and, respectively, the baseline counts (red) and the probabilistic counts (blue) for the different look-ahead times (lower scores mean better results).

**FIGURE 9 — RPS means and standard deviations for baseline (red) and probabilistic (blue) counts**

![Figure 9](image)

**TABLE I — Means and standard deviations of the baseline and probabilistic scores**

<table>
<thead>
<tr>
<th>( l ) (h)</th>
<th>Baseline mean</th>
<th>Baseline stdev</th>
<th>Probabilistic mean</th>
<th>Probabilistic stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.0</td>
<td>2.20</td>
<td>1.155</td>
<td>0.89</td>
<td>0.507</td>
</tr>
<tr>
<td>-4.5</td>
<td>2.12</td>
<td>1.177</td>
<td>0.97</td>
<td>0.579</td>
</tr>
<tr>
<td>-4.0</td>
<td>2.41</td>
<td>1.500</td>
<td>1.11</td>
<td>0.701</td>
</tr>
<tr>
<td>-3.5</td>
<td>2.00</td>
<td>1.963</td>
<td>1.18</td>
<td>0.804</td>
</tr>
<tr>
<td>-3.0</td>
<td>1.63</td>
<td>1.606</td>
<td>1.19</td>
<td>0.914</td>
</tr>
<tr>
<td>-2.5</td>
<td>1.93</td>
<td>1.574</td>
<td>1.15</td>
<td>0.821</td>
</tr>
<tr>
<td>-2.0</td>
<td>2.13</td>
<td>1.586</td>
<td>1.26</td>
<td>1.100</td>
</tr>
<tr>
<td>-1.5</td>
<td>2.31</td>
<td>1.306</td>
<td>1.08</td>
<td>0.878</td>
</tr>
<tr>
<td>-1.0</td>
<td>2.39</td>
<td>2.106</td>
<td>1.07</td>
<td>0.789</td>
</tr>
<tr>
<td>-0.5</td>
<td>2.59</td>
<td>1.987</td>
<td>1.15</td>
<td>0.861</td>
</tr>
<tr>
<td>0.0</td>
<td>1.57</td>
<td>1.290</td>
<td>1.08</td>
<td>0.910</td>
</tr>
</tbody>
</table>
Statistical tests applied to both the means and standard deviations show, at a 5% significance level, that:

- The baseline and probabilistic score standard deviations are significantly different for all the look-ahead times. The probabilistic prediction scores being less spread, there is a reduction in uncertainty on the count predictions which, according to the Capacity Buffer theory [2], would lead to a capacity increase.

- The baseline and probabilistic means are significantly different for all look-ahead times except for predictions made at time $t$ (0 hour look-ahead time) and at $t - 3h$: In the majority of the cases, the probabilistic count predictions are more accurate.

The stability over look-ahead time of the probabilistic prediction score has also to be noted as it would mean that probabilistic prediction provides more accurate and stable count predictions earlier in time.

### VII. Conclusion and further research

Based on the probabilistic trajectory model and the algorithm presented in [1], the paper described a flexible and extensible approach based on historical data to attach uncertainty to traffic demand: The empirical take-off delay, delta entry time and crossing time distributions conditioned on selected parameters are modelled by Gaussian Mixture Models (GMM). The approach allows to model the multimodal entry and crossing time distributions observed in ATM. When applied, the GMM-approximated distributions can be however represented as a combination of probable trajectories each having Gaussian entry and exit time distributions which are more tractable. These sets of trajectory combinations are input to a polynomial time algorithm computing occupancy count distributions.

The flexibility of the approach allows to select the set of most relevant model parameters and while being exponential in the number of parameters, it is shown that simple models (3 parameters modelled by GMM with up to 3 components) are already able to provide sensible results: Current results show not only a significant reduction in the spread of the prediction scores but also a general improvement of the quality of the predictions made using the probabilistic models.

The approach, limited here to the crossing of a single elementary sector, is being combined with probabilistic sector sequences necessary to model the uncertainty on the actual route flown by the aircraft.

Detailed performance analysis is on going. At this stage however, all computations related to the usage of the model always remained below one minute on standard desktop computers.

The work presented here opens several further research questions, amongst which:

- How to determine the most significant inputs or parameters to be taken into account when building traffic models?

- Is modelling time (or other) distributions in ATM using Gaussian mixture applicable/useful to other purposes than occupancy count distributions as described here?

- The use of Bayesian modelling should be explored when fitting the mixture models to the historical data. It would allow to inject some prior knowledge while determining the mixture parameters and avoid the tendency of BIC-based model determination to lead to models with a high number of components.
REFERENCES


[7] H.S. Bhat, N. Kumar, On the derivation of the Bayesian Information Criterion


TOWARDS SIMPLIFIED OPTIMAL SECTOR SPLITTING

Leonid Sedov and Valentin Polishchuk
Communications and Transport Systems ITN, Linköping University, Sweden
firstname.lastname@liu.se

Billy Josefsson
LFV
firstname.lastname@lfv.se

Abstract—Merging and splitting control sectors is a certified way to address demand-capacity imbalances in an airspace: during high-traffic hours a sector is split into two or more smaller sectors, while in low-traffic hours the parts are merged back. In this paper we explore ways of splitting a sector with simple straightline cut so as to balance, as well as possible, the number of controlled aircraft in the obtained smaller sectors. The approach is verified by computational experiments with enroute traffic over Swedish airspace.

I. Introduction and Related Work

Demand-capacity imbalances are a major challenge in airspace management, and initiatives like dCDB (dynamic Demand Capacity Balancing) within SESAR and DAC (Dynamic Airspace Configuration) in NextGen are specifically addressing it. The issue is that the traffic density is changing over time, so static control sectors may become under- and overloaded during a day. To alleviate the pressure on the air traffic controllers (ATCOs), sectors are being split during high traffic hours; to save the operational costs, the sectors are merged back when the traffic density decreases.

Finding optimal ways to split and merge sectors is a fundamental research problem in ATM, which has attracted a lot of attention over the years. One of the most important distinction between algorithmic sectorization paradigms is whether the resulting sectors are built up by gluing together some “elementary bricks” (hexagons [1], square pixels [2], 3D voxels [3], etc.), or are obtained by directly cutting the airspace into the sectors [4], [5], [6] (see survey [7] for a full taxonomy of approaches to sectorization). Work of the former type employs an Integer Program (IP) to build up the sectors optimally – this synthesis approach provides a lot of means to fine-tune sector design, but gives little control over the sectors shape (the sectors are produced by the “blackbox” IP); therefore additional tricks and/or postprocessing are then required to even keep each sector (simply) connected, and further care is needed to ensure that the sectors boundaries are smooth. On the contrary, direct partitioning of the airspace into sectors allows one to influence the sectors geometries within the algorithm (producing, in particular, convex sectors), but gives little flexibility in taking care of complex constraints and objectives.

The research area remains active: a new form of sectorization IP was developed in [8], [9], and novel tools, like simulated annealing [10], are also being explored.

Our contribution

We consider a sectorization approach falling into the second [geometric decomposition algorithms] category from the two paradigms outlined above. In a nutshell, the major difference between this paper and earlier research is that previous work aimed at
producing sectors of potentially complicated shapes (including 3d) taking into account the whole variety of sectorization quality measures; naturally, due to the complexity of the considered model, it was impossible to require that the obtained sectors are truly optimal according to some objective. Contrary to that, we consider only two KPIs (which, moreover, are based on a single input – the number of flights in the sector), but aim at attaining a perfect sector split into two parts in terms of the objectives. We show that, surprisingly, it is possible to find sectors which simultaneously balance the maximum and the average aircraft count, using the simplest straightline cut through the airspace. As a by-product, we obtain sectors of simple (convex) shapes and simple (straightline) boundary between the sectors.

The rest of the paper is organized as follows: The next section outlines the model. Section III presents problem formulation and our solution to it, culminating in our main result: simple straightline cut suffices to perfectly balance traffic between the sectors (Theorem 4). Section IV illustrates the use of our method on a synthetic example and reports on computational experiments with real-world flight data. The last section discusses possible extensions.

II. Modeling

This section describes our abstraction of the sectorization problem.

A. Complexity based on density

Prior research has singled out several desirable properties that a good sectorization should possess, see e.g., [7], [11]. The foremost requirement is the balance of control between the sectors – no sector should be overloaded with traffic while other sectors see only few airplanes. We therefore aim at balancing two measures of traffic complexity between the sectors:

1. The maximum number of aircraft in the sector over time
2. The average number of aircraft in the sector over time

The former is a standard measure of sector load, used in many prior works; it corresponds to the Monitor Alert Parameter (MAP) – the number of aircraft in the sector that should not be exceeded. The latter measure is proportional to the average dwell time of aircraft in the sector [12]. While the maximum aircraft count represents the peak complexity of the sector, the average count represents the total complexity over the day. Balancing the two objectives tends to give a fair distribution of the traffic between the sectors.

We acknowledge that looking only at the number of aircraft (and not at flight directions) is a delimitation of our model, as it ignores the control complications caused by converging aircraft trajectories (the celebrated Dynamic Density [13], [14], [15]). We believe that our setup nevertheless provides a good starting point for more elaborated models. This was confirmed with operational experts from Nordic Unified Air traffic Control (NUAC) and Swedish ANSP Luftfartsverket (LFV) who manage enroute traffic over Sweden – the subject of our experiments (Section IV): the need for complex deconfliction maneuvers is rare in the uncongested Nordic airspace. We remark that aircraft count and, in particular, its maximum over time, was, in fact, in focus also in most of the earlier work on geometric airspace partitioning.
B. Binary split

In a generic sectorization problem, the input are flight trajectories through an airspace, and the goal is to partition the airspace into some number $K$ of sectors, while redirecting a set of constraints. We consider the simplest case of splitting into only $K = 2$ sectors. While in most practical cases, one would be looking for a larger number of sectors, our split into 2 parts may be used as the basic ingredient for a more complex sectorization: since the binary split increases the number of parts by 1, it can be applied recursively to obtain an arbitrary number $K > 2$ of sectors (however, if $K$ is not a power of 2, some parts may not be perfectly balanced). The recursive splitting of high-traffic sectors may be motivated also by the subsequent recursive merging: when the sectors are to be merged during low-traffic hours, it is important that the merged sectors also have balanced workload and bounded maximum complexity.

C. Flight segments

We assume that aircraft trajectories are given as a set of time-stamped segments – the standard data format in EUROCONTROL’s demand data repository (DDR2) which we use in our experiments (Section IV).

D. Straight boundary

It is commonly acknowledged that the existing sectors have suboptimal geometries, often following state boundaries (which were optimized for anything but ATC); in fact, SESAR Joint Undertaking puts a lot of effort into removing country boundaries from consideration when designing the sectors – one important initiative is establishing FABs (functional airspace blocks) over the harmonised European skies. A good sector should have “nice” shape and “simple” boundary – subjective criteria which are often formalized by requesting that the sectors are convex [7], [11], [9] (convexity is a desirable property also from the operational point of view because, by definition of convexity, any straightline-segment flight intersects a convex region at most once, thus never re-entering the sector). Note that if the boundary between two sectors has a vertex (i.e., is not a single straightline segment), then one of the sectors is not convex [Fig. 1, left]. Thus, to ensure convexity, we require that the sector boundary is a straightline segment, or equivalently that the sectorization is obtained by a single straightline chord joining two points on the boundary of the airspace [Fig. 1, right].

FIGURE 1 — Left: A rectangular airspace (dotted) partitioned into sectors $L$ and $R$; a “kink” (black dot) in the boundary between the sectors makes sector $R$ non-convex – a flight segment (blue) may exit the sector and re-enter it again. Right: Chord $C$ (bold) is the boundary between sectors $LC$ and $RC$; blue segments are flights.

As an additional constraint we require that the chord does not go through a flight segment, i.e., no flight segment should belong to the chord. This restriction makes sense from the operational perspective – the control over the plane should not be shared between the sectors (the only time when two ATCOs have to attend to the flight is when it crosses between the sectors). In fact, even stronger forms of this restriction may be imposed – that any flight segment lies somewhat “deep” inside within the sector (the distance between the flight and the sector boundary is larger than a given tolerance parameter), that the angle at which the flight segment crosses the boundary is not too small, etc.
III. Problem formulation and Solution

Formally, the input to our problem consists of:

- The region of interest \( P \) (the airspace) which is to be split into sectors. We assume that \( P \) is convex (for otherwise, decomposing \( P \) into convex sectors may not be possible).

- A set \( S \) of \( n \) straightline segments, where each segment \( s \in S \) is a 6-tuple \((a_s, b_s, c_s, d_s, e_s, f_s)\) where \((a_s, b_s), (c_s, d_s)\) are the coordinates of the segment start and end points resp., and \(e_s, f_s\) are the times when the aircraft enters and leaves \( s \) resp. For simplicity we assume that the segments are traversed by the planes with the same speed (the assumption is not essential and can be easily lifted). Equivalently, each element of \( S \) is a segment in the \((x, y, t)\)-space (Fig. 2, left). We will also make the General Positioning Assumption (GPA) that \( S \) is nondegenerate in the sense that no 3 segments go through a common point, no 3 segment endpoints are collinear, etc. (we will include other assumptions into GPA as needed) – one can make the GPA hold by perturbing the input. We will often identify a segment with the flight following it and speak e.g., about the time when a segment enters a region, and such.

Our output is a chord \( C \) – a (directed) straightline segment with endpoints on the boundary of \( P \). The chord splits \( P \) into two parts – the sector \( L_C \) to the left of the chord and \( R_C \) to the right of \( C \).

Let \( L_s(t) \) be the number of flights inside the left sector at time \( t \), or equivalently the number of segments intersecting \( L_C \) lifted to height \( t \) in the \((x, y, t)\)-space (Fig. 2, right); define \( R(t) \) analogously. Let \( ML(C) = \max_{t \in [0,T]} L_C(t) \) denote the maximum aircraft count in \( L_C \), where \( T \) is the given time horizon. The maximum count in \( R_C \) is defined analogously: \( R_C(C) = \max_{t \in [0,T]} R_C(t) \).

The average number of aircraft in \( L_C \) is

\[
AL'(C) = \frac{1}{T} \int_{\Delta_{L_C}} L_C(t) \, dt
\]  

(1)

By ergodicity, this average number is proportional to the total length of the segments from \( S \) inside \( L_C \). Indeed, for a segment \( s \in S \), let \( 1(s, t) \) be the indicator function of the presence of \( s \) in the sector over time: \( 1(s, t) = 1 \) if \( s \) is in \( L_c \) at time \( t \), and \( 1(s, t) = 0 \) otherwise. Then \( \int 1(s, t) \, dt \) is the time that the aircraft, following \( s \), spends in \( L_C \):

\[
\int 1(s, t) \, dt = |s \cap L_c|/v
\]  

(2)

where \( |s \cap L_c| \) is the length of \( s \) inside the sector and \( v \) is the speed of the aircraft.

Counting separately the contribution of each aircraft, we can write \( L_C(t) = \sum_{s \in S} 1(s, t) \) and rewrite (1) as

\[
AL'(C) = \frac{1}{T} \int_{s \in S} 1(s, t) \, dt
\]  

(3)
Interchanging the integration with the summation and using (2), we obtain

\[
AL'(C) = \frac{1}{IV} \sum_{s \in S} |s \cap L_C|
\]

Since \(v\) and \(T\) are independent of \(C\), we use

\[
AL(C) = \sum_{s \in S} |s \cap L_C|
\]

i.e., the total length of the segments in \(L_C\), as the measure of the average aircraft count in the sector. We define

\[
AR(C) = \sum_{s \in S} |s \cap R_C|
\]

analogously.

**The objectives**

As the measures of how good a split a chord makes, we use the differences in the maximum and average numbers of aircraft in the left and right sectors. Specifically, for a chord \(C\), the max-imbalance or M-imbalance is defined as \(M(C) = \|ML(C) - MR(C)\|\); similarly, the avg-imbalance or A-imbalance is \(A(C) = \|AL(C) - AR(C)\|\). With the above defined notation, our problem can be formally stated as

\[
\text{Given the airspace } P \text{ and the set } S \text{ of flight segments, find the chord } C \text{ balancing the maximum and average aircraft count between the sectors } L_C \text{ and } R_C, \text{ i.e., the chord minimizing the } M \text{- and } A \text{-imbalances } M(C) \text{ and } A(C).
\]

Since \(A(C)\) and \(A'(C)\) differ by a factor of \(vT\), minimizing \(A(C)\) is equivalent to minimizing \(A'(C)\).

**A. Listing M-imbbalances**

We prove that even though there are uncountably many chords, the number of possible max-imbbalances is only quatic in the number of segments, because many of the chords lead to the same M-imbalance:

**Theorem 1.** The number of distinct max-imbbalances is \(O(n^4)\).

**Proof.** Let us look closer at the interaction between chords and segments in \(S\). In this proof we will work in the \(\{x, y, t\}\)-space, so by a chord \(uv\) we will actually mean the vertical plane in the space, going through the points \(u\) and \(v\) on the boundary of \(P\) in the \(\{x, y\}\)-plane.

Let \(H\) be the set of the horizontal planes \(t = e_s, s \in S\) through the starting and ending points of all segments in \(S\). We call the points in \(H \in S\) critical. That is, critical points are intersections of the planes with the segments (e.g., the hollow circles in Fig. 2, right are critical points).

Define the combinatorial type of a chord to be the subset of the critical points to the left of the chord and say that a chord is canonical if it passes through two critical points (Fig. 3). It is easy to see that canonical chords define combinatorial types: consider any chord and transfer it parallel to itself until it hits a critical point, and then rotate the chord around the hit point until it hits another critical point (i.e., until the chord becomes canonical); during the transfer and the rotation, the combinatorial type of the chord does not change (since changing the type requires moving over a critical point). Moreover, chords of the same combinatorial type have the same M-imbalance. To show this, use the same transfer-and-rotation to move any chord \(C\) onto the canonical chord \(C'\) of the same type. During the
motion, the sector boundary does not intersect any segment, hence \( L_+ \) and \( L_-^* \) contain the same subset of the aircraft at any time \( t \), meaning that also the maximum number of aircraft in the sectors are the same \( M L \left( C \right) = M L \left( C^* \right) \). By the same argument, \( M R \left( C \right) = M R \left( C^* \right) \), and so the max-imbalance of all chords having the same type as \( C^* \) is the same.

**FIGURE 3** — \( C \) (dashed) is moved (1) and rotated (2) until becoming canonical (bold). Circles are critical points

Finally, since \(|S| = n\), there are \( O(n^2) \) critical points and \( O(n^4) \) canonical chords. Hence there are also \( O(n^4) \) different possible values for \( M \left( C \right) \) — one per chord combinatorial type; moreover, they all can be listed algorithmically by computing the canonical chords.

It follows from the above proof that the M-imbalance is a “continuous” function of \( C \):

**Corollary 2.** As the chord \( C \) moves infinitesimally, \( M \left( C \right) \) changes by at most 1.

**Proof.** Changing the maximum count \( M L \left( C \right) \) or \( M R \left( C \right) \) by 2 would require the chord to jump over 2 critical points simultaneously, which is not possible by the General Positioning Assumption. Moreover, as \( C \) moves over a single critical point, only one of \( M L \left( C \right) \), \( M R \left( C \right) \) changes by 1; thus also the difference \( M \left( C \right) = \left| M L \left( C \right) - M R \left( C \right) \right| \) changes by 1.

We use the corollary in the next subsection to prove our main result (Theorem 4).

### B. Minimizing both M- and A-imbalance

Previous subsection showed how to find all possible max-imbbalances. However, it did not show how to balance the maximum counts, and did not consider avg-imbalance at all. The reminder of this section fills both gaps: we “scroll through” all A-imbbalances and show that the maximum and the average may be balanced simultaneously.

The crucial fact is that for any position of one endpoint of the chord, there exists a unique position of the other endpoint that vanishes the A-imbalance:

**Lemma 3.** \( \forall u \in \partial P \exists v : A(uv) = 0 \)

**Proof.** If \( v \) is immediately clockwise from \( u \) along the boundary \( \partial P \) of the airspace, then \( L_{uv} \) and hence \( AL(uv) \) are very small (essentially \( AL(uv) = 0 \)); on the contrary, if \( v \) is immediately counterclockwise from \( u \), then \( L_{uv} \) is (almost) the entire \( P \), and hence \( AL(uv) \) is large (essentially equal to the total length of all segments). Since \( AL(uv) \) changes continuously and monotonically as \( v \) moves along the boundary of the airspace, by the Intermediate Value Theorem, there exists a unique point \( v \) where \( AL(uv) \) passes over half of the length of all the segments.

In fact, the explicit dependence \( v(u) \) of \( v \) on \( u \) may be derived using the Implicit Function Theorem: One may redefine the combinatorial type of a chord as the set of segments endpoints that are to the left of the chord (here, unlike in there previous subsection, we work in 2d); similarly to the previous section, it can be shown that there is only a quadratic number of the types, and that they are defined by the set \( L \) of \( \Theta(n^2) \) lines passing through all pairs of segments’ endpoints. One may scroll then through all combinatorial types. Chords from the same type have their endpoints on two fixed arcs of \( \partial P \) (the arcs are defined by lines from \( L \)), and intersect the same set of segments from \( S \) (since the set of intersected segments is uniquely determined by the endpoints to the left of the chord). Thus, for a fixed combinatorial type, and hence fixed arcs to which \( u \) and \( v \) belong, the total length \( AL(uv) \) of segments in \( L_{uv} \) is given by the same formula which may be explicitly written. If \( u \) and \( v \)
are such that \( AL(uv) = AR(uv) \), then \( AL(uv) = \text{const} \) (the common value of \( AL(uv) \) and \( AR(uv) \) is equal to the half of the total length of all segments). Setting the full differential to 0

\[
\frac{\partial AR(u,v)}{\partial u} du + \frac{\partial AR(u,v)}{\partial v} dv = 0
\]

we obtain, in accordance with the Implicit Function Theorem,

\[
\frac{dv}{du} = \frac{\partial AR(u,v)}{\partial v} \cdot \frac{\partial AR(u,v)}{\partial u}
\]

Equation (4) may be solved for \( v(u) \) (analytically or numerically).

We are now ready to prove our main result – existence of the “doubly-balanced” split:

**Theorem 4.** \( \exists C: M(C) = A(C) = 0 \)

**Proof.** For a point \( u \in \partial P \) on the airspace boundary, let \( C(u) = C(uv) \) be the chord that vanishes the A-imbalance: \( AL(uv) = 0 \) (existence and uniqueness of \( C(u) \) is guaranteed by Lemma 3). Let \( M(u) = M(C(u)) \) be the M-imbalance of the chord.

Take an arbitrary point \( u_1 \in \partial P \), and let \( u_2 = v(u_1) \) be such that \( AL(u_1u_2) = 0 \). By symmetry, \( u_1 = v(u_2) \), i.e., \( C(u_1) \) is the same chord as \( C(u_2) \), just going in the opposite direction [Fig. 4]. In particular, the left and right sectors of the chords swap places (\( L(u_1u_2) = R(u_2u_1) \)), implying that

\[
ML(u_1u_2) - MR(u_1u_2) = MR(u_2u_1) - ML(u_2u_1)
\]

(5)

**FIGURE 4 — As the A-balanced split rotates, the sectors swap places**

Now, suppose that \( M(u_1) \neq 0 \), e.g., that \( ML(C(u_1)) > MR(C(u_1)) \). Move \( u_1 \) along \( \partial P \). By Corollary 2, during the motion \( M(u_1) \) changes continuously “in integers”, i.e., changes in increments/decrements of 1. By equation (5), when \( u_1 \) gets to \( u_2 \), the M-imbalance changes the sign.

Thus, somewhere between \( u_1 \) and \( u_2 \) there exists a point \( u \) where \( M(u) \) crosses 0.

One may find the doubly-balanced split using equation (4) to “watch” how \( C(u) \) changes as \( u \) moves along the boundary of \( P \), and using Theorem 1 to keep track of \( M(u) \).

**IV. Experiments**

Figure 5 shows a small synthetic example. On the left, the red circular stepwise curve is the graph of \( ML(uv) - MR(uv) \) as the function of \( u \). For visualization clarity, the location of \( u \) is represented by points on a circle surrounding the rectangular airspace (instead of points on the boundary of the airspace itself). That is, the “\( u \)-axis” of the graph is the blue circle around the airspace, and the value of the function (the red curve) is given by the distance from the center of the circle. In fact, the function takes only three values \([-1,0,1]\) – the circles represent the levels of the function. It can be seen that the difference \( ML(uv) - MR(uv) \) indeed changes by 1 at certain points and, in fact, stays equal to 0 at quite many angle ranges, and that the graph is centrally symmetric about the circle center, in accordance with the theory developed in the previous section. On the right, the times of maximum aircraft counts are shown.
For a real-world example, we took airspace over Sweden. The Swedish Flight Information Region (FIR) was enclosed into a rectangle $P$, to avoid dealing with the complicated outer boundary, which mostly follows the state boundary. A busy day ($T=24$ hrs) of flight data was downloaded from Eurocontrol’s DDR2 repository. We restricted attention to the FIR overflights: our focus is on partitioning the upper airspace, and those aircraft that landed and/or took off within the region would mostly be handled in the transition airspaces. All the flights were clipped off at the points where they crossed $P$ (each of our flights had a unique entry and a unique exit point from $P$). We replaced every flight trajectory with the single segment between the crossing points, thus emulating free routing through the airspace (which is in place in the Danish–Swedish FAB for several years). Even though some of the flightplans did contain waypoints inside $P$, using the direct segments allowed us to exclude the turning points as workload hotspots (attention attractors) for ATCOs, providing yet another justification for using only aircraft count as the complexity indicators. (Alternatively, our algorithms could take as the input the flightplans together with the turning points.)

We then used Theorem 1 to zoom in on cuts with 0 maximbalance: we drew canonical chords through pairs of critical points, which split $\partial P$ into intervals (Fig. 6). For every pair of intervals $I_1, I_2$ such that for any $u_1 \in I_1, u_2 \in I_2$ the combinatorial type of the chord $u_1 u_2$ is the same, we evaluated $M(u_1 u_2)$ and kept only pairs with no M-imbalance (their existence is guaranteed by Theorem 4).

Finally, we searched for avg-balanced splits of the airspace into two parts as per Theorem 4 (since we handled traffic over the whole FIR, our parts rather represent control centers than sectors). To avoid going around the full boundary of the airspace, we looked for the A-balanced chords $A(u_1 u_2)$ only for $u_1, u_2$ lying within the pairs of intervals $I_1, I_2$ that have $M(u_1 u_2) = 0$ for all points $u_1 \in I_1, u_2 \in I_2$ (i.e., the intervals identified as described in the previous paragraph).

Solving the differential equation (4) may be a daunting task in practice; instead, we used monotonicity of $AL(u_1, u_2)$ as the function of $u_2$ and performed a binary search for the A-balanced chord $A(u_1 u_2)$. Specifically, for each pair of intervals $I_1, I_2$, we took a large set $U \subset I_1$ of points laid out regularly in the first interval (essentially $U$ is a dense 1d grid within $I_1$). For each point $u_1 \in U$ we evaluated $AL(u_1, I_2), AR(u_1, I_2), AL(u_1, j)$ and $AR(u_1, j)$ where $i$ and $j$ are the endpoints of $I_2$. If the differences $AL(u_1, I_2) - AR(u_1, I_2)$ and $AL(u_1, j) - AR(u_1, j)$ had the same sign (both positive or both negative), we discarded $u_1$; otherwise we knew that

**FIGURE 5** — Left: Black circles are the critical points; the M-imbalance changes in increments or decrements of 1 (the red stepwise function). Right: The sectors shifted up to the times at maximum aircraft counts

**FIGURE 6** — Two canonical chords (dashed) and points $u_1, u_2$ in the intervals $I_1$ and $I_2 = (i, j)$ on $\partial P$. If $AL(u_1 u_2) > AR(u_1 u_2)$, the next candidate location for $u_2 = v(u_1)$ is tried clockwise of the current $u_2$; otherwise, the search for $u_2$, delivering $A(u_1 u_2) = 0$, continues counterclockwise.
there exists a point \( u_2 \in I_2 \) such that \( A(u, u_2) = 0 \). We searched for \( u_2 \) with the binary search: given a candidate location for \( u_2 \), we computed the length of segments inside \( L_{uu_2} \), and directed the search clockwise or counterclockwise around the boundary depending on whether \( AL_{uu_2} \) was smaller or larger than the total length of all segments [see Fig. 6].

The pseudocode for finding the balanced cut is presented in Algorithm 1. The computational complexity of the algorithm is \( O(n^3) \).

Figure 7, left shows the resulting split of the airspace. The chord cuts out Sweden off the congested area southeast, suggesting that there is essentially as much control work to do in Sweden as there is in the adjacent European airspace. It is worth noting that this is obtained fully automatically with our algorithms, without any human looking over the map.

**FIGURE 7 — The M- and A-balanced split of \( P \). Left: The rectangle is the bounding box of Swedish FIR. Right: A smaller rectangle, more tightly focusing on the flights over Sweden**

Of course, Swedish aviation authorities are less excited about separating Sweden from Europe; the interest is in partitioning the Swedish airspace itself. To accomplish that, we shifted the right boundary of the rectangle \( R \) to the left, so that it runs along the longitude of 18 degrees, and recomputed the split. (Our techniques work for partitioning non-rectangular airspaces just as well, but for this experiment we kept \( P \) a rectangle – shifting the boundary to the left made a northern part of Sweden fall out to the right of the rectangle, but the fallen out contained almost no traffic anyway.) Figure 7, right shows the result. Again, without any human oversight, our algorithm separated the truly more congested southern part of the Swedish FIR from the north – a reasonable division.

As mentioned in Section II-B, one natural way to proceed with our method is to do recursive balanced partitioning: split a large airspace into two and continue splitting the parts. In terms of the considered example, this would mean that both of our chords could be used simultaneously: first a SW–NE chord (like in Fig. 7, left) cut out Sweden from Europe, and then a SE–NW chord (ala Fig. 7, right) partition Swedish airspace itself.

Figure 8, left shows the two chords combined: it can be seen that the combination results in a high-degree vertex of the sectorization [at the lower side of \( P \), around the longitude of 18 degrees] where several sector boundaries meet – an undesirable artifact. To remedy this, we note that our technique produces all perfectly balanced cuts – and there may be more than one [see e.g., our synthetic example in Fig. 5]. Figure 8, right shows all doubly-balanced cuts of the airspace over Sweden (the part on the left of the cut in Fig. 7, left), and Figure 9 shows two of the cuts used to divide the Swedish part. The cuts were picked manually as the ones making the largest angle with the primary cut from Fig. 7, left i.e., so that the chord at the next level of the recursion is as perpendicular as possible to the previous-level chord) – designing algorithms for automated choice of the perpendicular cuts is outside the scope of this paper.
**Algorithm 1: Balanced cut**

<table>
<thead>
<tr>
<th>Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$I \leftarrow \emptyset$ $\triangleright$ Set of points on $\partial P$;</td>
</tr>
<tr>
<td>2</td>
<td>$C \leftarrow$ critical points of $C$;</td>
</tr>
<tr>
<td>3</td>
<td>foreach pair of points $(c_1, c_2) \in C$ do</td>
</tr>
<tr>
<td>4</td>
<td>$(i, j) \leftarrow \partial P \cap$ line through $c_1, c_2$;</td>
</tr>
<tr>
<td>5</td>
<td>$I$.append$(i, j)$;</td>
</tr>
<tr>
<td>6</td>
<td>end</td>
</tr>
<tr>
<td>7</td>
<td>Sort $I$ in counterclockwise order;</td>
</tr>
<tr>
<td>8</td>
<td>foreach pair of intervals $(I_1, I_2) \in I$ : combinatorial</td>
</tr>
<tr>
<td>9</td>
<td>type of $u_1, u_2$ is the same $\forall u_1 \in I_1, u_2 \in I_2$ do</td>
</tr>
<tr>
<td>10</td>
<td>$M \leftarrow M$-imbalance$(u_1, u_2)$;</td>
</tr>
<tr>
<td>11</td>
<td>if $M = 0$ then</td>
</tr>
<tr>
<td>12</td>
<td>if $\exists$ A-balanced cut $u_1, u_2$ then</td>
</tr>
<tr>
<td>13</td>
<td>return $C \leftarrow u_1, u_2$;</td>
</tr>
<tr>
<td>14</td>
<td>end</td>
</tr>
<tr>
<td>15</td>
<td>end</td>
</tr>
</tbody>
</table>
V. Conclusion

We showed how to split a sector into two parts while balancing the traffic density between the obtained sectors. Our sectorization uses only single segment as the boundary between two sectors, which provides for sectors of noncomplicated (in fact, convex) shape – a useful property [7], [9], [11] if applied recursively, in order to obtain an arbitrary number $K > 2$ of sectors, the simple 1-chord split is reminiscent of the binary space partition sectorization [6]). Simple splits are advantageous for subsequent dynamic sectorization, when the sectors are split and merged on a regular basis: we believe it will make it easier for the ATCOs to qualify to handle sectors with the simple boundary (i.e., to be certified both for split and for merged sectors).

Future work may extend our solution in many ways: (recursively) splitting into multiple sectors, using more complicated boundary between the sectors, caring about directions of the flights (we looked only at the aircraft counts), taking into account interaction between the flights and the sectors boundaries (e.g., minimizing the number of sector changes), etc. In some cases there may exist more than one sectorization that simultaneously balances the maximum and the average number of flights; in such situations it would be of interest to choose the best sectorization based on some other KPI (in fact, as our method produces all balanced binary splits, the subsequent choice of the best sectors may be left to the human designer). We hope that despite the relative simplicity of our approach, it can be embedded as a basic unit into more sophisticated sectorization tools.

Acknowledgements

We thank Anders Nyberg (NUAC) for discussions on the topic of the paper. This work is part of the Swedish Enroute Airspace Optimization (SWEAO) project supported by the Swedish Transport Administration (Trafikverket) via the Swedish Air Navigation Service provider LFV (Luftfartsverket).

REFERENCES


Big data
and liability
AUTONOMOUS SYSTEMS IN AVIATION: BETWEEN PRODUCT LIABILITY AND INNOVATION

Ivo Emanuilov  
Centre for IT & IP Law-imec (CITIP)  
KU Leuven  
3000 Leuven, Belgium  
ivo.emanuilov@kuleuven.be

Abstract—Increasingly autonomous cyber-physical systems based on self-adaptive software are making their way into the aviation domain. However, the combination of their adaptive learning properties and the safety goals of aviation create unique legal and regulatory challenges for the manufacturers and regulators of such systems alike. This paper argues that some of the fundamental concepts of the product liability regime in the EU and their interpretation deprive manufacturers of autonomous systems of two essential defences: the ‘state of the art’ defence and the regulatory compliance defence. The hesitation in the direction of the overall approach to regulating and certifying autonomous systems in aviation induces legal uncertainty which can only be overcome through surgical legislative intervention. The paper formulates recommendations for amendments in light of the ongoing evaluation and pending review of the Product Liability Directive.

I. Introduction

Software plays a critical role in commercial aviation. Navigation, aircraft control and other functions of flight management systems are now largely automated by software. With pilots’ role becoming mainly supervisory in nature, the trend over the last few decades has been clearly one of steady growth of automation. While automation has undoubtedly improved aviation safety, it has also been a contributing factor to several fatal incidents [1].

The continuous increase in air traffic has called for transformation of the aviation industry. The disruptive power of new technologies, such as increasingly autonomous cyberphysical systems and machine learning, promises to improve the capacity and profitability of air services and contribute to improving safety, security, environmental protection and infrastructure modernisation [2]. However, the “coupling” of cyberspace with the physical world gives rise to significant challenges in terms of reconciling the distinct features of the two environments. Cyberspace, an ideal environment governed by the rules of software code, is opposed to the physical environment of aviation governed by the laws of physics, linear time and stringent safety rules. Connected aircraft and digital air traffic management (ATM) systems are just beginning to leverage the benefits of this cyber layer. At the same time, increasingly autonomous systems, such as unmanned aircraft systems, relying on self-adaptive software and machine learning, promise cost savings and facilitate new opportunities for air carriers [3], [4].

Against the background of fast-paced developments and growing complexity of software-intensive aviation socio-technical systems [5], regulators and legislators are

This research has received funding from the imec-ICON project SafeDroneWare.
facing the challenges of the growing divide between technology and regulation. The conservative nature of aviation safety regulation is now confronted with the influx of a wide range of commercial off-the-shelf technologies which cannot be certified using the safety standards for aircraft and ATM systems. Furthermore, existing certification and standardisation processes are based on the assumption that a system’s correct behaviour “must be completely specified and verified prior to operation” [6]. This constitutes a significant barrier to the development of new autonomous systems relying on adaptive software and machine learning algorithms as they are intrinsically self-directed and non-deterministic.

Certification and standardisation also provide the manufacturers of such systems with a certain level of assurance regarding the compliance of their products. More specifically, under product liability law, manufacturers can be held liable for damages to third parties caused by defective products. Until recently, manufacturers’ liability in the aviation industry was considered more the subject of academic debate rather than a practical issue. However, in 2015, the Spanish Supreme Court held liable the manufacturers for product defects of the collision avoidance system (TCAS) installed on board the aircraft involved in the Überlingen mid-air collision accident [7]. A significant precedent “reaffirm[ing] product liability in the aviation domain”, this decision also demonstrates the difficulties in allocating liabilities in socio-technical systems and the role of certification as a determining factor [8].

At EU level, Directive 85/374/EEC on liability for defective products (“Product Liability Directive”) establishes a harmonised strict liability regime which holds the producer liable for damage caused by a defect in their product. The applicability of this directive in the realm of aviation was recognised from the very beginning [9]. The general rule is qualified by six exceptions, two of which are particularly important for manufacturers of autonomous systems in aviation. According to Article 7 (d) and (e), a manufacturer may escape liability if they prove that:

- the defect is due to compliance of the product with mandatory regulations issued by the public authorities (known as ‘regulatory compliance defence’); or
- the state of scientific and technical knowledge at the time when they put the product into circulation was not such as to enable the existence of the defect to be discovered (known as ‘development risk’ or ‘state of the art defence’).

This paper argues that these two exceptions can hardly be invoked by manufacturers in the context of increasingly autonomous systems which in turn induces legal uncertainty. More specifically, for manufacturers of hardware and software for the nascent market of commercial unmanned aircraft systems in the EU this might act as a barrier stifling innovation. In light of the ongoing evaluation of the directive, the paper briefly discusses several possible solutions.

The paper is organised as follows:

- section II provides an overview of the specific legal and regulatory challenges of increasingly autonomous adaptive systems;
- section III discusses the central notion of ‘defect’ and the applicability of the notion of ‘state of the art’ to increasingly autonomous adaptive systems;
- section IV focuses on the difficulties in the certification of increasingly autonomous adaptive systems and their impact on the exercising of the regulatory compliance defence;
- section V outlines the impact of these challenges on the developing European market for unmanned aircraft systems;
- section VI looks at the possible solutions to mitigate the legal uncertainty and create enabling conditions for the development of autonomous systems in aviation.
II. Automation and autonomy

The terms automation and autonomy are often confused and even used interchangeably. However, while automated technologies have been used in aviation for quite some time now, this is not the case for autonomous technologies. Automation refers to a system performing its function with little or no human involvement where the system’s performance is limited to its predefined tasks. Unlike automation, autonomy refers to systems which exhibit self-directed behaviour and can dynamically respond and adapt to events which have not been pre-programmed [6].

As future cyber-physical systems in aviation will increasingly rely on autonomous technologies, this will put to the test some established regulatory conventions. From a legal and regulatory perspective, at least two groups of challenges draw attention.

The first challenge is linked to the basic question of when a product is defective, and whether autonomous systems’ learning and adaptive capabilities render the safety expectations test and the manufacturer’s state of the art defence unfit. These problems will be discussed in section III.

The second challenge is related to how certification authorities verify and certify autonomous systems, given predictability and certainty are at the core of the current certification process. Authorities would have to assure the new systems are at least as safe as the existing ones. The US Federal Aviation Administration (FAA) and the European Aviation Safety Agency (EASA) can certainly rely on the ‘special conditions’ for such ‘non-conventional aircraft’, as established in FAR 21, Paragraph 16 and EASA Part 21, Paragraph 21.A.16B, to add safety standards ensuring equivalent level of safety to the one in the airworthiness regulations/certification standards [10]. However, this does not resolve the issue of how acceptability and practicability would be balanced to establish a sufficient level of safety of autonomous systems. The uncertain context and independence of autonomous systems could render existing performance standards inapplicable to them. While some solutions to this problem have already been proposed, their impact on producers’ liability will be discussed in section IV.

III. Defects: between consumer expectations and the state of the art

Under the Product Liability Directive, an injured party can assert a claim against a producer on three conditions: there must be a defect, a damage and a causal relationship between the two.

Article 6 of the Product Liability Directive provides that a product is defective when it does not provide the safety which a person is entitled to expect. It is beyond any doubt that the reasonable safety expectations for an aviation product will be very different from, for example, those for a smartwatch. Furthermore, it should be understood that ‘products’ here in the context of autonomous systems in aviation may include hardware components as well as software applications etc. Thus, the preliminary question of whether an autonomous system is a ‘product’ in the first place must be answered.

The Product Liability Directive defines in Article 2 products as all ‘movables’, even when incorporated in immovables. Considering the software-intensive nature of autonomous systems in aviation, it is reasonable to ask if software fits within this definition. The European Commission’s view is that the directive “applies to software in the same way [as to other movables], moreover, that it applies to handicraft and artistic products” [11]. However, it is uncertain whether the directive applies solely to embedded software and software available on tangible medium or also to “software as a service” [12]. Legal uncertainty remains high, however, given that up to date there has been no case law on
whether the Product Liability Directive applies to software, a point also acknowledged by the Commission itself [13]. In any case, as most autonomous aviation systems are likely to be cyber-physical systems operating embedded software, they should fall within the ambit of the definition of a ‘product’.

It was mentioned that for a product to be defective, it must fail to meet a person’s safety expectations. Thus, the notion of ‘defect’ in EU product liability law is not grounded in technical defects but rather in the general public’s expectations of the required degree of safety. In the recent Boston Scientific case, the Court of Justice of the EU (“CJEU”) held that “[t]he safety which the public at large is entitled to expect […] must therefore be assessed by taking into account, inter alia, the intended purpose, the objective characteristics and properties of the product in question and the specific requirements of the group of users for whom the product is intended” [14]. In his Opinion, the Advocate General (“AG”) Bot suggested that a product defect “can exist irrespective of any internal fault in the product concerned” and that the “triggering factor does not reside in the product fault, but in the fact that the product does not provide the safety which a person is entitled to expect” [15]. This implies that a product which is technically sound may still be defective if it fails to meet the expectations in terms of safety.

Concurring with the AG Opinion, the court went further and found that even the potential lack of safety may give rise to producer’s liability which, in the case’s context of medical devices, stems from “the abnormal potential for damage which those products might cause to the person concerned” [§ 40]. Thus, the court ruled that “where it is found that such products belonging to the same group or forming part of the same production series have a potential defect, it is possible to classify as defective all the products in that group or series, without there being any need to show that the product in question is defective” [§ 41]. While the court did not go as far as AG Bot to highlight the “preventive function” assigned to product liability law [§ 38] [15], it grounded its findings in teleological interpretation of the directive’s objective of fair risk apportionment and the high priority of consumer protection [§ 42, 47].

The decision has been criticised for its “counter-productive effects” in creating liability for potentially defective products and effectively rewriting the directive [16]. Despite some commentators’ opinion that the case has implications only for the medical devices industry, the broad policy objectives upon which the decision is based and the court’s use of broad language suggest otherwise [17]. Furthermore, provided the safety requirements for aviation systems are very high, an analogy to the “abnormal potential for damage” stemming from defects in such products may not be that far-fetched.

The criterion of ‘safety expectations’ merits special attention in the context of autonomous aviation systems. Autonomous systems are inherently self-adaptive; they control their behaviour in accordance with “context-relevant norms, constraints, or desiderata” [18]. Their objective characteristics come into conflict with the concept of ‘safety expectations’ which is based, inter alia, on the expectations at the time when the product was put into circulation and which cannot go beyond that point in time on grounds of subsequent better products (Article 6(1)(b) in relation to Article 6 (2) Product Liability Directive). Thus, it is not clear whether, for the purpose of assessing these expectations, the alteration of an autonomous system’s behaviour or an update/upgrade of its functionality could be considered a defective product if it fails to meet the safety expectations or as a new product that does not affect them. In addition, the public’s safety expectations are also determined by the presentation of the product [Article 6(1) [a] Product Liability Directive] which means that as well as generally known risks, any risks of which the public has been specifically informed by the manufacturer are also relevant. As a matter of fact, in the case of Boston Scientific, the claims were based precisely on alerts made by the manufacturer. Given the broad interpretation of potential defects and the likelihood of claims being asserted on the basis of a mere notice, manufacturers are likely to become discouraged to share information with the public where the risk is perceived to be considerably small [16]. This is even more so when it comes to autonomous systems in aviation where the very behaviour of the system
coupled with the high safety expectations in the sector may lead to it being considered a product of "abnormal potential for damage".

The liability for potential defects established in Boston Scientific raises troubling concerns as it cannot be easily reconciled with the adaptive behaviour of autonomous systems. For example, two autonomous systems may operate in two different contexts changing their behaviour according to different conditions and constraints. If one of them changes its behaviour in a way that does not meet the public’s safety expectations and causes damages, the potential defect liability doctrine would likely automatically render any other instance of the same system operating in a different context defective as well. As a result, this opens the door to an unlimited chain of claims against the manufacturer based on the potential lack of safety of the autonomous system.

Another interesting point could be made with respect to the public’s safety expectations for software-intensive systems in general. It is well-known in the software industry that commercial software is often shipped with flaws and defects and this more or less has been accepted by the public when it comes to standard software packages, such as word processors [19]. In the aviation domain, however, the software assurance process is much more thorough so that it can ensure the product is developed in line with very specific requirements and it does not exhibit any unintended behaviour [6]. Nevertheless, cases such as the Spanish case against the manufacturers of the TCAS involved in the Überlingen accident show that even this software can fail. A reasonable question, then, is how the presence of ‘bugs’, as inherent and unavoidable ‘features’ of present-day software, impacts the public’s safety expectations. In the English case of A and Others v National Blood Authority and another, the court referred to an exchange between Mrs Flesch MEP and the European Commission in June 1980, Viscount Davignon, on behalf of the Commission, stated that “nobody can expect from a product a degree of safety from risks which are, because of its particular nature, inherent in that product and generally known, e.g., the risk of damage to health caused by alcoholic beverages. Such a product is not defective within the meaning of ... the ... Directive” [20], [16]. It is interesting to see if manufacturers would employ a similar reasoning to argue that the fact software is never flawless effectively lowers the public’s safety expectations. Furthermore, a related question concerns the extent to which software’s inherent flaws impact producers’ liability for potential defects.

The liability for potential defects in Boston Scientific operationalises the risk of malfunction to become a defect in the future and not the occurrence of an actual defect [16]. This distinction is critical for manufacturers as it determines whether they may invoke the state of art defence under Article 7 (e) Product Liability Directive. As the public cannot have legitimate expectations of 100% safety, this simply means that if the risk of a product malfunctioning in the future reveals an “abnormal potential for damage”, without any specific indications of current or imminent malfunctioning, the manufacturer may not be able to rely on the state of the art defence.

As construed by the CJEU, the defence refers to the “objective state of scientific and technical knowledge of which the producer is presumed to have been informed” to the extent that this knowledge is “accessible at the time when the product in question was put into circulation” [21]. This means the state of the art must be objectively examined through the lens of the most advanced level of knowledge, regardless of the industrial sector concerned (§ 26). Thus, for a manufacturer to exonerate themselves from liability for potential defects, they must prove that they could not have known about the risk of product malfunctioning in the future, even with the most advanced level of scientific and technical knowledge.

Such an interpretation implies a very high standard for exoneration which is likely to leave manufacturers with no effective defence for the development risks they undertake. This is especially so in the case of autonomous aviation systems, where the learning and adaptation feedback loops may lead to changes in a system’s behaviour that creates new risks that could have been neither known nor foreseen. The problem is further exacerbated by the interaction and exchange of data between autonomous systems which
give rise to new and more complex safety risks. Without an objective standard for what an “abnormal potential for damage” constitutes, especially in the case of autonomous systems, the state of the art defence could easily become a thing of the past.

The state of the art is a moving target. In principle, this should mean that for the new risks emerging with the technology’s continuous development, manufacturers should be covered by the state of the art defence. The far-reaching implications of Boston Scientific, however, suggest that the product liability regime in the EU has a “preventive and prophylactic function that goes beyond merely reacting to damage materialization” [17]. While it is true that the court explicitly says that the products “may be”, and not “must be”, considered defective [16], the very possibility that a national court may concur with this reasoning produces legal uncertainty. Consequently, the product liability’s deterrence function may equally end up deterring innovation if applied broadly and indiscriminately.

IV. Certifying autonomy: between producers’ compliance and standard-setters’ liability

Standardisation and certification are essential for safety in aviation. It was already mentioned that the current certification standards are rooted in determinism and predictability which means that the correct behaviour of a system must be completely specified and verified prior to operation. However, regulatory authorities are now facing the challenges of introducing commonplace cyber technologies in aviation. This trend not only further the state of the art but also renders this traditional mindset largely inapplicable to autonomous systems based on adaptive software and machine learning algorithms, such as the ‘sense and avoid’ algorithms in unmanned aircraft systems.

The need of a new approach to the certification, verification and validation of increasingly autonomous systems in aviation has long been recognised as a critical bottleneck [3]. And while a good amount of research on the certification challenges is already underway, the implications of these challenges for the manufacturers’ liability for defective products has remained largely unexplored.

The link between certification and liability is evident from Article 7(d) Product Liability Directive which establishes the so-called ‘regulatory compliance defence’. A manufacturer can invoke this defence to exonerate themselves from liability if they prove that the defect is “due to compliance of the product with mandatory regulations issued by the public authorities” (emphasis added). Thus, if a manufacturer proves that the defect in their product is the result of their compliance with a mandatory (ie, binding) norm adopted by an authority with regulatory powers, they shall not be liable.

It is generally accepted that this defence has a rather narrow scope and can rarely be invoked successfully [22]. This is so because, first, in most cases legislation establishes minimum standards which provide manufacturers with a wide margin of appreciation and, second, the defect itself must be the result of compliance with the mandatory rule which would rarely be the case.

In a comparative perspective, the regulatory compliance defence in most European countries is of rather limited application.

First, the scope of the term “mandatory regulation” is construed restrictively. For example, in France, Spain, Austria, Germany and the UK compliance with norms establishing minimum legal or regulatory requirements, voluntary standards, private norms and technical standards issued by national or international standardisation organisations are excluded from the scope of the defence. Standards, however, may have a role in determining the public’s legitimate safety expectations in Austria, Spain and Norway. [23] The general stance is that for the “regulations” to be considered
“mandatory”, they should be embodied in legal provisions that force the producer to manufacture defective products (Germany, § 1 (2) no 3 ProdHaftG), constitute structural standards of production that cannot be disregarded (Italy, Art. 118 Consumer Code), or cover the design and/or composition of the product (Netherlands, Art. 6:185(1)(d) BW) [23].

Second, the case law dealing with the regulatory compliance defence is scattered and has until recently been of interest mostly to the pharmaceuticals and food industry. Thus, in the case of Pollard v Tesco Stores, the English court held that a violation of non-binding British standard is not conclusive proof of defect [24]. At the same time, in the case of Haribo, the Cologne Court of Appeal held that while compliance is not an automatic defence, it is a strong evidence that the product is not defective [25]. Broadly speaking, while compliance with standards does not amount to a regulatory compliance defence, it may still act as a presumption that the product is compliant [26]. Conversely, non-compliance with standards may be interpreted as failing to meet the legitimate safety expectations of the public.

The recent case of Überlingen (Manufacturers) breathed new life into the discussion on product liability in the aviation industry. In the case, the manufacturers of the collision avoidance system, which was installed onboard the two aircraft involved in the Überlingen accident, successfully invoked the defence because of compliance with mandatory requirements imposed by the FAA. The case concerned, inter alia, a software update that was available but could not be installed as the standard mandated the use of specific algorithms, ie the standard did not leave any margin of appreciation to the manufacturers. Remarkably, the court applied the Hague Convention on the Law Applicable to Products Liability of 1973 which, based on the principal place of business of the defendants, determined as applicable the law of the US states of Arizona and New Jersey. Nevertheless, authors agree that the essence of the regulatory compliance defence in US law is in line with its embodiment in the Product Liability Directive [8].

The decision in Überlingen (Manufacturers) is important in at least two directions: (1) it reinforces the applicability of the regulatory compliance defence to manufacturers in the heavily regulated aviation industry; and (2) brings to the fore the discussion of holding standard-setters and regulators liable for the choices they make in adopting mandatory standards. However, with the advent of autonomous systems, both ‘victories’ may prove to bring only short-lived comfort for the aviation industry.

First, the regulatory compliance defence would be hardly applicable to autonomous systems in the absence of sufficiently precise mandatory regulations. As rightly noted in literature, the current way of drafting standards can at best ‘codify’ the “desire that such systems be reliable” [18]. The certification challenges would likely require an entirely new regulatory approach. Several ways have been proposed in literature [18],[6]:

a) Employing existing regulatory approaches by limitation of the autonomy’s scope either by (1) placing a human in the loop or (2) making the operational context uniform.

b) Data-driven multistage approach modelled after the approval process for drugs for medical devices.

c) Modification of certification standards to enable a more dynamic software structure while keeping the existing safety principles.

d) Development of new verification approaches where testing is replaced by formal methods.1

e) Certification of adaptive functions providing advanced capabilities (eg, recovery from aircraft upset etc.) by treating the system differently based on the time at which it executes (eg, take-off, cruise, landing).

1 Mathematically based techniques for the specification, development, and verification of software aspects of digital systems which allow automated and exhaustive verification of properties [27].
f) Development of a licensing mechanism for autonomous systems based on the pilot licencing regime which, after demonstration of extensive knowledge and skills, leads to certification.

Nearly all proposed solutions have a bearing on the standardisation and certification process. The solutions in (a) and (c) seemingly involve minimal modifications to the certification process, while those suggested in (b), (e) and (f) will require either major changes or a complete shift of the certification paradigm.

In any case, from a legal point of view, if a certification authority and/or a standardisation body has issued a mandatory regulation which has been complied with by the manufacturer, the regulatory compliance defence will be available. This will not be the case, however, in the testing-based validations in (b) and (f) since their focus would be on the acceptable behaviour of a system which could easily be rendered invalid by a single change to the system [6]. A potential solution to the problem of unanticipated changes is the integration of the approach of machine-driven adaptation with human-driven evolution of the system [28]. It is questionable whether a mandatory regulation for autonomous systems could be precise enough as to prescribe specific design choices without leaving any margin of appreciation. Furthermore, any testing-based validation, no matter how comprehensive, could fail to account for the potential of a system learning new behaviour that has not been anticipated during the development. Legally speaking, any such unanticipated change, to the extent it fails to meet the safety expectations of the public, is likely to be treated as a defect. In such cases, even though the manufacturer would have complied with the mandatory regulation of a certification body, their product would still not qualify for the regulatory compliance defence because the defective behaviour would not be the result of compliance with the mandatory regulation and there would be no causal link between the two. This situation raises the reasonable question of whether the certification body could be held liable for its failure to conduct a comprehensive testing-based validation and against what standard its conduct would be judged.

It would be no exaggeration to say that the liability of certification and standardisation bodies is enigmatic. As reported in literature, the cases of standardisation and certification bodies held accountable for adopted mandatory standards are “notoriously rare” [8]. Claims that standard-setters and regulators are “not regularly accountable through legal mechanisms” [8], though, are imprecise. In the case of mandatory standards imposed by a state, the state could be challenged and the rules for liability of public authorities exercising regulatory functions should apply [26]. Admittedly, the situation is more complicated when it comes to the liability of standardisation and certification bodies at EU level.

In the EU, the EC has recognised EUROCAE as the competent body to cooperate with the European Standardisation Organisations (ESOs) in the preparation of European Standards (ENs) and Community Specifications (CS). EASA has equally recognised EUROCAE for its role in developing aviation safety technical documents. While these standards are considered soft law from a legal perspective, as they set minimum performance requirements, their recognition by certification bodies such as EASA in mandatory regulations essentially transforms them into hard law. In doing so, however, EASA does not become accountable for the choices made in these standards [8].

In principle, as a body of the EU, EASA’s decisions concerning airworthiness and environmental certification, pilot certification, air operation certification etc. may be appealed before a Board of Appeals. Furthermore, actions may be brought before the CJEU for the annulment of acts of EASA which are legally binding on third parties, for failure to act and for damages caused by the agency in the course of its activities.

The case law on the liability of EU institutions suggests that as a general rule liability could be attached only exceptionally, in cases involving a sufficiently serious breach

---

1 Eg the Radio Technical Commission for Aeronautics (RTCA) and the European Organisation for Civil Aviation Equipment (EUROCAE).
2 Eg FAA and EASA.
of a rule of law for the protection of an individual [29]. The court does not seem to
distinguish between liability for legislative and liability for administrative measures,
focusing instead on the criterion of the degree of discretion enjoyed by the institution
[30]. The difficulty here is that the intensity of the judicial review is often described as
“peripheral”, especially in cases where the body in question enjoys a wide margin of
appreciation [31]. Given the margin of appreciation enjoyed by EASA, one can imagine
establishing a manifest breach would be a cumbersome task.

It follows that holding the certification body liable for its failure to conduct a comprehensive
testing-based validation of an autonomous adaptive system could be a hard case. While
strengthening the accountability of standard-setters and certification bodies for their
design choices could be made by legislative amendments, given the maturity of the
current system, this is unlikely to happen in the near future. Thus, calls for considering
the potential impact of a certain standard on the liability of manufacturers in the drafting
process are a welcome ‘soft’ measure that could improve legal certainty [8].

V. Autonomy in the sky: unmanned aircraft systems

The challenges of liability of manufacturers for ‘defective’ autonomous systems in the
aviation industry may not be as remote as some would think. Considering the fast-
growing market of unmanned aircraft systems and the ensuing development of new
commercial services, these challenges may turn to be just around the corner.

In the recently proposed draft regulation laying down rules for unmanned aircraft
operations in the open and specific category, EASA explicitly provides for autonomous
operations in the specific category.5 ‘Autonomous operation’ is defined as an operation
during which an unmanned aircraft operates without the possibility for remote-pilot
intervention in the management of the flight. The proposal foresees that aircraft in the
specific category, which covers the majority of the commercially viable operations, are
organised around the concepts of operational authorisation issued to the operator by
a national aviation authority based on a risk assessment process. Thus, in line with
EASA’s operation-centric, proportionate, risk- and performance-based approach, in the
specific category it is the operator that is responsible for compliance with the technical
requirements laid down in the authorisation or the expected standard scenarios. Unlike
the specific category, in the open category the manufacturer is responsible for compliance
with the technical requirements based on the regime of essential requirements and
conformity assessment (CE marking).

As the operator would be the ultimately responsible for the technical requirements of
the unmanned aircraft in the specific category, the liability of the manufacturer could
only be engaged indirectly. For example, in the acceptable means of compliance6 listed
in an annex to the proposal, in the case of autonomous operations, the operator should
ensure that the UAS complies with the instructions provided by the manufacturer. These
instructions would certainly play a role in determining the public’s safety expectations.
Furthermore, any such ‘instructions’ may, in their own right, be treated as products for the
purposes of product liability.7 In this case, provided the acceptable means for compliance
and the guidance material are non-binding, the regulatory compliance defence cannot
be invoked. Similarly, manufacturers of software and software frameworks could be
held liable for defects in the provided software and the accompanying instructions.
In light of the issues with the definition of defect and the liability for potential defects,
discussed in section III above, software companies engaged in the development of
software frameworks or applications may face challenging legal uncertainty.

5 Article 2(1)(d) and UAS.SPEC.10 EASA Notice of Proposed Amendment 2017-05 (A).
6 Article 2(1)(b) of the proposal provides that acceptable means of compliance are non-binding standards which may be used
to demonstrate compliance.
7 The question of whether information as such falls within the ambit of the Product Liability Directive is subject to discussion
in literature. Strong arguments as to why information, especially when ‘materialised’ on a tangible medium, should be
treated as a product could be found in [23].
In addition to the open and specific categories, EASA foresees a third category of UAS operations (ie the certified category) which is not subject to regulation by the proposal. This category will require certification of the aircraft and licencing of the flight crew. Examples of such operations include, *inter alia*, large or complex UAS operations over assemblies of people, large or complex UAS operating beyond visual line of sight in high-density airspace, UAS used for transportation of people etc. Thus, potentially, any large-scale UAS operation would fall within the certified category. While EASA is planning to propose first rules for the certified category in the beginning of 20188, it is apparent that the regime will be based largely on the model of manned aviation. Conversely, this means that the challenges of the existing product liability regime, particularly with respect to the regulatory compliance defence, will also persist in the certified category.

VI. Recommendations and Conclusions

The digitalisation and increasing autonomy of aviation systems disrupts the traditionally conservative domain of aviation safety and puts to the test the limits of existing product liability rules and certification mechanisms.

While there are no doubts that the Product Liability Directive applies to the aviation domain, reasonable concerns have been raised as to whether its ‘strong-arm’ power can reach the major aviation technology producers which are currently mostly US-based.9 The stated aim of competitiveness and global leadership of the EU in the development of a “drone ecosystem” [32] is echoed in the regulatory actions of EASA. Furthermore, the rapidly growing number of companies and research organisations from the EU engaged in development of software for unmanned aircraft systems is a strong indicator of the EU’s innovative potential in developing autonomous systems for the aviation sector. This delicate balance, however, could be easily distorted by legal uncertainty induced by the fact that the existing product liability regime is arguably unfit for such purposes.

The analysis in the previous sections demonstrates that autonomous systems’ learning and adaptive capabilities are a significant challenge for the product liability regime. Most of these issues could only be resolved with legislative intervention. To this effect, the following amendments could be suggested:

- **Software should be included explicitly as a product and the definition should extend to cover both non-embedded software and ‘software as a service’.**

  This measure is critical for reinforcing the deterrent role of product liability, particularly for entrants that are new in the aviation industry, such as young companies and research organisations developing software for unmanned aircraft systems.

- **Objective standard for measuring the “abnormal potential for damage” must be crafted to restrain the otherwise broad scope of the liability for potential defects and to prevent innovation from stifling.**

  In the absence of an objective standard against which the criterion of “abnormal potential for damage” could be measured, the liability for potential defects introduced with *Boston Scientific* could have serious repercussions for producers engaged in autonomous systems development. The European legislator should also consider the legal nature of the liability engaged in these cases since the criterion of knowledge on the part of the manufacturer adds a negligence twist to the otherwise strict liability based regime. Furthermore, the regulatory impact of liability based on potential defects and potential damages should be carefully evaluated as it could have the negative effect of discouraging manufacturers to share information on potential risks, especially when

---

8 Based on the presentation of Yves Morrier of EASA during the UAS Open and Specific Category Workshop hosted by EASA in Cologne on 5th July 2017.

9 This was also the case with Honeywell and ACSS in Überlingen (Manufacturers).
the risks are considered to be minor, out of fear of claims. In the domain of aviation, this could have catastrophic consequences.

▶ The state of the art defence must be reassessed in light of the very high standard for exoneration in the case of autonomous systems.

The learning and adaptation feedback loops in an autonomous system can lead to changes in its behaviour that may create new risks which, by their nature, cannot be known or foreseen. Thus, in light of the liability for potential defects, in order to rely on the state of the art defence, a manufacturer must prove he could not have known about the risk of product malfunctioning in the future, even with the most advanced level of scientific and technical knowledge. Provided the leading role of the criterion of ‘safety expectations of the public’ in determining whether a product is defective, the defence may be rendered effectively useless to manufacturers of autonomous systems.

▶ The regulatory compliance defence’s role must be reassessed in light of the certification challenges experienced by certification authorities regarding autonomous systems

If manufacturers cannot rely on the regulatory compliance defence for autonomous systems certified for their ‘proper’ behaviour on the basis of testing-based validation, then the accountability of certification bodies and standard-setters for their design choices and verification and validation mechanisms should be made more explicit.

The highly regulated environment and the paramount importance of safety in aviation have had impact on the liability for defective products which reveals certain specifics compared to other domains. This has led some authors to call for the adoption of a special (possibly international) legal instrument for product liability in aviation [8]. However, given the state of international affairs and the difficulty in promoting a new legal instrument in a field as conservative as aviation, this proposal is unlikely to see the light of day anytime soon.

Even if no international product liability regime for aviation could be agreed in the near future, the aviation industry in the EU has a unique opportunity to participate in the drafting of the new rules of product liability in the EU and to state in a loud voice its concerns and propose solutions to issues which otherwise threaten to suffocate its innovation potential.

Acknowledgment

The author would like to thank his colleagues from the SafeDroneWare project for the inspiration and encouragement to write this paper.

REFERENCES


NEW DATA SOURCES TO STUDY AIRPORT COMPETITION

Riccardo Gallotti, Marc Fuster and José J. Ramasco
Instituto de Física Interdisciplinar y Sistemas Complejos IFISC [CSIC-UIB]
07122 Palma de Mallorca, Spain
Email: riccardo@ifisc.uib-csic.es, jramasco@ifisc.uib-csic.es

Abstract—Traditionally, there is a lack of detailed information on passengers’ movements from and to the airports. This is due to the limitations in accuracy and coverage of methodologies like local surveys commonly used to obtain data in this context. As a consequence, managers and policy makers must take decisions based on partial information on passengers’ transport demands. Recent developments and popularization of the use of Information and Communication Technologies (ICT) provide new alternative data-sources allowing for the precise derivation of individual mobility at different spatial scales. This data may pose some challenges in terms of correcting potential biases, but it overcomes many of the traditional methods limitations. Here, we investigate how the availability of ICT data depicts a new comprehensive perspective on door-to-door air transport mobility. We do this by proposing three case studies involving three new sources of data: i) GPS records of taxi pickups; ii) a database of geolocated tweets including 10 million users tracked for two years in Europe; and iii) the travel-times between the user’s home and the alternative airports (provided by Google’s API). By integrating this data into simplified discrete choice models, we exemplify how the description of airport catchment areas can be treated in large cities served by more than one airport. This works illustrates how the air transportation system interacts with other transport modes in the passengers decision process. While passengers can still be described within the classical rational choice paradigm, new models must be developed to include the influence of ground transportation aspects in the passenger’s travel decisions.

I. Introduction

The increasing availability of data offered by the explosion of the Information and Communications Technologies, together with the raise in computational power and methodological tools necessary for their elaboration, enables us to study socio-technical systems with unprecedented detail [1]. This possibility propelled a new wave of studies that touched several aspects of human long-range mobility, like seasonal changes in population distribution [2], migration [3], tourism [4], [5], and including air passenger flows [6], [7]. Long-range airline traffic however, interacts with short-range ground transportation. The spreading of epidemics, for instance, is strongly shaped by the interaction of international passenger flows and urban commuting [8]. For this reason, in the effort of expanding our knowledge on the behavior of air passengers, it becomes important to integrate the currently used models with a better understanding of the impact of ground transportation. Also in this case, the recent years have shown a lot of novel data-informed results based on the analysis of cars [9], [10] or taxis [11] GPS traces, mobile phones [12], micro-blogging [13], location based social networks [14], and public transportation timetables [15], [16]. Data by itself is not sufficient and must be supported by adequate methodological foundations to correctly improve our understanding on the evolution of any system and our possibility of forecasting it [17]. In particular, the economical dimension of transportation must be taken into account [18], [19], [20], [21].
In this paper, we propose three case studies to showcase the potentiality of publicly available data sources in describing the effects of the competitions between airports serving the same urban area. We will use the GPS record of taxi pickups in New York City (NYC), geo-referenced tweets in London and Paris, and the trajectories suggested by the Google Maps API [22] for reaching the airports in the same two cities. Our analysis aims at highlighting the role of ground transportation in the choice between alternative airports. In support of this data-driven perspective, we use rational choice theory [23] to model the decision behavior and to point out which aspects are more relevant for the travellers when they face the option of more departure airports.

II. Results

A. Taxi Pickups

As a consequence of the Freedom of Information Law, in 2013 the New York City Taxi and Limousine Commission shared the content of its database to anyone who requested and agreed to physically go to copy the data available in their facilities [24]. For this case study, we use in particular the dataset released by the University of Illinois at Urbana-Champaign [25], but we remark that since then the New York City Taxi and Limousine Commission simplified the access to this type of data which can be now downloaded directly from their website [26]. We limited our analysis to data of the year 2013, which includes over 173 million taxi trips across all the New York state. For each trip, the pickup and dropoff coordinates, and timestamps are recorded, together with the fare charged including tips, tolls, taxes and the surcharges characteristics of the trips to and from the airports. More details can be found in [25].

NYC is served by three airports: John F. Kennedy International Airport (JFK), LaGuardia Airport (LGA) and Newark Liberty International Airport (EWR). JFK is the main international airport and the one with the largest number of passengers (60 million per year), it also has a good public transport connection by train with the island of Manhattan and with Brooklyn. LGA hosts mostly national flights (24 million passengers per year). LGA is also the closest airport to Manhattan but the public transport connection is only by bus. Newark is both national and international, but is has less passengers than JFK (35 million per year). In the data, we identify the trip from and to the airports as if the pickup or dropoff coordinates fall in a box around the airport. Being EWR not in the New York state, only dropoffs are recorded in that airport. For this reasons, the following analysis is limited to dropoffs.

We use here taxi data to model and test the passenger choice selection between airports using a multinomial logit model [23], [27]. We first select an area of analysis comprising the three airports (see Fig. 1 up left) and divide it in bins of approximately 400m × 400m. For each bin i, we evaluate: i) the fraction $F_i(a)$ of pick-ups having as destination one of the airport a; ii) the average travel-time $t_i(a)$ to the airport; and iii) the average cost $c_i(a)$. The travel time and the monetary cost of the taxi travel to the airport allow us to estimate cost associated to the trip as a combination $C_i(a) = c_i(a) + V_t t_i(a)$, where $V_t$ is a constant value-of-time. The total utility $U$ associated to a trip should in principle include also the generalized utility gained by performing the trip, and the cost associated to the plane ticket. Here, we simplify of the problem by ignoring these two factors, which are known to be relevant in transport analysis and planning, implicitly inducing the naive assumption that all three airports offer similar flights at similar times, with similar quality and costs of the trips, in order to focus on the effect of ground transportation. Under these assumptions, we can model the probability of choosing the airport $a$ from $i$ as:

$$P_i(a) = \frac{\exp(-C_i(a)/k)}{\sum_{i} \exp(-C_i(a)/k)}$$

where $k$ is a free parameter representing uncertainty of information [27]. These probabilities can be compared with the observed fractions $F_i(a)$. By minimizing the
total error \[ \sum_{a \in A} (P(a) - F(a))^2 \], we identify the optimal values for \( k \) and for the value of time \( V_t = 0.35 \) USD/minute \( (R^2 = 0.975) \). These value allow us to reproduce the observed catchment areas for taxi users with surprisingly good precision (see Fig. 1 up right and the flow comparison below).

FIGURE 1 — Real (up left) and estimated (up right) fraction of travellers going toward one of the three NYC airports, represented with an RGB scale (Red: La Guardia LGA, Blue: JFK, Green: Newark NWA) for cells with more than 3 journeys. The majority choice is, therefore, the dominant color. Below, a comparison between model and data through a scatter plot of the fraction of users going from a cell to a each given airport.

B. Geo-referenced tweets

Geo-located tweets have been continuously recorded by querying the Twitter API [28]. The system we implemented allows us to capture a good part of the entire streaming of geo-located tweets [29]. For this work, we filter only the countries where air traffic is handled by the European Civil Aviation Conference (ECAC). For this selection, we find a total of 9.8 Million users observed during the two-years period of analysis considered (2015-2016).

By tracking the movements of the individual (anonymized) users, it is possible to approximately reconstruct their home country as the country where they have been observed for the longer time. This allow us to distinguish between tweets of locals and tourists in the same area. In Fig. 2 (a), we can see, for instance, the distribution of tweets of locals (UK) and tourists of two different nationality (France FR and Spain ES) within the metropolitan area of London.
London is served by six airports: Heathrow (LHR), Gatwick (LGW), City (LCY), Luton (LTN), Stansted (STN), and Southend (SEN). These airports are characterized by a different user base. For studying this difference, we define for each airport a polygon describing its contour [see Fig. 3 (a)] and isolate the tweets performed inside this polygon [see as example the tweets distribution in Heathrow in Fig. 3 (b)]. This permits us to assign a set of users to the airport they tweeted from. It is important to note that we filter out users tweeting frequently in different days from the same airport in order to exclude workers and population living around airports. The first thing we can quantify is what proportion of users observed in the airport are local (residents in the London area) or non-local travelers (having their residence within the rest of the ECAC area) (see Fig. 2 (b)). The airport most used by tourist is Stansted, which is indeed a base for a number of major European low-cost carriers, while the one least used in proportion (and also in total) is the Southend airport.

As one could notice in Fig. 2, locals and tourists are distributed differently in London’s metropolitan area, with the tourist mostly concentrated in the central districts. This is reflected also by the subset of users we observed tweeting from within the airports. For each of these users, we can approximate the position of the home-place (for locals) or the final urban destination (for tourists) as the area from where the user tweets the most. In this approximation, we first divide the area of the analysis [Fig. 2] in a number $N$ of sub-areas commensurate to the total number $n$ of tweets recorded [using the rule $N = \sqrt{n}$], and identify the home/destination as the center of mass of the points in such sub-area. This procedure has of course some limits, but assures that the approximated location is in an area that has been largely visited, which is not true if one uses the alternative option of computing directly the center of mass of the tweets.
The spatial distribution of the final urban destination of the tourist observed in the different airports show very small differences with the exception of the City and Southend airport from where most often the travelers go to ‘The City’ of London. A clearer evidence of a spatial optimization in the choice of the airport can be observed in Fig. 4 for local travelers. In these maps, it appears evident that the residence of the travelers observed in an airport are typically closer to that airport than to others. In the choice between the alternative airports, local travelers are therefore minimizing the travel-time [and cost] between home-place and airport. We will build on this observation in the following section, integrating Twitter data with travel-time information extracted from Google maps in two case studies: London and Paris.
The observed difference in the catchment area of the six airport between can be quantified using the L2-distance between the distribution of Fig. 4. The L2-distance is computed as the sum of the square differences between the value of each cell: $D_{L_2}(p_1,p_2)=\sum_{i,j} (p_1(i,j) - p_2(i,j))^2$, where $p_1$ and $p_2$ are two probability density distribution, and $i$ and $j$ respectively the row and column index of the cell. In Fig. 5, we represent the L2-distances with a color-scale. In the (a) panel, we can observe the distances for locals and in the (b) panel for tourists. We observe that in both cases the two airports with the most peculiar catchment areas are City and Southend, and that for tourists in the (b) panel all differences are less pronounced than for locals in (a).

**FIGURE 4** — Empirical distribution of the approximate home location for the local passengers observed in the six London airports.

**FIGURE 5** — L2 distance between the distribution of home location of local passengers (a) and destination of tourists (b).
We can finally associate to each sector of the area analyzed in Fig. 4 the most common airport used. This permits us to outline the empirical catchment areas of each airport. These are shaped differently for locals and tourists (see Fig. 6). In both cases, the largest and relatively central airport of Heathrow is the dominant option in the center of London. It is noteworthy the effect of low cost companies, which are most used by tourists. Consequently the yellow area representing Stansted is wider for tourist than for local travelers. Conversely, the central and more expensive City airport, in red, is more widely used by locals.

**FIGURE 6 — Empirical catchment areas in London for locals (up) and tourists (bottom). The black line represents the London borough. In the top panels each square represents an area of 10x10 km\(^2\). In the bottom panels we zoom in the two most central sections and each square is of 1x1 km\(^2\).**

---

### C. Google Maps API

The results illustrated in the preceding section show how the choice between alternative airports is dictated by their accessibility, together naturally with the offer of flights and their cost. In this section, we want to investigate how information on the travel-time of ground transportation acquired through the google maps API [22], integrated with spatial information on the distribution of population in a city (Fig. 2) and ticket costs
extracted by the Sabre market intelligence database [30] allows for the modeling the
catchment areas of cities with more than one airport such as London and Paris.

**FIGURE 7 — Transit time to Gatwick airport (black square).** We remark that for many areas in
the centre of London transit (left) is faster than driving (right). In white areas the API does not
provide any possible trajectory for reaching the airport from the centre of the box of 1x1 km.

We divide the area including the six airports of London (see previous section) and the
three airports of Paris: Charles de Gaulle CDG, Orly ORY, and the low cost Beauvais BVA
(situated in the far north) in cells of a square kilometer, each identified by a couple of
indexes for row and columns $i, j$. We approximate the real destinations distribution $\text{Pop}_i$ with Twitter data by associating each user to the cell where he/she tweets the most,
and the travel-times $t_{i,j}(a,m)$ to the airport $a$ and a mode of transport $m$ as the time given
by Google Maps for a trip from the cell to the airport at 8am of a Monday (see Fig. 7 for
an illustration of trips to London Gatwick). The mode of transport we considered in this
analysis are cars (‘driving’ option in Google API), and public transport (‘transit’ option in
Google API). From the Sabre dataset we obtain the average price of a ticket $c_{a,b}$ between
the origin airport $a$ in London or Paris and a destination $b$ in the year 2014. For this study,
we considered a set of $\approx 200$ destinations $b \in B(a)$ within the ECAC area and for which
more than an origin/destination $a$ was available in the city.
Similarly to what proposed for taxi, we use a multinomial logit to model the decision between alternative airports. We define a generalized cost function as 

$$C_{ij}(a, b, m) = c(a, b) + V_{ij}(a, m).$$

This cost function would predict that for the travelers departing from cell \((i, j)\) having final destination \(b\) and using the mode of transport \(m\) for getting to the airport the probability of using airport \(a\) is

$$P_{ij}(a; b, m) = \frac{\exp\left(-\frac{C_{ij}(a, b, m)}{k}\right)}{\sum_j \exp\left(-\frac{C_{ij}(a, b, m)}{k}\right)}$$

where \(k\) is a free parameter. From this, we can estimate the fraction of passengers that would choose to go to the airport \(a\) for reaching their destination \(b\) in the area of analysis as

$$F(a, b, m) = \frac{\sum_{ij} P_{ij}(a; b, m) Pop_{ij}}{\sum_{ij} Pop_{ij}}$$

and compare them with the empirical fractions \(F^*(a, b)\) that can be extracted from Sabre data, from where we can obtain number of passengers \(pass(a, b)\) that have flown from \(a\) to \(b\) in 2014. This approach is naturally based on a series of simplifying assumptions: i) the cost of ground transportation is totally represented by travel-time, ignoring the monetary cost of the trip to the airport; ii) travel-times are set as constant regarding the days and hours of possible departure; iii) the price of the air ticket is constant in time (notice also that our price data does not include promotions); iv) the choice of the
travel destination $b$ and the mode of transport $m$ is independent on where the traveller starts the trip (cell $i, j$); vi) the value of time $V_t$ and the parameter $k$ are constant across the population. vii) we do not consider the option of choosing an alternative destination $b$ (e.g., another airport in the same city) or another mode of transport alternative to the flight.

**FIGURE 9 — Map of the modelled catchment areas in Paris for two possible destinations: Madrid (a), Palma de Mallorca (b). The black lines represent the Paris arrondissements.**

A last strong assumption we are making at this stage is that we propose here two different reconstruction of the catchment areas for the two alternative means of transport $m$ (cars, or public transport) that can be used for reaching the airport. A more parsimonious way of modeling, for instance, would be for instance to separate the population $\text{Pop}_i$ in two subgroups using different transportation means, but this would require further input information not available at the moment, or alternatively to couple the modeling of the decision between the airports with a second modal-split model describing the decision among the available modes of transport. This alternative modeling option would be in our opinion too refined at this point, since the current model is already based on the aforementioned list of very important assumptions and finds most of his strength in its relatively simple interpretation. Therefore, we compare directly the empirical fraction $F^*(a,b) = \frac{\text{pass}(a,b)}{\sum_b \text{pass}(a,b)}$ with the theoretical fractions $F(a, b, m)$ obtained from our models, that explicitly depends upon the mode of transport chosen for reaching the airport $m$.

For each departure city, we obtain the values of $V_t$, $V_T$, and $k$ that better approximate the passenger’s behavior by minimizing the mean square deviation

$$\text{err}(m) = \sum_{a \in A(b), b \in B} \left( F^*(a,b) \right)^2 - \left( F(a,b,m) \right)^2$$

for a set of destination airports $B$ where the set of possible departure airports $A(b)$ includes more than a single airports.

The key factor to interpret, with this model, the passenger’s choice behavior and the differences between cities is the value of time $V_t$. We show here two case studies.
describing the London airports reached via public transportation (Fig. 8) and the Paris airports reached via private transportation. In both cases, the value of time we found is remarkably high: 150 USD/h for London and 190 USD/h for Paris. This high value suggests us that the more central airports offer probably some further advantages that exceeds simple accessibility, such for example a better choice of flighting time. In reality the collectivity of passenger studies would be better characterized by a distribution of value of times: for some passengers money is more an issue than time, while for others, like for instance business flyers, time is more relevant. This variability, neglected by assumption v) is probably a very important aspect that requires further investigation.

The different availability of departure airports and difference in ticket costs induce a particular outline of the airport catchment areas for every destination b. In Fig. 8 (a) and (b) we can appreciate how, as a consequence of the topology of the underlying transit network and of the radial distribution of the airports, the visual representation of the most frequently used airport seems to cut the center of London “as slices of a cake”. More in detail, we see that the area of influence low cost airport of Stansted (purple) for passengers using public transport is expected to expand for trips to the touristic destination of Palma de Mallorca (PMI, panel b) as compared to the state capital of the same country Madrid (MAD). Comparing these with the catchment areas for trips to Zurich (ZRH) displayed in panel (c), we observe how for this in general more expensive destination the role of the central City airport is expected to become more important. In Fig. 8 (a) and (b) we propose two catchment areas for the Paris airports, if reached by car, proposing a similar comparison between an important touristic destination (PMI) and a state capital (MAD) as destination. In this second case, the visual representation of the most frequently used airport represented in Figure 9a shows the two main airports (CDG and ORY) splitting horizontally in half the municipality area of Paris. For the touristic destination PMI (Fig. 9b), the low cost options in ORY expands its area of influence to the whole Paris municipality. The influence of the low-cost airport of Beauvais seems to be very marginal under this perspective. We expect that the same study focused on transit travel-times would have differed from this picture, but we discovered that information on the shuttle service from the center of Paris to the BVA airport is not provided by the Google Maps API, practically reducing the study of the French capital to its main two airports.

III. Conclusion

In this paper, we assessed a series of novel modelling opportunities provided by three new sources of ICT data with application to the description of the catchment areas in large metropolitan areas served by many airports. For these cities, we have shown that the way air transport system interacts with other transport modes can be an important factor in the decision process of air passengers. To include the influence of ground transportation, we modeled the passengers choices with an utility dominated by a cost function that includes the travel-times necessary to reach the airport. Our modeling approach involved series of strong simplifying assumptions in the choice modeling, but it still allows us to illustrate how new types of data can be used to study mobility and travelling behaviour in air transport.

From the three case-studies proposed, we can already reach some preliminary conclusions at a relatively coarse level of granularity. The fact that it has been possible to successfully reconstruct the proportion of taxi passengers going the different NYC airports using only information on the ground transportation, without the need of introducing a different cost associated to each airports, suggests that the advantages and dis-advantages associated to the different airport might be very balanced once one aggregates over all possible destination offered (as it is implicitly done in our study).

Detailed insight on the passenger’s behavior is now available thanks to open-access individual data such as geo-located tweets. For instance, we observe a clear difference between locals and tourists behavior: London citizens more often choose the closest
airports, while this behavior is less remarkable when tracking tourists. However, the quantity of Twitter records available after selecting the passengers observed within the airport is often insufficient for high resolution statistics. For this reason, in our third case study we relied on the Sabre dataset for validation, while Twitter data were used to reconstruct the final destinations within the city.

As extremely rich spatial data such as those offered by mobile phone records [31] are becoming progressively more accessible, the individual trajectory data can successfully be integrated with other rich and high definition sources of data on ground transportation such as public transport timetables [16], open street map [32], or the Google Maps API [22]. Finally, all these sources of mobility and transportation data can be enhanced thanks to more traditional models such as rational theory. Indeed, as we have shown in our last case-study, even within the clear assumption shortcomings of the chosen model, this approach is useful to identify patterns in the data that can inform on the systems’s characteristics. This data assimilation is naturally limited of the assumptions behind the model we chose. For instance, the fact that in our analysis some airports have clearer catchment areas than others is probably generated by the fact that the assumption of similar service offerings is violated: whilst some airports may compete for passengers with a similar service offering, others do not and are unique.

Acknowledgment

The project leading to these results has received funding from the SESAR Joint Undertaking under grant agreement No 699260 under European Union’s Horizon 2020 research and innovation programme.

REFERENCES

UNDERSTANDING DOOR-TO-DOOR TRAVEL TIMES FROM OPPORTUNISTICALLY COLLECTED MOBILE PHONE RECORDS
A Case Study of Spanish Airports

Pedro García-Albertos, Oliva G. Cantú Ros, Ricardo Herranz
Nommon Solutions and Technologies Madrid, Spain
pedro.garcia@nommon.es

Carla Ciruelos
ISDEFE Madrid, Spain cciruelos@isdefe.es

Abstract—A strategic objective of the European transport policy is the so-called 4-hour door-to-door target, according to which, by 2050, 90% of travelers within Europe should be able to complete their journey, door-to-door, within 4 hours. However, information on door-to-door travel times is scarce and difficult to obtain, which makes it difficult to assess the level of accomplishment of this ambitious target. In this paper, we present a methodology for the measurement of door-to-door travel times based on the analysis of opportunistically collected data generated by personal mobile devices. Anonymized mobile phone records are combined with data from the Google Maps Directions API to reconstruct the different legs of the trip and estimate the travel times for the door-to-kerb, kerb-to-gate, gate-to-gate, gate-to-kerb and kerb-to-door segments. The proposed methodology is illustrated through a case study focused on the door-to-door journeys of passengers traveling to Adolfo Suárez Madrid-Barajas airport from the rest of the Spanish airports. We finish by discussing the room for improvement of the proposed approach and outlining future research directions.

I. Introduction

The European Commission’s 2011 White Paper on Transport [1] puts particular emphasis on the need for a multimodal, passenger-centric transport system, able to provide seamless door-to-door travel and facilitate better modal choices. In line with this vision, one of the high-level objectives of the European aviation policy is to ensure that air transport is seamlessly integrated into the European transport network, with the goal of taking passengers and their baggage from door to door predictably and efficiently while enhancing air transport experience and rendering the transport system more resilient against disruptive events [2].

In contrast with this high-level vision, air transport in general, and ATM in particular, have so far, to a large extent, lacked this multimodal perspective, with performance objectives and decision-making processes not always taking into account the ultimate impact on passengers. One of the reasons behind this lack of passenger-centric
indicators is the difficulty to gather accurate and reliable data on passengers’ behavior, needs and choices. Traditional methods, based on observations and surveys, provide rich travel and demographic information, but they also suffer from a number of major shortcomings: they are expensive and time-consuming, they depend on users’ availability and willingness to answer, and data acquisition needs to be planned in advance, which prevents the study of unpredicted events.

The vast amount of spatio-temporal data generated by the use of different types of personal mobile devices in our daily lives (smartphones, public transport smart cards, credit cards, etc.), opens new opportunities to collect rich passenger-focused data and complement traditional data acquisition methods. This paper explores the potential of mobile phone records to extract relevant insights about door-to-door mobility. In particular, we present a methodology to detect the origin and the destination beyond the airport and to measure the experienced travel times during each segment of the trip.

The rest of the paper is organized as follows: Section II introduces the problem of door-to-door mobility characterization and discusses the current approaches for the evaluation of door-to-door travel times; Section III describes the case study selected to develop and evaluate the proposed methodology, which has focused on Spanish domestic flights directed towards the Madrid-Barajas airport; Section IV describes the datasets used in the study and the algorithms developed to integrate and analyze such datasets; Section V presents the main results of the case study; Section VI concludes and discusses future research avenues.

II. The Door-to-Door Problem

The basis for the Door to Door (D2D) concept were settled in the 2020 Vision for European Aeronautics developed by ACARE in 2001 [3], which, in addition to highlighting the need for improving air transport punctuality and reliability of airline timetables, introduces the idea of reducing the time spent at the airport, establishing an objective of no more than 15 minutes in the airport before departure and after arrival for short-haul flights and 30 minutes for long-haul flights. In 2011, a new long-term vision was outlined in the report ‘Flightpath 2050 - Europe’s Vision for Aviation’ [4], which lays out how and where the European research priorities should be set to preserve EU growth and competitiveness worldwide, whilst meeting market needs as well as energy and environmental challenges. One of the high-level goals defined in the Flightpath 2050 report is that, by 2050, "90% of travelers within Europe are able to complete their journey, door-to-door within 4 hours", extending the concept of “time spent in the airports” to a wider and multimodal concept that includes all the stages of the passengers’ travel from their origins to their destinations.

The D2D concept can be split into different segments depicted in Figure 1:

- Door-to-kerb: segment of the trip that involves the access to the airport from the origin location of the passenger.
- Kerb-to-gate: segment of the trip that involves the airport processes: check-in, luggage drop-off, security checks and buffer times.
- Gate-to-gate: segment of the trip that involves the airside processes: from the boarding to the disembarking.
- Gate-to-kerb: segment of the trip that involves the arrival airport processes: passport control (if needed), customs and luggage pick up.
- Kerb-to-door: segment of the trip that involves the exit from the arrival airport to the final destination of the passenger.
One of the problems of this D2D concept is the complexity of measuring the D2D travel time and the duration of each segment, which, in turn, makes it difficult to assess the impact of policies and measures aimed to reduce D2D travel times. Passenger surveys are used to collect information about aspects such as the modes of transport used for airport access and egress, airport catchment areas, waiting times and travel habits, all of them directly related to the D2D concept. Airports (e.g., U.K. CAA Departing Passengers Survey, AENA EMMA surveys), airlines (e.g., Lufthansa Balance Customer Satisfaction Survey), national statistical offices (e.g., International Passenger Survey of the UK Office for National Statistics), governments (e.g., Survey on Tourist Habits conducted by the Spanish Ministry of Tourism) and international organizations (e.g., IATA Satisfaction Survey) collect information about passengers’ behavior and experience, including their travel times. However, surveys are costly and time consuming, which reduces the size of the sample and the frequency with which information is updated, limiting the information to punctual observations rather than continuous monitoring. Additionally, traditional surveys do not allow the collection of data about unexpected events such as extreme weather events, strikes, etc., due to the need for data collection to be planned in advance, as well as to the difficulties to interview passengers under certain types of special circumstances (as an example, on the 20th of August 2008, when the fatal Spanair incident occurred, a passenger survey was coincidentally planned in Madrid-Barajas, but the survey was cancelled due to the circumstances). Other data sources such as air traffic databases, travel reservation systems and market intelligence data services also provide valuable information to characterize the gate-to-gate segment, but they typically fail to capture door-to-door origin-destination pairs and travel times.

This lack of D2D related data has recently led to the launch of several European projects aiming to measure D2D travel times and assess the level of accomplishment of the 4-hour D2D target. Examples of relevant projects in this field are META-CDM, TRIMODE, DORA, Mobility4EU, and BigData4ATM. Some of these projects (e.g., DATASET2050) are trying to develop new modelling techniques to fill the gaps in traditional data collection methods, while other projects (e.g., DORA) are proposing the use of innovative technologies for data acquisition, such as app-based surveys. The BigData4ATM project, in which the present work is framed, adopts a different approach, focusing on the extraction of D2D passenger mobility information from large-scale, passively collected datasets generated by personal mobile devices.

III. Case Study

The goal of the present study is to evaluate the potential of mobile phone records to extract information about passengers’ door-to-door mobility. More specifically, we aim to develop a methodology allowing the identification of the passengers’ origin and final destination and the measurement of the duration of the different D2D trip segments. Taking advantage of the fact that the BigData4ATM project has access to a dataset of anonymized mobile phone records provided by Orange Spain, the data analysis methodology has been developed and tested in the context of a case study focused on Spanish domestic flights arriving in the Madrid-Barajas airport.

Adolfo Suarez Madrid-Barajas is the first Spanish airport in terms of passenger traffic, transport of goods and operations. With 50.4 million passengers in 2016, it occupies the
6th position in the ranking of European airports, and is the largest airport in Europe by physical size along with Paris Charles de Gaulle. To delimit the case study, we focus on the passengers flying to Madrid through direct flights from other Spanish airports in July 2016. The study considers the top 25 Spanish airports, which, according to EUROCONTROL’s Demand Data Repository (DDR) [5], account for around 98% of all the domestic flights arriving to Madrid (Figure 2).

**FIGURE 2 — Airports considered for the case study**

The case study gives us the opportunity to analyze both short-haul and medium-haul flights:

- A typical domestic flight from an airport in the Iberian Peninsula to Madrid takes around 1 hour, similar to many intra-European flights, like Paris-London or Amsterdam-Frankfurt. This kind of connections are the ones with more options to achieve the 4 hours door-to-door objective.

- In order to analyze flights’ of a higher duration, the connections with the Canary Islands are also studied, which allows us to analyze flights with a duration of between 2 hours 30 minutes and 3 hours, similar to the connections between Rome and London or Warsaw and Paris, to mention some examples.

### IV. Data and Methodology

#### A. Datasets

1. **Mobile phone data**

The mobile phone data used for this study consist of a set of anonymized Call Detail Records (CDRs) provided by Orange Spain. Orange is currently the second largest mobile network operator in Spain, with a market share of around 27%. A CDR is a data record produced every time a mobile phone interacts with the network through a voice call, a text message or an Internet data connection. CDRs are stored by the mobile network operator for billing purposes. Each of the records available for this study contains an anonymized identifier of the user together with the time when an interaction with the network occurred (a phone call starts/ends, an Internet data session is opened, etc.) and the cell tower to which the user was connected at that particular moment. The registers do not provide the exact location of the users, but the location of the tower to which they are connected, which typically provides an accuracy of around 100-200 meters in urban environments and up to a few kilometers in rural areas, where the mobile network is less dense. To refine the estimation of the user position inside each of these areas, the cell plan has been integrated with a layer of land use information, which assigns users to different areas in a cell with a probability that depends on the type of land use (residential, commercial, industrial, etc.). As for the temporal granularity
of the data, it varies from user to user, since the number of registers depends on the level of usage of the mobile device. However, for smartphone users, which constitute the vast majority of the sample, registers corresponding to data sessions are usually generated on a periodic basis without any user interaction, e.g. due to apps running in the background. Figure 3 shows the probability density function of the time between two consecutive data sessions registers. Taking into account these data registers and the registers corresponding to calls, SMS, etc., an average smartphone user who has his mobile phone switched on and with a data connection enabled typically produces a register at least every 30 to 60 minutes, which provides a reasonably good resolution for the analysis of the user’s mobility patterns.

FIGURE 3 — Probability density function of the time between two consecutive data registers

2. **Google Maps Directions API**

This API calculates route options and their associated travel times given an origin and a destination. Routes can be obtained for different transport modes, such as walking, private vehicle and public transport. Also, the API accepts waypoints to better characterize the trip. As explained before, mobile phone data do not provide neither a continuous monitoring of the mobile phone user nor exact location information. The Google Maps Directions API can be used to retrieve the different travel options for the different trips (or trip legs) detected for a user, in order to estimate the selected transport mode(s), the chosen route(s), and the associated travel times.

3. **Demand Data Repository (DDR2)**

EUROCONTROL’s DDR2 [4] has been used to obtain the average flight duration and the number of flights for each Spanish domestic route with destination the Adolfo Suarez Madrid-Barajas Airport during July 2016.

**B. Methodology**

1. **Generation of activity diaries for long distance travels**

The first studies that explored the use of CDRs for the analysis of activity and mobility patterns were focused on urban environments ([6], [7], [8]). Although the basic principles of activity detection from mobile phone registers still apply, when analyzing long distance journeys the problem shifts from obtaining an activity diary with as much spatial detail as possible to obtaining an accurate characterization of the semantics of the detected stays allowing the identification of long distance journeys and their decomposition into their constituent legs. For instance, a long stay detected at the airport may be characterized as part of a long distance trip or as an activity on its own (e.g., work activity, in the case of the airport employees), depending on the previous and next positions of the mobile phone. In the present study, a two-step approach has been followed:
First, a diary of stays is obtained for each agent. The condition to identify a stay is that the traveler spends a certain amount of time within a defined radius of a location. The Stay Point Detection Algorithm (SPD) described in [9], originally developed to obtain activity diaries from GPS registers, was adapted to reconstruct long distance trips. The process involves two clustering phases: first, we detect coarse, high-level clusters far away from each other; second, we dig into the details of each high-level cluster to obtain the sequence of activities. When only one clustering phase is used, the stay diaries may not produce accurate results: if thresholds are too permissive, detailed activities may be lost, and information about D2D origin/destination may be blurred; if thresholds are too restrictive, multiple false stops may be detected during trips. This phenomenon is illustrated in Figure 4, where a car trip from Madrid to Barcelona with some intermediate stops is shown. In Figure 4.a, the registers left by the mobile phone user during the trip are shown. Figure 4.b shows the results of a single, coarse clustering process, while Figure 4.c shows the results of a single fine clustering process. It can be seen that neither of both approaches satisfies our needs, as the coarse clusters are not detailed enough while the fine cluster artificially splits the trip. The results for the two-step approach are shown in Figure 4.d, where trip is split at typical stopping places for Madrid-Barcelona trips.

Next, using the Google Maps Directions API and data about the locations of ports, airports and long distance rail stations, the nature of the stays obtained in the previous step is determined. Stays are classified into activities and stops. Activities are those stays that primarily motivate the displacement of the user (e.g., going to work), while stops are stays between the different legs of the same trip (e.g., a stay at the airport previous to a flight, a stop to refuel or have lunch during a car trip, etc.). In the case of plane trips, a very characteristic pattern is observed in the mobile phone registers: the user disappears from the network during a time interval and re-appears at a long distance from the location of the last register before disappearing. Additionally, before and after the interval without registers, the user locations are close to an airport. As a result of the process, the diary of stays is transformed into a diary of activities and intermediate stops, which identifies the origin and destination of each long distance trip and the corresponding intermodal stops, if any (e.g., stops at the airport or high-speed train stations). In the example shown in Figure 4.d, the stays detected in Zaragoza (which is a typical stop, as it is in the middle of the way) and Igualada (around 75 kilometers North-West of Barcelona) are classified as stops.

**FIGURE 4 — Madrid-Barcelona road trip: use of different cluster approaches**

![Figure 4](image)

2. **Identification of plane trips**

Once the diary of activities and stops is obtained for each of the users in the sample, those users that are relevant for the case study are selected. Since the case study
focuses on those passengers that travel from any Spanish airport to the Madrid airport by plane through a direct flight, we select the users whose activity-stop diary includes a long distance trip with two stops at two different Spanish airports, being Madrid-Barajas the destination airport. The following information about each plane trip is obtained:

- **Origin of the trip**, i.e., the activity previous to the trip. It contains information about the location, the start time and the end time of the activity. The end time corresponds to the start time of the door-to-door trip.

- **Intermediate registers associated with the trip from the origin to the departure airport** (door-to-kerb leg). The purpose of storing these registers is to estimate the time of arrival to the airport and explore whether we are able to determine the transport mode and the route chosen to access the origin airport.

- **Stop at the departure airport**: it contains information about the location of the stop (in some cases allowing us to identify the terminal from which the flight departs), the time of arrival to the airport, and the last register at the airport. This last register does not necessarily correspond to the departure time of the flight, as it depends on the mobile phone activity of the user and the moment he/she switches off the phone.

- **Intermediate registers associated with the flight**: for most of the users, there will be no registers during the flight, either because they have switched off their mobile phone or because for most part of the flight there is no mobile phone coverage. However, for some users that do not switch off their phones, a small set of registers appears along the take-off and descent trajectories.

- **Stop at the arrival airport** (Madrid Barajas): similar to the stop at the origin airport, it contains information about the location of the stop and the times of the first and last connections at the airport.

- **Intermediate registers associated with the trip from the arrival airport to the final destination** (kerb-to-door segment). As in the case of the departure airport, the purpose of storing these registers is to estimate the time of departure from the airport and characterize the egress leg.

- **Destination of the trip**, i.e., the activity next to the trip. It contains information about the location, the start time and the end time of the activity. The start time corresponds to the end time of the door-to-door trip.

3. **Corrections**

Due to the characteristics of the mobile phone registers and the way these registers are generated when the user performs a plane trip, some adjustments are still needed such that the extracted information can be used to obtain accurate and relevant results.

- There are some cases where the user switches off the mobile phone some time before the flight departs and/or after the flight lands. This leads to an estimated flight duration that is considerably longer than the actual duration. There are other cases where the user does not switch off the mobile phone during the flight, which may lead to the opposite situation. For both situations, the approach followed is to correct the estimated flight durations by using the actual flight durations extracted from DDR.

- Some cases were detected where the door-to-kerb or kerb-to-door duration was considerably long. These cases were studied in detail. For some of them, the obtained values were consistent with the travel time estimations provided by the Google Maps Directions API. However, for some users disabling their data connections during the night time hours and then reappearing in the network once they have started their trip to the airport, the proposed algorithm erroneously assigns the end time of the activity to the moment the device was switched off,
artificially shortening the activity at the origin and increasing the duration of the trip to the airport. To avoid these errors, in these cases the travel time estimations provided by the Google Maps Directions API were used.

V. Results and Discussion

A. Spatial Distribution of Origins

First, we obtained the origins of all the trips considered in the case study. The heat map for these origins is shown in Figure 5. The main trip generators are the airports of Palma de Mallorca, Barcelona and the Canary Islands. The heat maps represent the catchment areas of each airport for flights directed to Madrid. As expected, the catchment areas of the airports located in both the Balearic and the Canary Islands are significantly bigger than the rest, which has implications for the planning of ground connections with airports. Also, it can be observed that the demand for touristic airports seems to be more spatially spread than for other airports.

FIGURE 5 — Heat map of the D2D trip origins

Next, the distribution of the distance for the door-to-kerb segments was calculated. Figure 6 shows this distribution for all origins and for the specific cases of Barcelona, Mallorca and A Coruña. Some interesting observations regarding airport catchment areas can be derived. For example, it seems that most trips from A Coruña are generated near the city, in a range of around 10km, while in the case of Barcelona this range increases up to 20km. The case of Mallorca is very interesting, as the majority of the trips are originated in the range of 10-20km, while there is an appreciable group of trips generated in the range of 40-50km. This is consistent with what was enunciated before about airports situated in touristic islands having more spread catchment areas.

FIGURE 6 — Door-to-kerb distance distribution
B. Spatial Distribution of Destinations

The heat map of the final destination of travelers with origin at all airports, as well as the heat maps of the final destination of the trips with origin in Barcelona, Mallorca and A Coruña airports are shown in Figure 7. The airport with a higher dispersion in final destinations is Palma de Mallorca, while the final destinations for the Barcelona and A Coruña airports are more concentrated in the Madrid Metropolitan Area. This can be explained by the fact that many of the flights from Palma de Mallorca are return flights of tourists. This reinforces the hypothesis that the door-to-door origins and destinations for touristic trips are much more spread than for other trip purposes. Kerb-to-door distance distribution is shown in Figure 8.

FIGURE 7 — Heat map of the D2D trip destinations: a) all origins; b) trips from A Coruña; c) trips from Barcelona; d) trips from Palma de Mallorca
C. Door to Door Travel Times

Now we analyze door-to-door trips from a temporal perspective, in contrast with the spatial approach taken in the previous section. Figure 9 presents the distribution of the door-to-door travel times for all the possible origins and for the same three airports analyzed before (Barcelona, Mallorca, A Coruña). The figure shows that the vast majority of trips last more than 4 hours door-to-door.

D. Travel Times per Trip Segment

To gain insights into the distribution of door-to-door travel times, we analyze the distribution of the duration for each segment (see Figure 10). By looking at the distribution for the door-to-kerb segment, we can observe that there is not such a clear pattern as in the case of the door-to-kerb distance distribution. Although Mallorca and Barcelona attract passengers from further distances than A Coruña, this is not reflected in terms of travel time, which may be an indicator that both airports are well integrated into the ground transport network.

When looking into the kerb-to-gate segment, differences start to appear: the distribution for Barcelona presents lower stays times at the airport, followed by A Coruña and finally Mallorca. This may be due to both the efficiency of airport processes and the different profiles of passengers. This may also indicate that leisure travelers spend more time at the airport (e.g., due to the check-in process, travelling with children, different perception of the value of time, etc.).
Regarding the gate-to-gate segment, we can observe that A Coruña has the shortest flight time with a time distribution centered between 0.75 and 1.25 hours, followed by Barcelona (1-1.25 hours) and Mallorca (1.25-1.5 hours). This is consistent with the DDR data, where the average duration for the flights is 0.9 hours for A Coruña, 1.2 hours for
Barcelona and 1.33 hours for Mallorca. If we look at the distribution for all the possible origins, we can observe two groups: one centered at around 1.25 hours and the other centered at 2.75 hours, the last one corresponding to the flights from the Canary Islands.

Gate-to-kerb distributions are very different from kerb-to-gate distributions, due to the different nature of the processes of each phase. Passengers from Mallorca present a higher duration of the gate-to-kerb segment: assuming that most of them were tourists coming back to their homes, this is likely to be driven by the waiting at luggage delivery. Passengers from A Coruña spend significantly less time than the rest, which may be indicative of business passengers travelling without luggage.

Finally, in the kerb-to-door time distribution it can be observed that trips from A Coruña present lower travel times. This is consistent with the smaller geographical dispersion previously observed. Although centered at around 0.25-0.5 hours, kerb-to-door durations higher than 2.5 hours are observed for all origins plotted. Flights coming from Mallorca are the most significant of the tail of the distribution, which is again consistent with the higher dispersion for touristic trips described in previous subsection.

By looking at the previous findings from a passenger perspective, several conclusions can be extracted about the observed door-to-door journeys. Passengers from Mallorca can be classified as tourists, with a high dispersion on door-to-door origin and destination locations, which leads to high kerb-to-door travel times but not significantly high door-to-kerb travel times, which suggests a good integration of the Mallorca airport with the ground transport network. These passengers also spend more time at the airport. Passengers from A Coruña are mainly business passengers, with door-to-door origin and destination close to both airports, which leads to low kerb-to-door and door-to-kerb travel times and short stays at the airport. Another fact that can be relevant when analyzing the catchment area for A Coruña airport is that Santiago airport is very close, only 60km far away, thus presumably taking passengers from it. Passengers from Barcelona are probably a mix of tourists and business passengers, presenting intermediate characteristics between Mallorca and A Coruña. Also, kerb-to-gate times in Barcelona are the lowest of the three airports, which may be a sign of efficiency in the airport processes.

VI. Conclusions and Future Directions

The case study presented in this paper shows that mobile phone data can be a useful source of fine grained passenger information. By analyzing the registers produced by mobile phone users, it is possible to obtain valuable insights into door-to-door travel times, which are very difficult to measure by using more conventional methods. Future research directions are outlined below:

- The estimation of short distance travel modes and route choices may be relevant, for example, for demand forecasting models that consider different choices for long distance trips (e.g., different competing airports, competition between air transport and high speed train, etc.) where airport accessibility is a key determinant of travelers decisions. The mobile phone registers generated by the user during his airport access/egress trip can be used to produce information about the transport mode used and the routes followed by airport users, e.g. by using map matching techniques. In the cases where the spatio-temporal resolution provided by mobile phone records is not enough to identify modal/route choices, a data fusion approach could be explored, by merging mobile phone records and Google Maps Directions API data with other sources, such as public transport smart cards.

- In this study, data from DDR was used to obtain an estimation of the average flight duration from each origin airport to the Madrid airport. However, as mobile phone data provides data about individual trips, it may be possible to extract a sample of passengers for each flight. This would add another level of disaggregation to the
door-to-door study, allowing us to extract mobility patterns at different times of the day (e.g., do people increase their buffer time at the airport for early morning trips?). Also, it would be possible to analyze how door-to-door travel times are impacted by air traffic delays.

In the next stages of the BigData4ATM project, the following research questions will be addressed:

- Extend the analysis to other airports. It is reasonable to assume that the results presented here are not directly applicable to other Spanish airports. It is therefore interesting to extend the analysis to other airports in order to analyze how different types of airports interact with the ground transport network.

- Analyze off-peak period. In this paper we have studied the month of July 2016. It is expected that, if the same methodology is applied to another period, the results will vary, not only due to the different availability of flights, but also to the change in the type of passenger, with a lower share of leisure travelers.

- Analyze the impact of disruptions. The methodology presented in this paper allows the assessment of the effect of air traffic disruptions on every segment of the door-to-door trip, which would help us gain a more comprehensive understanding of how door-to-door passenger journeys are affected by delayed flights.

- Evaluate the 4-hour door-to-door target. The proposed methodology will be used to evaluate the level of achievement of the 4-hour-door-to-door target for Spanish domestic flights. It is expected that this will help improve the knowledge about how the air transport network is embedded into the transport network, providing useful inputs for the planning of transport policies, infrastructures and services.

**Acknowledgment**

The project leading to these results has received funding from the SESAR Joint undertaking under grant agreement No 699260 under European Union’s Horizon 2020 research and innovation programme. The authors would like to thank the BigData4ATM project team, as well as the members of the BigData4ATM Advisory Board, for their valuable inputs. We would also like to thank the SESAR JU Project officer, Alessandro Prister, for his continuous and timely support throughout the project.

**REFERENCES**


Conflict detection and resolution
IDENTIFICATION OF SPATIOTEMPORAL INTERDEPENDENCIES AND COMPLEXITY EVOLUTION IN A MULTIPLE AIRCRAFT ENVIRONMENT

Marko Radanovic, Miquel Angel Piera Eroles, Thimjo Koca
Dept. of Telecommunications and Systems Engineering
Autonomous University of Barcelona
Barcelona, Spain
marko.radanovic@uab.cat, miquelangel.piera@uab.cat, thimjo.koca@uab.cat

Christian Eduardo Verdonk Gallego, Francisco Javier Saez Nieto
Centre for Aeronautics, SATM
Cranfield University
Cranfield, Bedfordshire, United Kingdom
ce.verdonk@cranfield.ac.uk, p.saeznieto@cranfield.ac.uk

Abstract—To support future automated transitions among the ATM safety nets, this study elaborates identification of the complex traffic scenarios based on the concept of aerial ecosystems. As an extension of the TCAS operational domain and evolving from the separation management towards collision avoidance layer, the concept has been developed as a stepwise algorithm for identification of cooperative aircraft involved in the safety event – detected conflict, and negotiating their resolution trajectories before the ecosystem deadlock event occurs, in which at least one aircraft stays out of a conflict-free resolution. As a response to this threshold, the paper examines generation of both acceptable and candidate resolution trajectories, with respect to the original aircraft trajectories. The candidate trajectories are generated from a set of tactical waypoints and a return waypoint to the original trajectory. Described methodology has been practically implemented to one ecosystem scenario, characterizing its evolution in terms of the intrinsic complexity. By introducing the heading maneuver changes and delay in the resolution process, the results have shown how the scenario complexity is increasing, especially affected by the states of two aircraft in the initial conflict. Furthermore, it has been demonstrated an evolution in the amount of the acceptable and candidate trajectory solutions, for which the minimum complexity value is satisfied. A goal of the study was to explore the lateral resolutions capacity at certain moments and its timely decrement.

I. Introduction

An increased traffic demand and trajectory deviations due to environmental uncertainties impact on the ATC workload at tactical level [1], [2]. With respect to the closest point of approach (CPA) between two aircraft in conflict, this level is timely framed between two safety thresholds: the mid-term conflict detection (MTCD), that is activated approximately 15 minutes before the aircraft reach the CPA, and the short-term conflict alert (STCA),
Conflict detection and resolution triggered approximately 120 seconds before the CPA. This point is operationally defined as an estimated 4D point at which a distance between two conflicting aircraft reaches a minimum value. The air traffic control (ATC) system provides the separation management (SM) services by guiding one or more aircraft out of their trajectories. If the STCA fails, two conflicting aircraft potentially enter a collision avoidance (CA) layer that is characterized by a non-ATC separation provision, but directives coming from on-board the aircraft [3].

In near-term operations, the ground-based safety nets (STCA) need to work optimally in the future ATM environments. The Airborne Collision Avoidance Systems are globally operable and need to be optimized compatible with existing systems [4]. The Traffic Alert and Collision Avoidance System (TCAS), as an airborne autonomous system, demonstrates an excellent performance in the pairwise and multi-treat encounters, but suffers from a lack of an extended operational logic due to well-reported induced collisions in some complex scenarios [5]. Moreover, the TCAS resolution advisories [RAs] may be inconsistent with the standard ATC procedures [6], and produce a gap in integration of the SM, at the tactical level, and collision avoidance (CA), at the operational level. Therefore, new research lines are required towards development of the collaborative and decentralized SM layer, on which the human behavior and automation will be fully aligned. That anticipates an operational integration of the safety procedures in such a way that any pair of aircraft involved in a conflict, together with the surrounding trajectories of the neighboring aircraft, behave as a stable conflict-free air traffic system. Furthermore, the integration should include the critical information on the feasible resolution trajectories (RTs), proposed throughout development of decision support tools.

Potential incoherence between the SM and the CA could occur due to differences between the ATC directive after STCA, and a TCAS advisory. In many complex situations, the ATC system does not timely provide separation services after STCA that activates a TCAS alert. As a TCAS sense is based on a set of logic advisories, considering only nearby airspace volumes, the advisory is frequently opposite from an ATC directive, which is considered from a larger, sector-based volume. This situation may produce an ambiguity in the pilot-in-command decision process, and provoke a higher severity of the conflict event [7]. Moreover, TCAS advisories sometimes require more demanding manoeuvres for the crew, taking into consideration the flight efficiency aspects [8].

The proposed concept of aerial ecosystems - a tactical air traffic system - presents a new operational framework that intends to solve the time horizon paradigm in a multiple aircraft environment. The principal function is to identify the system causality and decrease a solution complexity at the SM level, not triggering the TCAS alerts for any potential state changes. An ecosystem, as a multi-agent system [9] presents a set of aircraft with the trajectory-amendment and decision-making capability, whose trajectories are identified inside a computed airspace volume - cluster - and are causally involved in a safety event through identification of the spatiotemporal interdependencies (STIs). The STIs present a product of potential avoidance maneuvers among the aircraft involved in the safety event. The ecosystem creation is based on a pairwise conflict detection, computation of the operational airspace volume (clustering), and search for the surrounding traffic (ST) aircraft who might have the STIs with conflicting aircraft. An ST exploration could be done timely, in advance, by applying the proper functional metrics at the certain timestamps, preceding the conflict event. This position should guarantee the coherence between the SM and CA layers and the functionalities before and after the STCA threshold. An important ecosystem objective is a deployment of the negotiation interactions among the ecosystem aircraft for finding the best comprise in the resolution process. From the flight-efficiency aspect, the RTs should be as closest as possible to the reference business trajectories. Furthermore, the multiagent, decision-making process inside the ecosystem should assure a reliable generation of the conflict-free resolutions comparing to the separation actions performed at the strategic level, that must include more intent data for the RT generation.

This paper elaborates the ecosystem identification procedure from an operationally created cluster, and detection of the STIs between each pair of aircraft, belonging to the ecosystem. The identification is performed using a specific set of the parametric values. A comprehensive state space analysis of the detected interdependencies has been used
for a method definition of the RT generation. Then, it is further described a selection of the candidate RTs among each pair of the ecosystem members, and analysis of their acceptance based on a given complexity value. Those are trajectories triggered only in case that agreed resolutions trajectories cannot be obtained before the ecosystem deadlock event is reached. For the flight efficiency purpose as well as coherence with the TCAS function in vertical plane, only the lateral resolutions are considered. Explained methodology has been practically implemented on a real traffic scenario and obtained results have been gathered for a post-analysis and the potential improvements.

In addition to this introductory section the article comprises five other sections. Section II defines the conceptual problem of complex traffic scenarios when severity of the conflict event arises. Section III describes methodology for the ecosystem identification and STI generation for development of the tactical conflict management, while Section IV describes the algorithm for generation of the candidate resolutions before the deadlock event occurs. Section V analyses the simulation results of an ecosystem scenario at a certain complexity level, and compares the pairs of the candidate resolutions at three time stamps during the ecosystem evolution. Concluding remarks and further research notes are provided in Section VI.

II. Problem definition

This section describes a need for introduction of the ecosystem concept and challenges in generation of the RTs, when a time evolution affects an available conflict-free airspace, and results a decrement of the system solutions.

A. Time horizon problem

The transitions from the SM to the CA require a time capacity in which the standard separation minima (SSM) is fully maintained, i.e. $SSM_H = 5$ NM and $SSM_V = 1000$ ft, where $SSM_H$ presents the horizontal separation distance while $SSM_V$ denotes the vertical one. The resolution of a pairwise aircraft encounter in a multi-aircraft environment frequently meets the lack of a maneuvering time for a succeeding conflict event. In this case, the conflict usually evolves into an induced collision, which is a subject to the implementation of different TCAS RAs operable in the vertical, but also in the horizontal plane [10].

To illustrate the concept of an induced collision, it is first considered a simple traffic scenario, with two evolving and one cruising aircraft. Figure 1 illustrates a scenario with three aircraft, namely A/C01, A/C02 and A/C03. A/C01 and A/C02 are flying over trajectories that generate a predicted conflict, while A/C03 presents an ST aircraft.

**FIGURE 1 — Induced collision as a product of previously solved conflict**

A/C01 is in cruising mode while A/C02 starts descending in the opposite direction from A/C01, which assumes a direct approach to A/C01 with a loss of height. On the other hand,
A/C03 is climbing close to identified encounter. As it can be seen, based on TCAS logic [11], [12], the conflict between A/C01 and A/C02 is triggered after activation of the traffic advisories (TA), at the time stamps $t_{TA}^{01}$ and $t_{TA}^{02}$, and then followed by the corresponding RAs, successfully resolved at the time stamps $t_{RA}^{01}$ and $t_{RA}^{02}$. The minimal required vertical separation, ALIM, has been successfully achieved around the CPA.

As the CA layer activates in approximately 60 seconds before the CPA, once resolved conflicts produce very high uncertainty in guidance over the amending RBTs. After its amendment, A/C01 enters a protected zone of A/C03 [13], [14], and generates a new conflict, denoted as an induced conflict. It is characterized by an instantaneous RA alert, while the aircraft is still performing requested resolution maneuver and not resuming to its RBT [15]. In this case, A/C01 was automatically alerted by the succeeding RA at $t_{RA}^{01}$, but also A/C03 with an instantaneous RA at $t_{RA}^{03}$, with an advisory ”Descend”. However, due to insufficient time for the appropriate succeeding maneuvers two aircraft came into the induced collision. Therefore, the ST aircraft might introduce a higher level of uncertainty in geometry of the pairwise encounter.

B. Ecosystem evolution and deadlock event

The key issue in the resolution of an ecosystem is to identify the time limit above which an induced collision could emerge due to a conflict avoidance maneuver. This threshold is called the ecosystem deadlock event (EDE) and depends on the geometric profiles of the RBTs, the flight configuration (cruise, climb, descent) and the encounter dynamics (closure rates). The EDE is computed and triggered by the ATC. It presents a time instant at which at least one ecosystem aircraft cannot perform any feasible maneuver leading to the conflict-free solution. Instead, an induced collision could emerge. The time frame between the ecosystem identification instant and EDE is approved for the resolutions negotiation. This negotiation is implemented by means of the agent technology in which each aircraft is enhanced by an agent that follows the airline business model, used to identify preferred amending maneuver. This technology provides the right framework to support the negotiation between the ecosystem members to reach a resolution consensus avoiding the ATC intervention which does not consider the airline preferences. Therefore, the objective is to explore the STIs among each pair of aircraft and provide an information on the conflict intervals within the time frame, mentioned above.

The ecosystem evolution toward computed EDE is characterized by a continuously decreasing rate in the number of potential resolutions. In other words, the prolongation in the agents’ negotiation forces the aircraft to continuously follow-up their RBTs which negatively affects a total number of the conflict-free trajectory amendments. A time lost in the negotiation is indirectly proportional to the available ecosystem maneuvering space. Figure 2 illustrates the ecosystem evolution over three time windows, TW1, TW2 and TW3, in which each subsequent window is a sub-window of the previous one. TW3 denotes a CA window whose edges present the EDE moment. Aircraft reaching this ecosystem instant on their RBTs are not a subject to the ATC separation provision, but the TCAS activation. Therefore, any agreed (cooperative) maneuvers inside the TW3 will not provide the conflict-free amendments with respect to the SSM.

**FIGURE 2 — Ecosystem evolution towards EDE**
Figure 3 shows a theoretical decreasing rate of the conflict-free solutions $S$ over the ecosystem time. It can be noted a higher drop in the number of solutions that occur until the TW1, and then follow-up with a lower decreasing rate until the TW2. $S$ approaches to zero value when the ecosystem enters the TW3.

**FIGURE 3 — Rate of change in the number of resulting manoeuvres**

As a response to non-agreed resolutions before the EDE appears, the compulsory resolutions must be activated. The computational method continuously searches for them, aiming to eliminate TW3 in any preceding moment.

### III. Methodology for ecosystem identification

This section describes the ecosystem identification algorithm and method for the STI detection. The algorithm relies on the concept hotspot-cluster-ecosystem that foresees four steps [16]:

1. extraction of en-route traffic at tactical level,
2. creation of the ecosystem scenarios from detected pairwise conflicts,
3. clustering of the airspace volumes that comprises a set of the ST aircraft nearby detected conflicts, and
4. identification of those ST aircraft having the STIs with the conflicting aircraft; they become the ecosystem members.

#### A. Ecosystem identification

The following subsection describes Step 4 from the clustered aircraft. As a response to the time horizon problem, the ecosystem considers a longer operational time characterized by an advanced conflict prediction interval, i.e. lookahead time (LAT), in which the conflicting aircraft obtain a set of information on the ST aircraft in nearby airspace, and all together cooperatively interact in a decision-making process.

The aircraft clustering is built on a pairwise conflict using the spatial measures, the horizontal ($H_{cb}$) and the vertical ($V_{cb}$) cluster buffer. By default, $H_{cb}$ is set to 15 NM and $V_{cb}$ to 3000 ft. Within a box-shaped volume and by the filtering procedure of the corresponding traffic data, any waypoint belonging to the ST trajectory identifies a cluster aircraft [16], but also potential ecosystem member. Figures 4 and 5 illustrate one clustering configuration projected in the horizontal and vertical plane, respectively. There is a predicted conflict between aircraft A/C1 and A/C2 with three identified ST aircraft, i.e. ST1, ST2 and ST3. A/C1 and A/C2 are positioned at their conflict detection points from which the cluster volume has been created, in line with adopted spatial constructors.
The ecosystem algorithm determines if a cluster ST aircraft evolves into an ecosystem member for which a loss of the SSM with any of two conflicting aircraft would occur if this aircraft performs a given amending maneuver at any moment during the LAT. Considerably, the ecosystem identification is a spatiotemporal category as the applied maneuver generates conflict intervals with neighboring aircraft [17]. Maneuverability is applied in both the horizontal and the vertical plane and defined with the set of parametric values:

- \( m_1 \): Left heading change with a deflection angle \( \Delta h_{dL} = +30^\circ \);
- \( m_2 \): Right heading change with a deflection angle \( \Delta h_{dR} = -30^\circ \);
- \( m_3 \): Climb at vertical rate \( ROC = +1000 \text{ ft/min} \) and minimal flight path angle \( \gamma_C = +2^\circ \);
- \( m_4 \): Descent at vertical rate \( ROD = -1000 \text{ ft/min} \) and minimal flight path angle \( \gamma_D = -2^\circ \).

The specified values have been used for the testing of identified ecosystems. However, they might be a subject to changes in a further analysis. Figure 6 illustrates an example of the identification procedure where A/C1 and A/C2, being in predicted conflict, identify the ST aircraft, namely A/C3 and A/C4, by applying certain avoidance maneuvers, \( m_3 \) and \( m_4 \).
B. STI detection

The algorithm computes the time windows for each ecosystem member, inside which any potential cooperative or non-cooperative, horizontal or vertical, maneuver could produce a loss of the SSM. Those windows are sub-intervals of the LAT and the number of conflict maneuvers within each window is obtained as per defined time rate (by default, one second) along each RBT. Figure 7 shows an example of the conflict interval generated using left heading change. Conflict interval 1 denotes a period in which A/C1 performing given maneuver generates a continuous conflict with A/C3.

FIGURE 7 — Conflict interval for a single RBT applying $\Delta \text{hd}_{L} = +30^\circ$

The number of STIs ($N_{STI}$) between the pairs of aircraft is obtained using four types of amending maneuvers, explained above, and one additional, $m_0$: RBT follow-up. In this study, therefore, five types of maneuvers are counted for, i.e. $M = 5$. Each interdependency contains one or more conflict intervals, and a total number of the conflict intervals ($I$) must satisfy the following condition:

$$I \leq \frac{N_A(N_A-1)}{2} M^2$$

where $N_A$ denotes the number of ecosystem members, and $M^2$ is a derived property that presents total number of maneuvering combinations applied to one pair of aircraft. An example of the STI structure is presented in TABLE I. It consists of the STI identifier, the combinations of two interdependent flight identifiers, the maneuvering combination and the conflict interval. One STI among one pair of aircraft might generate more conflict intervals due to different maneuvering combinations.

TABLE I — STI structure

<table>
<thead>
<tr>
<th>STI_ID</th>
<th>Interdependent aircraft</th>
<th>Maneuvering combination</th>
<th>Conflict interval [sec]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>$m_3 - m_0$</td>
<td>$t_{11} - t_{12}$</td>
</tr>
<tr>
<td>STI 1</td>
<td>A/C1 – A/C2</td>
<td>$m_1 - m_3$</td>
<td>$t_{11} - t_{12}$</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>$m_0 - m_1$</td>
<td>$t_{11} - t_{12}$</td>
</tr>
<tr>
<td>STI 2</td>
<td>A/C1 – A/C3</td>
<td>$m_3 - m_0$</td>
<td>$t_{11} - t_{12}$</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C3</td>
<td>$m_0 - m_1$</td>
<td>$t_{11} - t_{12}$</td>
</tr>
</tbody>
</table>
IV. Method for generation of the candidate RTs

A complexity of the ecosystem evolution is evaluated based on the decreasing (perishable) rate in number of the candidate RTs over the ecosystem time. A resolution candidate trajectory is defined based on generation of a set of the tactical waypoints (TWPs) and a return waypoint to the RBT.

Those TWPs are calculated from an ellipse-based trajectories scheme, in which the aircraft is placed at one foci (a starting point) and a returning point is allocated to the opposite foci. Thus, the TWPs are placed on the different ellipses generated by fixing a certain amount of delay to be introduced to the flight [Fig.8]. Then, a pair of the candidate trajectories is evaluated one against another by computing the evolution of the intrinsic complexity as defined in [19]. If two candidate trajectories have a complexity value larger than the values analogous to the TCAS TAs, it is rejected.

In addition, if the proposed trajectories result in the separation infringements, they are also rejected. The generation of the RTs is limited to a set of heading changes, including maintaining the RBT. These heading changes vary from -30° to +30° for each aircraft, with steps of 10°. In addition, the delays that could be introduced can go up to 4 minutes, with a 1-minute step. Finally, the number of the available RTs in each timestamp includes those that can be issued at that specific moment, and all available RTs that are computed for the future timestamps until the end of the conflict interval.

V. Simulation and analysis of results

The main data source for simulation and verification of the obtained results was Demand Data Repository 2 (DDR2), developed and maintained by EUROCONTROL. Traffic scenarios are generated using historical data, the selected flight plans (planned 4D trajectories) in the so-called s06 model 1 (m1) data format [20]. In this study, s06 trajectories are considered as RBTs. The following data have been used for testing:

► historical traffic dated on 24/08/2017;
► traffic extraction in the selected period, 08:00 – 09:00 (28800 – 32400 seconds);
► operational environment above FL300.

After analysis of the simulated traffic, an ecosystem scenario with 5 members has been selected [TABLE II]. Four interdependencies have been detected among the ecosystem members, as structured in TABLE III. The time frame for the ecosystem process was slotted between 29159.00 and 29421.29 seconds. For the graphical presentation purpose, these time thresholds are converted in such a way that 29159.00 corresponds to 0 seconds, and 29421.29 to 262.29 seconds.

**TABLE II — Ecosystem trajectories**

<table>
<thead>
<tr>
<th>Flight ID</th>
<th>4D structure of ecosystem trajectories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\phi_1$ [°]</td>
</tr>
<tr>
<td>A/C1</td>
<td>39.0000</td>
</tr>
<tr>
<td>A/C2</td>
<td>39.0462</td>
</tr>
<tr>
<td>A/C3</td>
<td>39.1109</td>
</tr>
<tr>
<td>A/C4</td>
<td>39.7103</td>
</tr>
<tr>
<td>A/C5</td>
<td>38.9277</td>
</tr>
</tbody>
</table>
TABLE III — STI structure

<table>
<thead>
<tr>
<th>STI_ID</th>
<th>Interdependent aircraft</th>
<th>Maneuvering combination</th>
<th>Conflict interval [sec]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>0 – 0</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>0 – 3</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>0 – 4</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td>STI_1</td>
<td>A/C1 – A/C2</td>
<td>2 – 1</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>3 – 0</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>3 – 3</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>3 – 4</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>4 – 0</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>4 – 3</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td></td>
<td>A/C1 – A/C2</td>
<td>4 – 4</td>
<td>0.00 – 188.26</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C3</td>
<td>0 – 0</td>
<td>0.00 – 51.48</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C3</td>
<td>0 – 3</td>
<td>0.00 – 51.48</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C3</td>
<td>0 – 4</td>
<td>0.00 – 51.48</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C3</td>
<td>3 – 0</td>
<td>0.00 – 51.48</td>
</tr>
<tr>
<td>STI_2</td>
<td>A/C2 – A/C3</td>
<td>3 – 3</td>
<td>0.00 – 51.48</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C3</td>
<td>3 – 4</td>
<td>0.00 – 51.48</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C3</td>
<td>4 – 0</td>
<td>0.00 – 51.48</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C3</td>
<td>4 – 3</td>
<td>0.00 – 51.48</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C3</td>
<td>4 – 4</td>
<td>0.00 – 51.48</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C5</td>
<td>2 – 0</td>
<td>120.00 – 130.28</td>
</tr>
<tr>
<td>STI_3</td>
<td>A/C2 – A/C5</td>
<td>2 – 3</td>
<td>120.00 – 130.28</td>
</tr>
<tr>
<td></td>
<td>A/C2 – A/C5</td>
<td>2 – 4</td>
<td>120.00 – 130.28</td>
</tr>
<tr>
<td>STI_4</td>
<td>A/C3 – A/C4</td>
<td>2 – 4</td>
<td>90.00 – 91.01</td>
</tr>
</tbody>
</table>

Figure 9. describes the evolution of the acceptable and the candidate RTs over the LAT. The horizontal axis represents the larger conflict interval given by TABLE III. On the left-hand axis, it is represented the number of acceptable RTs (columns on red) and a total number of the candidate ones that have been generated. The axis is represented on a log10-scale. The complexity for the solution that provides the minimum one is plotted in black, with its values on the right-hand side. If there is not an acceptable resolution trajectory, then the complexity is marked as 10 (maximum acceptable complexity).

FIGURE 9 — Evolution of acceptable and candidate RTs and complexity of the the minimal complexity solution

The situations at relevant timestamps are presented on the following figures [Fig. 10, Fig. 11 and Fig. 12]. Figure 10 represents the situation at the initial timestamp, Figure 11
the situation when 100 seconds passed, and finally, Figure 12 analyzes timestamp after 160 seconds. The chosen set of the RTs have been plotted in each figure, representing the 48-seconds projection when the resolution should be implemented. The aircraft position and trajectories are represented by a stereographic projection where the tangential point is located at the initial point of A/C1.

**FIGURE 10 — Resolutions scenario I: Timestamp 0, lower complexity level**

**FIGURE 11 — Resolutions scenario II: Timestamp 100-seconds, medium complexity level (A/C1 and A/C2)**

**FIGURE 12 — Resolutions scenario III: Timestamp 160-seconds, maximum complexity level (A/C1, A/C2 and A/C3)**

It can be observed how the complexity trend increases from the timestamp of 20 seconds, after a peak that is generated as a result of the initial states of Aircraft 1 and Aircraft 5. It is arguable that the solution provided at $t = 160$ sec would solve the situation, as the RTs have not been generated with realistic models for the aircraft dynamics, so the no-solution timestamp would be even earlier.
It can be observed in which period the aircraft could be allowed to negotiate among themselves for finding a solution compromise, but also how it is mandatory to maintain the possibility for the compulsory resolutions if the ecosystem members do not agree on a set of trajectories. This time instant should be located before the complexity evolution changes its behavior from the linear to the exponential one.

VI. Conclusions

This paper describes the automation-based conflict management process for a smooth transition from SM to CA layer by introducing the ecosystem concept. The goal of developed methodology was to determine the existence of ecosystems at the tactical level in the monitored airspace volume and their complexity levels coming from different traffic scenarios. For this purpose, the stepwise approach has been deployed by identifying the safety events in a high-dense traffic environment, in which the causal analysis could be performed. The exploration of the available and acceptable RTs, using the ellipse-based scheme, has been further elaborated with respect to the SSM, and by introducing some dynamic properties, like the heading changes and delay in the resolution initialization. Smooth transition from the ecosystem membership identification to the acceptable candidate resolutions generation provides very valuable insight of the STI structure and a complexity level at a certain moment in the ecosystem evolution.

The results show how the number of the available RTs perishes over time, for a fixed returning point of the intended trajectory. They also illustrate an exponential evolution of the complexity, due to chosen metric for its evaluation. The projected figures at the relevant timestamps display how the RTs become tighter and more complex. Taking into consideration the certain aircraft maneuverability, tested within computed conflict intervals over the ecosystem LAT, the solutions can be compared on basis of the heading changes and delay propagation, followed by the minimal complexity value. Nevertheless, more results obtained from the study and like those presented in this paper, have demonstrated that solutions not only prevent the separation infringements in the horizontal plane, but also provide the compatible aircraft states with TCAS function in which the TAs would not get triggered.

Further research needs to be carried in more directions: an analysis of the multi-thread conflicts with respect to time to the CPA, a reduction of the computational time and an incorporation of the fine trajectory predictions for the ecosystem detection and resolution algorithms, as well as an extension of the parametric values for more robust STI testing. Moreover, some research efforts will be made toward development of the agents’ negotiation process and prediction of the deadlock instant.

Acknowledgment

This research is partially supported by the H2020 Research and Innovation Programme, the project: “Adaptive self-Governed aerial Ecosystem by Negotiated Traffic” (Grant Agreement No. 699313). Opinions expressed in this article reflect the authors’ views only.

REFERENCES


Conflict detection and resolution


PROBABILISTIC AIRCRAFT
CONFLICT DETECTION AND
RESOLUTION CONSIDERING
WIND UNCERTAINTY

Eulalia Hernández-Romero*, Alfonso Valenzuela† and Damián Rivas‡
Department of Aerospace Engineering
Universidad de Sevilla, 41092 Seville, Spain
Email: *e.hernandez9@us.es, †avalenzuela@us.es, ‡drivas@us.es

Abstract—A probabilistic aircraft conflict detection and resolution method considering
wind uncertainty is proposed in this paper. The wind components are modeled as
random variables, described by a joint probability density function. The case of two
en-route aircraft flying with constant airspeed in the same airspace and flight level with
approaching multi-segment trajectories and affected by the same wind is considered.
The conflict is characterized by the minimum distance between the aircraft and the
probability of conflict. The probabilistic conflict detection is performed using the
Probabilistic Transformation Method. The conflict resolution problem is formulated as
a parametric optimization problem subject to constraints, being the optimality criterium
the minimization of the spatial deviation from the nominal trajectory. Numerical
results are presented for statistically-independent and uniformly distributed constant
winds obtained from Ensemble Weather Forecasts.

I. Introduction

Improving capacity, efficiency, and safety levels of the Air Traffic Management (ATM)
system are main goals of both the Single European Sky ATM Research (SESAR) and
the Next Generation Transportation System (NextGen). The ever increasing levels of
air traffic challenge these goals, making the development and integration of innovative
automated decision support tools a necessity.

An approach that can improve prediction and optimization mechanisms is to model,
analyze, and manage the uncertainty present in ATM. Among the many sources of
uncertainty, the effects of weather uncertainty on the ATM system are of utmost
importance, see Rivas and Vazquez [1]. The limited knowledge about future meteorology
conditions, such as wind velocity and direction, fog, snowfall or storms, is responsible
for much of the delays and flight cancellations, which negatively affects ATM efficiency
and translates to extra costs for airlines and air navigation service providers.

Attending to weather conditions, two types of uncertainty can be identified. Firstly,
hazardous weather which may lead to no-fly zones affecting the ATM at a network level;
and, secondly, some atmospheric properties, particularly the wind, which may affect
aircraft individual trajectories. The inclusion and analysis of weather uncertainty into ATM-
related problems has been addressed by many authors. For example, Zheng and Zhao [2]
developed a statistical model of wind uncertainties and applied it to stochastic trajectory
prediction in the case of straight, level flight trajectories at constant airspeed with different
guidance laws, which affect the distributions of dispersions of the trajectory attributes.
Nilim et al. [3] proposed a dynamic routing strategy for an aircraft that minimizes the
expected delay when the aircraft’s nominal path may be obstructed by bad weather,
obtaining significant improvements when compared with more conservative strategies.
In this paper, a methodology to include the effects of wind uncertainty on the problem of aircraft conflict detection and resolution (CD&R), focused on the cruise phase of the flight, is developed. The CD&R problem under uncertainty has been addressed in the past by several authors. A widely used approach is to consider the aircraft expected position as a random variable and assume it follows a probabilistic distribution: Paielli and Erzberger [4] estimated the probability of conflict for pairs of aircraft whose trajectories were uncertainly predicted, considering the trajectory prediction errors as Gaussian variables; Krozel and Peters [5] presented a non-deterministic conflict detection model for two en-route aircraft, considering their positions, velocities and headings as Gaussian distributions; and Prandini et al. [6] also obtained expressions for the probability of conflict, applying two different models for the tracking errors. Another approach is to propagate the weather uncertainty into the trajectory prediction: Hu et al. [7] and Chaloulos and Lygeros [8] used this approach to study how wind correlation affects conflict probability; Matsuno et al. [9][10] proposed probabilistic aircraft conflict detection and resolution algorithms in the presence of spatially correlated uncertain winds and applied them to the two-dimensional aircraft-aircraft and aircraft-weather conflict resolution problem; and Rodionova et al. [11] developed and evaluated different conflict strategic resolution algorithms for wind-optimal trajectories considering wind uncertainties in the North Atlantic oceanic airspace. This second approach is followed in this paper.

In order to characterize and quantify the uncertainty in a weather forecast, it is convenient to use a probabilistic approach. This paper considers wind uncertainty provided by Ensemble Prediction Systems (EPS), a weather forecasting technique that characterizes and quantifies the uncertainty inherent to the prediction. The EPS has successfully been applied to Air Traffic Management problems in the past. González-Arribas et al. [12] generated wind-optimal cruise trajectories using pseudospectral methods and studied the sensitivity of the optimal flight paths to the numerical weather prediction uncertainty. Steiner et al. [13] presented an approach, focused on convective storms, of how high-resolution ensemble weather forecasts may get integrated with automated ATM decision support tools. Rivas et al. [14] analyzed the effects of wind uncertainty on aircraft fuel consumption in the case of multisegment cruise flight subject to average constant winds.

In this paper, the problem of two aircraft flying with approaching trajectories composed of several cruise segments at constant airspeed, constant altitude and affected by the same uncertain winds is tackled. The conflict detection is based on the Probabilistic Transformation Method (PTM), see for example Hogg and Craig [15]. A first step on the assessment of the impact of wind uncertainty on the problem of conflict detection using this method was presented in Hernández et al. [16]. The conflict resolution problem is formulated as a parametric optimization problem subject to constraints, being the optimality criterion the minimization of the spatial deviation from the nominal trajectory. The deterministic algorithm developed by Valenzuela and Rivas [17] is the base of the presented approach.

II. Ensemble weather forecasting

While deterministic meteorological forecasts have long been used in trajectory prediction and, as of today, are still the standard in ATM applications, a lot of efforts have been made to introduce uncertainty information in trajectory prediction systems. One of today’s most promising trends in probabilistic forecasting is the use of Ensemble Prediction Systems (EPS).

An ensemble forecast comprises multiple runs of a Numerical Weather Prediction (NWP) model, which differ in the initial conditions and/or the physical parametrization of the atmosphere; some ensembles use more than one NWP model [18]. The objective is to generate a sample of possible future states of the weather outcome. An ensemble forecast is a collection of 10 to 50 forecasts (referred to as members). Cheung et al. [19]
review some of them: PEARP, from Météo France, consisting of 35 members; MOGREPS, from the UK Met Office, with 12 members; the European ECMWF, with 51 members; and the multi-model ensemble SUPER, constructed by combining the previous three, forming a 98-member ensemble with the aim of capturing outliers and having a higher degree of confidence in predicting the future atmospheric evolution.

There are two approaches for trajectory prediction subject to uncertainty provided by ensemble weather forecasts:

1. Ensemble trajectory prediction, where a deterministic trajectory predictor is used for each member of the ensemble, leading to an ensemble of trajectories from which probability distributions can be derived; some type of postprocessing is required.

2. Probabilistic trajectory prediction, where probability distributions of meteorological parameters of interest (such as wind) are obtained from the ensemble forecast and evolved using a probabilistic trajectory predictor, leading to probability distributions of trajectory parameters of interest.

Cheung et al. [20] follow the first approach, and Rivas et al. [14] and this paper follow the second one.

The meteorological parameters considered in this work are the meridional (South-North), \( w_x \), and the zonal (West-East), \( w_y \), components of the wind. The approach to obtain the probability distributions of the wind components is as follows. Suppose that the ensemble has \( N \) members, then the first step is to obtain for each wind component and for a given location the \( N \) sample values \( \{w_{x1}, ..., w_{xN}\} \) and \( \{w_{y1}, ..., w_{yN}\} \). Next, one must assume that each wind component follows a particular distribution. Finally, the parameters of the chosen distribution are to be estimated from the sample.

Although the formulation presented in this paper is applicable to any probability distribution, the results shown in Section V are obtained for uniform distributions.

### III. CD&R Problem formulation

#### A. Assumptions

The study presented in this paper is analyzed under the following assumptions (see Fig. 1):

- A North-East reference system fixed to Earth is used.
- Two aircraft, A and B, fly in the same airspace and flight level with approaching trajectories composed of several cruise segments with different courses. Instantaneous turns are assumed for course changes of the aircraft.
- The initial position of the aircraft (\( \rho_{sA} \) and \( \rho_{sB} \)) are certain.
- The airspeeds (\( V_A \) and \( V_B \)) of both aircraft are constant and known.
- The two aircraft are affected by the same constant wind (\( \bar{w} \)). It is described by its meridional and zonal components (\( w_x \) and \( w_y \), respectively) which are uncertain.
- The initial separation between the aircraft is greater than a given horizontal separation requirement \( D \).
The absolute and relative motion of the aircraft are described next.

B. Aircraft motion

1. Absolute motion

Let us assume that the aircraft A and B follow trajectories composed on \( n \) and \( m \) segments, respectively. For each segment of the trajectory (see Fig. 2), the airspeeds and courses are constant and known, so the positions of the aircraft A and B in the segments \( i = 1, \ldots, n \) and \( j = 1, \ldots, m \) at time \( t \), \( \mathbf{s}_{A_i} \) and \( \mathbf{s}_{B_j} \), are given by

\[
\mathbf{s}_{A_i}(t) = \mathbf{P}_{A_{o i}} + \mathbf{V}_{gA}(t-t_{A_{o i}}) \quad t \in [t_{A_{o i}}, t_{A_{e i}}]
\]

\[
\mathbf{s}_{B_j}(t) = \mathbf{P}_{B_{o j}} + \mathbf{V}_{gB}(t-t_{B_{o j}}) \quad t \in [t_{B_{o j}}, t_{B_{e j}}]
\]

where \( \mathbf{P}_{A_{o i}} \) (\( \mathbf{P}_{B_{o j}} \)) is the origin waypoint of the \( i \) (\( j \)) segment; \( \mathbf{V}_{gA} \) (\( \mathbf{V}_{gB} \)) is the aircraft ground speed in this segment; and \( t_{A_{o i}} \) (\( t_{B_{o j}} \)) and \( t_{A_{e i}} \) (\( t_{B_{e j}} \)) are the times on which the aircraft A (B) starts and ends the segment \( i \) (\( j \)), respectively. The time that it takes for an aircraft to fly a segment can be expressed as the length of the segment divided by the ground speed of the aircraft, as follows:

\[
t_{A_{e i}} - t_{A_{o i}} = \frac{\mathbf{P}_{A_{e i}} - \mathbf{P}_{A_{o i}}}{\mathbf{V}_{gA}}
\]

\[
t_{B_{e j}} - t_{B_{o j}} = \frac{\mathbf{P}_{B_{e j}} - \mathbf{P}_{B_{o j}}}{\mathbf{V}_{gB}}
\]
Figure 2 — Absolute motion of aircraft $A$ in segment $i$. 

Taking this into consideration, the times when the aircraft reach the origin and destination waypoints of the segments ($t_{Ao}$, $t_{Ad}$, $t_{Bo}$, and $t_{Bd}$) can be expressed as:

$$t_{Ao} = \sum_{p=1}^{i-1} t_{A_p} + t_{Ao}, \quad t_{Ad} = \sum_{p=1}^{i} t_{A_p},$$

$$t_{Bo} = \sum_{q=1}^{j-1} t_{B_q} + t_{Bo}, \quad t_{Bd} = \sum_{q=1}^{j} t_{B_q}.$$  

If the initial position of the aircraft ( $s_{Ao}$ and $s_{Bo}$ ) are the origin waypoints of the first segments of each trajectory, then $t_{Ao} = s_{Ao} = 0$.

Because the wind is uncertain, the aircraft headings and the magnitudes of the ground speeds in each segment $V_{gA}$ and $V_{gB}$ are also uncertain. They can be obtained from the wind triangle (see Fig. 3) as follows:

$$V_{gA} = \begin{cases} V_A \cos (\psi_A - \arcsin \left( \frac{w_{cA}}{V_A} \right)) + w_x \\ V_A \sin (\psi_A - \arcsin \left( \frac{w_{cA}}{V_A} \right)) + w_y \end{cases}$$

$$V_{gB} = \begin{cases} V_B \cos (\psi_B - \arcsin \left( \frac{w_{cB}}{V_B} \right)) + w_x \\ V_B \sin (\psi_B - \arcsin \left( \frac{w_{cB}}{V_B} \right)) + w_y \end{cases}$$

where $\psi_A$ and $\psi_B$ are the aircraft courses, and $w_{cA}$ and $w_{cB}$ are the crosswinds affecting each aircraft in segments $i$ and $j$, respectively. In these expressions, the crosswinds are considered to be positive if they are from the left wing, and they are given by the following expressions:

$$w_{cA} = w_x \cos \psi_A - w_x \sin \psi_A,$$

$$w_{cB} = w_y \cos \psi_B - w_x \sin \psi_B.$$
Conflict detection and resolution

FIGURE 3 — Wind triangle for aircraft A in segment i.

The aircraft courses and for each segment can be obtained from:

\[ \psi_A = \arctan \left( \frac{y_{Ai} - y_{Ao}}{x_{Ai} - x_{Ao}} \right) \]  

\[ \psi_B = \arctan \left( \frac{y_{Bi} - y_{Bo}}{x_{Bi} - x_{Bo}} \right) \]  

where \( x \) and \( y \) are the coordinates of the origin and destination waypoints of segments \( i \) and \( j \), as depicted in Fig. 2.

2. Relative motion

The indicators defined in this work to characterize the conflict are only dependent of the relative motion between the aircraft, as it will be seen later. The relative position between the aircraft for segments \( i \) and \( j \) can be expressed as (from Eqs. (1) and (2)):

\[ \mathbf{s}_{ij}(t) = \hat{\mathbf{s}}_{ij}(t) = \begin{pmatrix} \hat{x}_{ij} \hat{y}_{ij} \end{pmatrix} + \mathbf{V}_{ij} \mathbf{t}, \quad t \in \left[ t_{A_{ij}}, t_{A_{ij}} \right] \cap \left[ t_{B_{ij}}, t_{B_{ij}} \right] \neq \emptyset, \]  

under the condition \( \left[ t_{A_{ij}}, t_{A_{ij}} \right] \cap \left[ t_{B_{ij}}, t_{B_{ij}} \right] \neq \emptyset \), that is, the aircraft fly the segments \( i \) and \( j \) at the same time. In the previous equation, the relative initial position, \( \hat{\mathbf{s}}_{ij} \), and the relative ground speed, \( \mathbf{V}_{ij} \), for each pair of segments are given by

\[ \hat{\mathbf{s}}_{ij} = \hat{\mathbf{p}}_{B_{ij}} - \hat{\mathbf{p}}_{A_{ij}} - \mathbf{V}_{gB} t_{B_{ij}} + \mathbf{V}_{gA} t_{A_{ij}}, \]  

\[ \mathbf{V}_{ij} = \mathbf{V}_{gB} - \mathbf{V}_{gA}. \]  

From equations (7) and (8), the relative ground speed can be expressed as:

\[ \mathbf{V}_{ij} = \begin{bmatrix} \cos \left( \psi_{ij} - \arcsin \left( \frac{w_{cB}}{V_B} \right) \right) \\ \sin \left( \psi_{ij} - \arcsin \left( \frac{w_{cB}}{V_B} \right) \right) \\ -V_A \cos \left( \psi_{ij} - \arcsin \left( \frac{w_{cA}}{V_A} \right) \right) \\ \sin \left( \psi_{ij} - \arcsin \left( \frac{w_{cA}}{V_A} \right) \right) \end{bmatrix}. \]
Notice that both $\vec{V}_{gij}$ and $\vec{s}_{gij}$ are uncertain due to the wind uncertainty.

C. Conflict detection

A conflict exists between two aircraft if a given set of separation minima is predicted to be violated in the future. In this project it is assumed that both aircraft are flying at the same flight level and that they are approaching, so a conflict will exist if the minimum distance between them, $d_{min}$, is found to be smaller than a given horizontal separation requirement. In this paper, the conflict between the aircraft is characterized by two indicators: the minimum distance between them, $d_{min}$, and the probability of conflict, $P_{conf}$. The probabilistic conflict detection problem is tackled using the Probabilistic Transformation Method (PTM), later described in Section IV. Using this method, it is possible to obtain the probability density function (PDF) of the conflict indicator $f_{d_{min}}$, as well as the value of $P_{conf}$.

The distance between the aircraft A flying in segment $i$ and the aircraft B flying in segment $j$, $d_{ij}(t)$, is the magnitude of the relative position, $d_{ij}(t) = \| \vec{s}_{ij}(t) \|$. It can be expressed as

$$d_{ij}(t) = \sqrt{\delta_{ij}^2 + 2\delta_{ij} \cdot \vec{V}_{ij} \cdot t + \vec{V}_{ij}^2 t^2}.$$  \hspace{1cm} (17)

A loss of separation would take place if $d_{ij}(t)$ were to be smaller than the separation requirement $D$. Because $\delta_{ij}$ and $\vec{V}_{ij}$ are uncertain, so they are $d_{ij}(t)$ and, therefore, the existence of a loss of separation at a given time. The indicators that characterize the conflict are described next:

1. Minimum distance

For each pair of segments $i = 1, ..., n$ and $j = 1, ..., m$, the minimum distance between the aircraft $d_{min,i}$ can be obtained from Eq. (17) as

$$d_{min,i} = \min \{d_{min,i,j}\}.$$  \hspace{1cm} (21)
2. **Probability of conflict**

The probability of existence of a conflict is given by the probability of \( d_{\text{min}} \) being smaller than the separation requirement \( D \):

\[
P_{\text{con}} = P \left[ d_{\text{min}} < D \right]
\]  

(22)

Notice that these two indicators depend on the wind, and therefore are affected by its uncertainty.

D. **Conflict resolution**

The conflict resolution process is designed in such a way that it generates a trajectory for each aircraft that present a probability of conflict smaller than a given threshold \( P_0 \). These resolution trajectories are defined by a new set of waypoints (vectoring). Considering that the nominal paths of the aircraft are the preferred trajectories, the deconflicted trajectories are determined so that they are as close as possible to the originals. The CR process herein presented is based on the deterministic algorithm developed by Valenzuela and Rivas in [17].

The conflict resolution problem is then formulated as a parametric optimization problem subject to inequality constraints:

\[
\begin{align*}
\text{minimize} & \quad F(x) \\
\text{subject to} & \quad c(x) \leq 0
\end{align*}
\]  

(23)

1. **Parameters**

The resolution trajectory is described in terms of a set of parameters \( x \), which correspond to the coordinates of the modifiable waypoints. Considering that the first and last waypoints of each trajectory remain fixed, the total number of modifiable waypoints is \( q = n + m - 4 \).

\[
x = [x_{A,1}, \ldots, x_{A,n-1}, y_{A,1}, \ldots, y_{A,n-1}, x_{B,1}, \ldots, x_{B,m-1}, y_{B,1}, \ldots, y_{B,m-1}]^T.
\]  

(24)

2. **Cost Function**

The optimality criterium in this problem is to minimize the deviation of the resolution trajectories from the nominal trajectories, which is defined as the following objective function:

\[
F(x) = \sum_{k=1}^{q} \left[ (x^k - x^k_n)^2 + (y^k - y^k_n)^2 \right]
\]  

(25)

where \((x^k_n, y^k_n)\) are the nominal coordinates of the modifiable waypoints.

3. **Constraints**

In the problem under consideration, the sole constrain is the condition for the probability of conflict \( P_{\text{con}} \) of being smaller than \( P_0 \). This inequality constraint can be expressed as:

\[
c(x) = P_{\text{con}}(x) - P_0 \leq 0.
\]  

(26)
4. Resolution Strategy

The resolution process has two phases:

1. **Search of a Feasible Starting Point**: In this phase, the objective is to find a starting point \( x_0 \) that satisfies the inequality constraint. This starting point is obtained by solving the following unconstrained optimization problem:

\[
\text{minimize } P_{\text{con}}(x).
\]  

(27)

Taking into account that the minimum of the function \( P_{\text{con}}(x) \) is zero, the constraint (26) is guaranteed to be satisfied by this solution. The nominal trajectory \( x_n \) is chosen as the starting point in this problem.

2. **Cost Function Optimization**: Once an initial feasible point is obtained, the aim of this second phase is to obtain the feasible solution with the smaller cost by solving the parametric optimization problem described in Eq. (23).

Considering that both the cost function and the constraint are non-linear, the optimization problem can be described as a nonlinear programming problem. Each phase of the resolution problem is carried out using MATLAB’s nonlinear programming solvers `fminunc` and `fmincon`, for phase one and two, respectively.

IV. Probabilistic transformation method

The probability density function (PDF) of the conflict indicator \( d_{\text{conf}} \), given a wind speed probability distribution, is obtained using the Probabilistic Transformation Method (PTM). An introduction to the application of this method to the problem of conflict detection under wind uncertainty was presented in Hernández et al. [16].

The basis of this transformation is as follows [see, for example, Ref. [15]]: Let \( u_1 \) and \( u_2 \) be two continuous-type random variables, statistically correlated or independent, having a joint PDF \( f_{u_1,u_2}(u_1,u_2) \), and let \( R \) be the two-dimensional region in the \( u_1,u_2 \)-plane where \( f_{u_1,u_2}(u_1,u_2) > 0 \). Let \( v_1 \) and \( v_2 \) be two random variables, \( v_1 = g_1(u_1,u_2) \) and \( v_2 = g_2(u_1,u_2) \), whose PDFs are to be found. Assuming that \( g_1(u_1,u_2) \) and \( g_2(u_1,u_2) \) define a one-to-one transformation of \( R \) onto a region \( S \) in the \( v_1,v_2 \)-plane, then \( u_1 \) and \( u_2 \) can be expressed in terms of \( v_1 \) and \( v_2 \) as \( u_1 = h_1(v_1,v_2) \) and \( u_2 = h_2(v_1,v_2) \). The joint PDF of \( v_1 \) and \( v_2 \) is then given by

\[
f_{v_1,v_2}(v_1,v_2) = f_{u_1,u_2}(h_1(v_1,v_2),h_2(v_1,v_2))|J(v_1,v_2)| \text{if (}v_1,v_2\text{) } \in S.
\]  

(28)

In this expression, \(|J(v_1,v_2)|\) is the absolute value of the Jacobian determinant

\[
J(v_1,v_2) = \begin{vmatrix}
\frac{\partial h_1(v_1,v_2)}{\partial v_1} & \frac{\partial h_1(v_1,v_2)}{\partial v_2} \\
\frac{\partial h_2(v_1,v_2)}{\partial v_1} & \frac{\partial h_2(v_1,v_2)}{\partial v_2}
\end{vmatrix},
\]  

(29)

which is assumed to be different from zero in \( S \).
In this work, the variables \( u_1 \) and \( u_2 \) are the two wind components \( w_x \) and \( w_y \), respectively; \( v_1 \) is the indicator \( d_{min} \); and \( v_2 \) is an auxiliary variable which, in order to simplify the Jacobian determinant, is chosen to be any of the two initial random variables \( u_1 \) or \( u_2 \) (i.e., the wind components \( w_x \) or \( w_y \)), resulting in \( J = \partial h_2 (v_1, v_2) / \partial v_1 \) or \( J = \partial h_1 (v_1, v_2) / \partial v_2 \), respectively.

The marginal PDF of \( d_{min} \), \( f_{d_{min}} (d_{min}) \), can be obtained from the joint PDF \( f_{d_{min}, v_2} (d_{min}, v_2) \) as follows

\[
f_{d_{min}} (d_{min}) = \int_{-\infty}^{\infty} f_{d_{min}, v_2} (d_{min}, v_2) dv_2.
\]

Once the PDF \( f_{d_{min}} (d_{min}) \) is known, one can compute the mean and the standard deviation of \( d_{min} \):

\[
E[d_{min}] = \int_{-\infty}^{\infty} \rho f_{d_{min}} (\rho) d\rho,
\]

\[
\sigma[d_{min}] = \left[ \int_{-\infty}^{\infty} (\rho - E[d_{min}])^2 f_{d_{min}} (\rho) d\rho \right]^{1/2}.
\]

This procedure is only to be applied when the transformations \( g_1 \) and \( g_2 \) are invertible functions in the domain of \( f_{u_1, u_2} \). However, it may happen that different wind values lead to the same value of the indicator and the auxiliary variable, thus having a more-to-one transformation. When this situation arises, the problem is divided into multiple sets \( S_j \) where the transformation is one-to-one, and the procedure is applied to each domain.

The probability of conflict can be computed from the PDF of \( d_{min} \) obtained with the PTM as

\[
P_{\text{conf}} = \int_{-\infty}^{D} f_{d_{min}} (\rho) d\rho.
\]

V. Results

Some initial results are presented for a particular scenario and arbitrary winds distributed uniformly.

The scenario under consideration presents two aircraft with segmented trajectories approaching to a common navigation point. The nominal trajectories of aircraft A and B are depicted in Fig. 4, and correspond to RNAVs routes UM192 and UN869, respectively. The waypoints latitudes and longitudes are collected in Table I; these coordinates can be consulted in [21] and [22]. The waypoints coordinates have been transformed to a North-East reference system fixed to Earth whose origin is the waypoint BLN, using an azimuthal equidistant projection. The initial position of aircraft A is \( \vec{s}_{0,A} = [-78.74, 66.63] \) NM, which corresponds to waypoint AMR; and the initial position of aircraft B is \( \vec{s}_{0,B} = [74.94, 27.71] \) NM, waypoint NASOS. The horizontal separation requirement D is set to 5 NM (9260 m) and the airspeeds of the aircraft A and B to \( V_A = 240 \) m/s and \( V_B = 230 \) m/s. The probability of conflict threshold set in the CR process is \( \rho_0 = 0.1\% \).
FIGURE 4 — Nominal scenario.

TABLE I — Waypoints designators and nominal coordinates.

<table>
<thead>
<tr>
<th>Aircraft A trajectory</th>
<th>Coordinates</th>
<th>Aircraft B trajectory</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMR</td>
<td>36 49 59.4 N</td>
<td>NASOS</td>
<td>39 23 56.9 N</td>
</tr>
<tr>
<td></td>
<td>002 15 33.9 W</td>
<td></td>
<td>003 01 40.0 W</td>
</tr>
<tr>
<td></td>
<td>37 11 44.5 N</td>
<td></td>
<td>39 00 00.0 N</td>
</tr>
<tr>
<td>AGIDO</td>
<td>002 37 37.0 W</td>
<td>ANZAN</td>
<td>003 13 17.2 W</td>
</tr>
<tr>
<td></td>
<td>37 24 56.3 N</td>
<td></td>
<td>38 09 09.1 N</td>
</tr>
<tr>
<td>ROLAS</td>
<td>002 51 15.7 W</td>
<td>BLN</td>
<td>003 37 30.0 W</td>
</tr>
<tr>
<td></td>
<td>37 34 47.1 N</td>
<td></td>
<td>36 48 51.5 N</td>
</tr>
<tr>
<td>ARPEX</td>
<td>003 01 27.1 W</td>
<td>MGA</td>
<td>004 22 10.5 W</td>
</tr>
<tr>
<td></td>
<td>37 44 03.9 N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAZAS</td>
<td>003 11 06.7 W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLN</td>
<td>38 09 09.1 N</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>003 37 30.0 W</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>39 00 00.0 N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MORAL</td>
<td>003 32 31.8 W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTB</td>
<td>39 46 50.6 N</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>003 27 51.1 W</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The wind components are considered to be statistically independent and distributed as uniform continuous random variables, whose PDFs are given by:

\[
f_{w_i}(w) = \begin{cases} 
\frac{1}{2\sigma_{w_i}} & \text{if } w \in \left[\bar{w}_i - \delta_{w_i}, \bar{w}_i + \delta_{w_i}\right] \\
0 & \text{otherwise,}
\end{cases}
\]

where \(i \in \{x, y\}\). Since the winds are independent, the joint PDF of the wind components is the product of these two PDFs,

\[
f_{w_x, w_y}(w_x, w_y) = f_{w_x}(w_x) f_{w_y}(w_y).
\]

The endpoints of the uniform distributions are chosen to be the maximum and minimum values of the forecasted winds provided by the EPS. In this application, the meteorological uncertainty data is retrieved from the European Centre for Medium-Range Weather Forecasts. In particular, the PEARP weather forecast, composed of 35 members, is considered. The wind available is the one given for a forecast horizon of 0 hours, released at 06:00 on 05-May-2016, for a pressure level of 250 hPa, and for the coordinates 38° North, 2.5° West, in the Southeast of Spain. The minimum and maximum values of \(w_x\) are 11.75 and 28.40 m/s, respectively, and the minimum and maximum values of \(w_y\) are 10.24 and 25.54 m/s, respectively. The corresponding values of the wind distributions are \(\bar{w}_x = 20.08\) m/s, \(\sigma_{w_x} = 8.33\) m/s, \(\bar{w}_y = 17.89\) m/s, and \(\sigma_{w_y} = 7.65\) m/s. Figure 5 depicts the winds probability density functions, along with the normalized bar charts that represent the PEARP weather forecast data.

**FIGURE 5 — Wind probability distributions (solid lines) and PEARP weather forecast normalized data (bars).**

In Figure 6, the PDFs of the minimum distance \(d_{\text{min}}\) before and after the conflict resolution process are presented. The PDF corresponding to the nominal scenario is depicted with a dashed line, and a solid line is used to represent the PDF for the resolution trajectories.

The expected value and standard deviation of \(d_{\text{min}}\) for the nominal trajectories are \(E[d_{\text{min}}] = 7044\) m and \(\sigma[d_{\text{min}}] = 3170\) m. The probability of conflict before the CR is 70.4%, which can be computed as the area under the PDF to the left of the dash-dot vertical line that represent the separation requirement \(D\). After the conflict resolution process, the expected value and standard deviation have changed to 14329 m and 2897 m, respectively. It is noticeable that the expected value has experienced a significant raise and that the dispersion of the indicator is now slightly smaller. The probability of conflict has dropped to \(P_{\text{con}} = 0.1\%\), which correspond to the constraint set in the CR process. As observed in the figure, the nominal and resolution PDFs are very similar in shape, being the main difference between the two of them their position on the x axis.

In Figure 6, the PDFs of the minimum distance \(d_{\text{min}}\) before and after the conflict resolution process are presented. The PDF corresponding to the nominal scenario is depicted with a dashed line, and a solid line is used to represent the PDF for the resolution trajectories.
The resolution trajectories for each aircraft are depicted in Figure 7, where the new trajectories have been depicted with solid lines. The cost of this solution [see Eq. 25] is $F = 6973$ m. The maximum deviation from its nominal trajectory is 4826 m for the aircraft A and 5029 m for aircraft B. This maximum deviation correspond to the waypoints closer to the navigation point BLN. It can be observed that the rest of the modifiable waypoints have not experienced noticeable changes.

**FIGURE 6 — PDF of $d_{\text{min}}$ for nominal (dashed) and resolution (solid) trajectories.**

**FIGURE 7 — Resolution trajectories.**
VI. Conclusions

In this work, a probabilistic method for conflict detection and resolution for en route aircraft under wind uncertainty has been presented. The proposed method allows the characterization of a conflict between two aircraft and it is able to propose new trajectories that lower the probability of conflict to acceptable levels.

The probabilistic conflict detection problem has been tackled using the Probabilistic Transformation Method. This methodology enables the assessment of the probability of conflict and other characteristics of the conflict for a given scenario. This approach is capable of taking as input any type of wind distribution derived from ensemble weather forecast. The minimum distance between the aircraft and the probability of conflict have been chosen as indicators to characterize the conflict.

The conflict resolution problem has been formulated as a parametric optimization problem and it is based on the modification of the nominal waypoints of the aircraft trajectories. With the proposed CR process it is possible to obtain a new set of trajectories with an acceptable value of the conflict probability and with a minimum deviation from the nominal trajectories.

The method has been applied to a particular scenario and some numerical results has been presented. A uniform wind distribution has been considered for the numerical application. The uncertainty information to determine the wind probability density function has been obtained from weather ensemble forecasting.

The consideration of different types of wind probability distributions or correlated wind-fields in which the wind velocity varies with the position is left for future work. The employment of wind information sources different from the EPS could also be of great interest. Next steps in this line of investigation also include the task of considering trajectories with altitude and velocity changes, of potential application to Terminal Maneuvering Areas.

Acknowledgments

The authors gratefully acknowledge the financial support of the Spanish Ministerio de Economía y Competitividad through Grant TRA2014-58413-C2-1-R, co-financed with FEDER funds.

REFERENCES


SIMULATED ANNEALING FOR STRATEGIC TRAFFIC DECONFLICTION BY SUBLIMINAL SPEED CONTROL UNDER WIND UNCERTAINTIES

Valentin Courchelle, Daniel Delahaye
ENAC - Ecole Nationale de l’Aviation Civile - Toulouse, France
Email: valentin.courchelle@gmail.com, delahaye@recherche.enac.fr

Daniel González-Arribas, Manuel Soler
Universidad Carlos III de Madrid
Leganés, Spain
Email: dangonza, masolera@ing.uc3m.es

Abstract—This paper introduces an algorithm that minimises conflicts between aircraft at the strategic level taking into account uncertainties on aircraft position due to errors into wind forecast. The strategy relies on subliminal speed control. Owing to the complexity of this kind of optimisation problem, a simulated annealing metaheuristic approach is employed. A scenario with four hours of traffic overflying the Spanish [structured, continental] airspace has been selected. Traffic has been retrieved from NEST Eurocontrol database with the corresponding wind ensemble probabilistic forecasts from the European Centre for Medium-Range Weather Forecasts. Due to uncertainties and to the little range of speed changing allowed by a subliminal control, it becomes not possible to resolve all conflicts. However, their number can be significantly reduced by slightly modifying flight plan speeds while not touching the selected route by the airspace user.

I. Introduction

An increase in global air traffic is foreseen in the coming decades. According to the International Civil Aviation Organization (ICAO), it will double by 2030 to reach 6 billion passengers. In order to increase airspace capacity and therefore avoid their saturation, the air traffic management (ATM) system needs to be improved. The development of advanced algorithms and tools capable of anticipating conflict detection and resolution are necessary as this would lighten future air traffic controller (ATC) workload. To this end, coping with uncertainty is absolutely paramount. The development of computer aided conflict resolution tools of this type is aligned with the goals and technological solutions within the future ATM system in Europe, built under the umbrella of the Single European Sky ATM Research (SESAR). The strategic (in this context, before departure) conflict resolution strategy seeks to deviate as little as possible aircraft from the original aircraft.
flight plan, minimising the impact of the separation maneuvers on the flight efficiency.

A large number of strategies have been proposed for so-called conflict detection and resolution problems; refer for instance to the non-exhaustive review provided in [1] or more recently [2]. According to its time horizon, conflict detection & resolution algorithm can be classified into tactical (real time algorithms within a sector) and strategic (planning level algorithm within a network).

For the former, the typical approach is to consider different separation maneuver, e.g., velocity changes [3], [4], heading changes [3], [5], or even combined actions on velocity and Flight Level (FL) changes [6]. Each of these used mixed integer optimisation models. Metaheuristics can also be effective, e.g., using ant colony [7] or genetic algorithm [8], both including heading changes. Aiming to provide robustness against uncertainties, some previous work has also considered different probabilistic approaches to the conflict detection and resolution problem at the tactical level, e.g., [9] (using Monte Carlo), [10] (Markov Chain) or other tools [11]. More recently, a two aircraft encounter was solved using wind uncertainties extracted from ensemble probabilistic forecasts [12].

Nevertheless, aircraft conflict resolution is a highly combinatorial problem that cannot be solved using classical optimisation techniques and realistic models when the number of aircraft becomes significant. This is the case when the problem is tackled at the strategic level, which imposes to consider a macro-scaled airspace and deal with thousands of flights. In this so-called strategic deconfliction context, previous work includes for instance [13] (with FL assignment and speed control) or [14] (with real traffic on a day in the European airspace and conflicts solved by heading changes). However, uncertainties are not taken into account, and they greatly affect traffic and thus potential conflicts. Other studies proposed models that consider uncertainties in the European airspace [15] or in the North Atlantic oceanic airspace [16]. Both resolved conflicts following a ground delay strategy and the modification of trajectory's geometrical shape, showing that it is possible to increase airspace capacity under uncertainties. Nevertheless, and to the best of author's understanding, the deconfliction using speed control on a macro-scaled traffic under uncertainties is an unexplored field.

ERASMUS is a related Eurocontrol funded project to study methods and technologies to increase levels of automation in ATM, in particular air traffic control. An important finding of ERASMUS is the so-coined ‘subliminal control’. In this approach, with minor speed control, significant portions of traffic could de-conflicted while ATCOs workload is reduced (since those minor speed modifications are not perceived by ATCOs). Publications related to the ERASMUS project and subliminal control include for instance [17], [18], [19].

In this paper we propose a strategic de-confliction method through subliminal speed regulations. Wind uncertainties are considered to be the unique source of uncertainty. Thus uncertainties on aircraft positions are taken into account, deduced from a real wind ensemble probabilistic forecasts. An application to real traffic into a structured, continental airspace is shown as case study. The number of conflicts is minimized by small speed deviations from that in the flight plan, while leaving the flight plan’s route untouched. Given the large number of planes that can transit into a given airspace, we resorted to a resolution by a metaheuristic approach using simulated annealing.

The next sections of this document are organised as follows: Section II elaborates on probabilistic wind forecast and associated uncertainties. Section III introduces the mathematical modelling. Section IV describes the simulated annealing algorithm to solve the problem. Section V presents the numerical results. Finally, some conclusions and future directions of research are drawn in Section VI.

mds
January 11, 2007
II. Wind uncertainty

Uncertainty of wind fields and convective regions will be derived from Ensemble Prediction Systems (EPS). Ensemble forecasting is a prediction technique that generates a representative sample of the possible future states of the atmosphere. An ensemble forecast is a collection of typically 10 to 50 weather forecasts (referred to as members) with a common valid time, which can be obtained using different Numerical Weather Prediction (NWP) models with varying initial conditions. The spread of solutions can be used as a measure of uncertainty. In this paper we focus on the output data of the global ensemble forecast system MétéoFrance PEARP EPS. Data can be accessed (among others) at the TIGGE dataset by the European Center for Medium-Range Weather Forecasts (ECMWF).

A. MétéoFrance PEARP EPS

The MétéoFrance PEARP (Prévision d’Ensemble ARPège) is the probabilistic form of the MétéoFrance global numerical weather prediction model ARPEGE. The EPS probabilistic forecast has been based on 35 integrations with approximately 10-km resolution in France (60-km at the antipodes) performing forecasts up to 4.5 days with 90 vertical levels.

The MétéoFrance PEARP EPS represents uncertainty in the initial conditions by creating a set of 34 forecasts starting from slightly different states that are closed, but not identical, to our best estimate of the initial state of the atmosphere (the control). Each forecast is based on a model which is close, but not identical, to our best estimate of the model equations, thus representing also the influence of model uncertainties on forecast error.

The divergence, or spread, of the control plus 34 forecasts -35 in total- gives an estimate of the uncertainty of the prediction on that particular day. On some days, the spread might be small implying that the atmosphere is very predictable and users can trust that the reality will fall somewhere in the narrow range of forecasts. On other days or in other areas, the 35 forecasts might diverge considerably after just a few forecast days, indicating that the atmosphere is especially unpredictable. The variable ensemble spread gives users potentially very useful information on the range of uncertainty. Having a quantitative flow-dependent estimate of uncertainty allows users to make better informed weatherrelated decisions.

III. Mathematical modelling

We assume all aircraft to fly at the same flight level. This hypothesis simplifies the implementation of the modelling because the algorithm deals only with two spatial dimensions. Notice however that it overemphasises the number of conflicts. Also, aircraft flying Eastwards are separated from those flying Westwards. Otherwise, face to face conflicts would not be solvable only by speed regulations. This is at the cost of running twice the algorithm: one time on flights Eastwards and the other on flights Westwards.

In the modelling, aircraft are assumed to fly a constant True Air Speed (TAS) profile, to be assigned by the algorithm within the sublimal speed control bounds. However, notice that the motion of the aircraft with respect to Earth is governed by its ground speed, which depends on its TAS and the existing wind according to the following function:

$$v_g = v_a + v_w \cdot \cos(\chi_w - \chi)$$

with $v_g$ the ground speed, $v_a$ the TAS, $v_w$ the wind speed, $\chi_w$ and $\chi$ the wind and aircraft tracks, respectively. Thus, wind uncertainties heavily affect aircraft positions.
A. Uncertainties modelling

In order to insert uncertainties into the model, we use the MétéoFrance PEARP ensemble forecast $E$ downloaded from TIGGE dataset. For each aircraft $a$ and each members $e$, knowing the departure time and the flight plan, we compute the arrival time $T^a_e$. As Figure 1 illustrates, from this set of possible arrival times, different metrics can be obtained, e.g., the mean time and the range of times. Consequently, for each aircraft $a \in A$ we associate a maximum error on arrival time as follows:

$$\Delta T^a = \max \left( \delta T^a_{\text{max}}, \delta T^a_{\text{min}} \right)$$

where

$$\delta T^a_{\text{max}} = \max_{a \in E} \frac{T^a_e - \sum_{a \in E} T^a_e}{34}$$

$$\delta T^a_{\text{min}} = \sum_{a \in E} \frac{T^a_e - \min_{a \in E} T^a_e}{34}$$

**FIGURE 1 — Illustration of aircraft's arrival time error.**

With the maximum error on arrival time, we can extend the protected area around an aircraft over time by considering an additional margin on the separation norm as illustrated in Figure 2. For each aircraft $a \in A$, two fictive positions $a^+$ (in front of $a$) and $a^-$ (behind $a$) delimit the segment of possible positions of the aircraft $a$ regarding uncertainties. If we call $T^a_0$ the departure time and $T^a$ the arrival time, for each time of the flight we can compute a protecting time gap as follows:

$$\forall a \in A, \forall t \in \left[ T^a_0, T^a \right], \Delta T^a(t) = \Delta T^a_0 \times \frac{t - T^a_0}{T^a - T^a_0}$$

Then, $a^+$ is the future position of $a$ at time $t + \Delta T^a(t)$ and $a^-$ is the previous position of $a$ at time $t - \Delta T^a(t)$. Note that for each aircraft $a$ the margin is zero at $T^a_0$ and grows over time to reach its maximum at $T^a$.

**FIGURE 2 — Illustration of the protected area around the aircraft $a$ at time $t$.** If two areas is overlapping, there is a potential conflict.

B. Conflict evaluation

Two types of conflicts can be distinguished. The first typology occurs at the intersection (called node) of two different routes and it will be coined node conflict. The second
typology occurs when two aircraft are in the same portion of route between two nodes (called link) and when one of the two aircraft is catching up the other. This type of conflict will be coined link conflict.

**Link conflict:**
A link conflict can occur at the entry of a link \( l \) (at the first node) and at its exit (at the second node). Let us consider two aircraft \( a \) and \( b \) flying on link \( l \) such that \( a \) is ahead of \( b \). Let \( v_a^g(l) \) and \( v_b^g(l) \) be the ground speeds of aircraft \( a \) and \( b \) on link \( l \), respectively. Two time intervals must be considered:

For the entry link, the first time interval is
\[
\left[ t_{v_a^g(l)}^{a_{in}^e} + \frac{S_0}{v_a^g(l)} \right] \text{ between the time } a^- \text{ is at the entry and the time it is at } S_0 = 5N \text{ M after the entry.}
\]
The second time interval \[ t_{v_b^g(l)}^{a_{in}^e} + \frac{S_0}{v_b^g(l)} \] is the equivalent for \( b^+ \). If these two intervals overlap, which means that \[ t_{v_a^g(l)}^{a_{in}^e} + t_{v_b^g(l)}^{a_{in}^e} + \frac{S_0}{v_a^g(l)} + \frac{S_0}{v_b^g(l)} < 0 \] there is a conflict. See Figure 3.

**FIGURE 3 — Illustration of a conflict at the entry/exit of a link.**

For the exit link, the same reasoning holds but replacing \( in \) by \( out \). To evaluate all conflicts which occur on link \( l \) let us define the following function:

\[
\forall l \in L, \quad \phi_{\lambda}^l(l) = \sum_{(a,b) \in A_{\lambda}^l} \left( t_{v_a^g(l)}^{a_{out}^e} + \frac{S_0}{v_a^g(l)} \right) - \sum_{(a,b) \in A_{\lambda}^l} \left( t_{v_b^g(l)}^{a_{out}^e} + \frac{S_0}{v_b^g(l)} \right)
\]

where \( L \) is the set of links and \( A_{\lambda}^l^l \) is the set of aircraft pairs \( (a, b) \) involved into a conflict at the entry link \( l \), giving that \( a \) flies ahead of \( b \). \( A_{\lambda}^l^l \) is the same for the link exit. By construction this function is positive.
Node conflict

The detection of a node conflict relies on the same principle. Three different cases need to be modelled to cover all configurations. These are illustrated in Figure 4.

For the first two configurations -upper sketches in Figure 4- aircraft \(a\) would be slightly behind or aircraft \(b\) would be slightly ahead. Around node \(n\), both \(a\) and \(b\) don’t follow the same link so their tracks are different. This difference, denoted \(\theta_{ab}^n\), has an impact on the required distance between \(a\)– and \(b\)+, which goes accordingly with their difference in velocities. The required distance is then not 5 NM anymore but \(S(a,b)\), where \(\alpha\) is the ratio between the ground speeds \([a\ over \ b]\). It can be proofed that [20]:

\[
S(a,b) = S_0 \times \frac{\alpha^2 - 2 \cdot \alpha \cdot \cos(\theta_{ab}^n) + 1}{\sin(\theta_{ab}^n)}
\]

As for the third configuration, the distance \(r(a,b)\) is chosen in order to have 5NM between \(a\)– and \(b\)+ when they are both at \(r(a,b)\) from the node. Then this distance has to be:

\[
r(a,b) = S_0 \times \frac{\alpha^2 - 2 \cdot \alpha \cdot \cos(\theta_{ab}^n) + 1}{2 \cdot \cos(\theta_{ab}^n)}
\]

As it is done for a link conflict, we focus, for each way of detection, on the overlapping of the specific intervals. If one of these detection ways reveals an overlap, aircraft \(a\) and \(b\) are in conflict. To evaluate all conflicts which occur on a node \(n\) we define the following function:

\[
\forall n \in N, \quad \phi_n(n) = - \sum_{(a,b) \in A^1_n} t^+_{n} - \frac{S(a,b)}{v^+_g(n)} - t^-_{n} - \sum_{(a,b) \in A^2_n} t^-_{n} - \frac{S(a,b)}{v^-_g(n)} + t^+_{n} - \sum_{(a,b) \in A^3_n} t^+_{n} - \frac{r(a,b)}{v^+_g(n)} - \frac{r(a,b)}{v^-_g(n)} + t^-_{n}
\]

where \(\ell\) is the set of nodes and \(A^1_n\) is the set of pairs \([a, b]\) of aircraft involved into conflicts in the first configuration at node \(n\), where \(a\) reaches \(n\) before \(b\). \(A^2_n\) reads the same but for conflicts in the second configuration. Finally \(A^3_n\) denotes the set conflicts detected in the third configuration. By construction this function is positive.

C. Mathematical modelling setting up

1. State space

The state space is the set of vectors \(X = \{x\}_{i=1,N}^N \in \mathbb{Z}^N\) with dimension \(N\) equal to the number of aircraft considered. Each component \(x_i\) of these vectors corresponds to a variation of TAS applied to the aircraft \(i\). These velocity variations are integers, something operationally consistent -pilots might set autopilot speed with precision 0.01-. Moreover, they can be positive or negative because the pilot can be asked to accelerate or decelerate. Knowing TAS and wind, one can readily get the ground speed used into the conflict evaluation function.
2. **Constraints**

**Variable definition constraint:**

\[ \forall k \in \{1, N\}, x_k \in \mathbb{Z} \]  

(6)

**Subliminal control constraint:**

As explained in Section I, subliminal control requires minor speed changes. A reasonable interval in which speed variations should be located could be \(-6\%\) to \(+3\%\) of the initial speed [4].

\[ \forall i \in \{1, N\}, -0.06 \times v_i \leq 0.03 \times v_i \]  

(7)

3. **Objective**

The aim is to minimise conflicts with the least impact on aircraft performance. We define the function which evaluates conflicts corresponding to the current state \(X\) as follow:

\[ \Phi(X) = \sum_{n \in N} \varphi_X(n) + \sum_{l \in L} \varphi_L(l) \]  

(8)

So the objective function is:

\[ \min f = M \times N \times \Phi(X) + \sum_{i=1}^{N} |x_i| \]  

(9)

where \(M\) is a coefficient used to weight the minimisation of conflicts w.r.t speed changes. The multiplication by the number of aircraft \(N\) plays an analogous role.

4. **Problem Statement**

All in all, the problem is stated as follows:

**Objective function:**

\[ \min f = M \times N \times \Phi(X) + \sum_{i=1}^{N} |x_i| \]

**Subject to:**

\[ \forall i \in \{1, N\}, -0.06 \times v_i \leq x_i \leq 0.03 \times v_i \]

Where:

\[ \Phi(X) = \sum_{n \in N} \varphi_X(n) + \sum_{l \in L} \varphi_L(l) \]  

(4)

\[ \forall l \in L, \varphi_L(l) \leftarrow \]  

(4)

\[ \forall n \in N, \varphi_X(n) \leftarrow \]  

(5)

\[ \forall k \in \{1, N\}, x_k \in \mathbb{Z} \]

D. **Complexity**

For a given flight plan, we can compute the associated time windows -with the uncertainty margins- for any given point in the route. Potential conflicts between two aircraft will be then detected. The relationship "is in conflict with", or "is in potential conflict with", ...
defines an equivalence relation coined “cluster”. As described in [21] “if we restrict ourselves to the horizontal plane with n airplanes, we can find the presence of \( \frac{n(n-1)}{2} \) potential conflicts”. It can be shown [22] that the set of permissible solutions contains \( 2^{\frac{n(n-1)}{2}} \) connected components, which implies that it requires as many executions of the search algorithm for a local search optimisation. Thus, for a cluster with 6 aircraft, this represents 32,768 related components. The presence of as many components without knowing which one contains the optimal solution make the problem highly combinatorial. That is the reason behind conflict resolution problems being hard optimisation problems. Metaheuristic are possibly more suitable.

IV. Simulated annealing (SA)

A. General description of simulated annealing

SA is a metaheuristic inspired by the annealing process in metallurgy. It consists in bringing the system, from a disordered random state, to a global-minimum energy state, involving heating process and cooling process. A global parameter, temperature \( T \) is applied to control these two processes. The objective function is analogical to the internal energy of the physical problem. SA compares the neighbouring state to its current state and moves from one to another probabilistically. When \( T \) is high, deteriorated solutions (with high energy) are more likely to be accepted. When \( T \) decreases, better solutions are found. At last, a state considered to be good enough is reached. SA is well known for its ability to trap out of the local minimum by allowing random neighbourhood changes. Moreover, it can be easily adapted to various kinds of problems with continue or discrete space states.

B. Adaptation of SA for our problem

In order to adapt the SA algorithm to our problem, several parameters and functions need to be considered.

1. Neighbourhood function:

A neighbourhood function is used to generate a local change from the current solution. Two criteria should be considered: the computational time should be low and the change should remain local, so as to avoid this change to resemble to a pure random search. The neighbourhood generation function is described in seudo code 1.

The fact that the neighbourhood choice is based on the conflict number count increases the likelihood that a flight involving many conflicts will be chosen. Moreover, such a neighbourhood function may preserve weak solutions, which in turn may include some components that could be useful later in the annealing process.

2. Initial temperature and acceptance probabilities:

The temperature parameter, \( T(k) \) - at iteration \( k \) of the SA - is used to control the acceptance of a solution’s degradation. If at step \( k \), \( T(k) \) is high, then all the neighbourhoods have almost the same probability to be accepted and large degradation are more likely to be produced. To the limit, when \( T(k) \) approaches infinity, all neighbours are systematically accepted. On the contrary, if \( T(k) \) is low, a movement that degrades the solution is unlikely to be kept. The slower the rate of temperature decrease, the better the chances of finding an optimal solution, but the larger the total number of SA iterations (thereby increasing the computational time). In order to determine the initial temperature, we
evaluate a temperature which can bring an acceptance rate of 80%. This evaluating method is described by the HeatUpLoop procedure of Algorithm 2.

FIGURE 5 — Simulated annealing algorithm.

3. Cooling loop

Among the different methods to decrease the temperature, we decide to use the geometric law which is a classical method for the SA.

\[ T_{t+1} = T_t \times \alpha, \quad 0 < \alpha < 1 \]

At each iteration, we get the new temperature via multiplying by a predefined coefficient \( \alpha \). The choice of \( \alpha \) is delicate because if \( \alpha \) is too large, the temperature decreases very slowly and the convergence toward the optimum is likely to be too long. However, if \( \alpha \) is chosen too small, the temperature decreases fast and the algorithm risks to be quickly blocked at a local optimum. That’s why this parameter has to be adapted to a problem. The precise cooling process is described by the CoolingLoop procedure of Algorithm 2.
Algorithm 1 Neighbourhood function

Require: the flight conflict count set conflictCount to record the sum of number of conflicts for a subset of aircraft

1: procedure GENERATE_NEIGHBOUR
2: Generate a random number p between 0 and 1;
3: Calculate the total number of conflicts, sumConf in the flight set
4: if sumConf > 0 then
5: target ← sumConf × p;
6: sum ← 0;
7: while sum < target do;
8: i ← iStart \( \triangleright \) iStart is the beginning index of flight set
9: sum ← sum + conflictCount[i];
10: i ← i + 1;
11: end while
12: else
13: i ← random number between iStart and jEnd; \( \triangleright \) jEnd is the ending index of active flight set
14: end if
15: Save the current decision variables;
16: Change the decision variable of flight i i.e. the speed change;
17: Update the flight set information;
18: end procedure

4. Stopping criterion

The termination criterion is set to be the final temperature reaching value \( T_{\text{init}} \times \varepsilon \), where \( \varepsilon \) is a predefined coefficient, and \( T_{\text{init}} \) is the initial temperature for cooling process. We set \( \varepsilon \) based on tests.
Algorithm 2 Simulated Annealing

Require: initial temperature $T$, number of transitions nbTransitions

1. procedure HEATLOOP

2: while $x_0 \times 0.8$ do $\triangleright$ the accepted rate is 0.8
3: acceptCount $\leftarrow 0$
4: $T \leftarrow T \times 1.1$ $\triangleright$ heat up
5: for $i = 0$ to nbTransitions do
6: initState($\hat{x}_i$);
7: CriterionCalculation $y_i = f(\hat{x}_i)$;
8: $\hat{x}_j = \text{generateNeighbour}(\hat{x}_i)$;
9: CriterionCalculation $y_j = f(\hat{x}_j)$;
10: if accept($y_i, y_j, T$, minimisation) then
11: acceptCount++;
12: end if
13: end for
14: $\chi_0 = \text{acceptCount/nbTransitions}$;
15: end while
16: $T_{\text{init}} = T$;
17: return $T_{\text{init}}$
18: end procedure

19. procedure COOLINGLOOP($T_{\text{init}}$)

20: $\alpha \leftarrow 0.95$; $\triangleright$ geometrical law
21: initState($\hat{x}_i$);
22: CriterionCalculation $y_i = f(\hat{x}_i)$;
23: $T = T_{\text{init}}$;
24: while $T > \varepsilon \times T_{\text{init}}$ do $\triangleright$ $\varepsilon$ defines ending temp.
25: for $i = 0$ to nbTransitions do
26: $\hat{x}_j = \text{generateNeighbour}(\hat{x}_i)$;
27: CriterionCalculation $y_j = f(\hat{x}_j)$;
28: if accept($y_i, y_j, T$, minimisation) then
29: $x_j = \hat{x}_j$;
30: $y_j = y_i$;
31: end if
32: end for
33: $T = T \times \alpha$;
34: end while
35: end procedure
V. Results

The proposed algorithm is implemented in Python and simulated on an Intel Core i5 2.4 GHz processor with 8 GB RAM. The data set (downloaded from the NEST Eurocontrol database) corresponds to air-traffic over Spanish airspace on 26th July 2016 between 12. am and 4. pm. Figure 6 shows the resulting 1060 flights together with the computed wind uncertainties according to the associated MétéoFrance PEARP EPS forecast.

**FIGURE 6 — Visualisation of the traffic (red) considered in Spanish airspace (green) under wind uncertainties — blues—.**

The simulated annealing parameters used are:

- Number of transitions: 200
- Geometric law coefficient \( \alpha = 0.96 \)
- Stopping criterion coefficient \( \varepsilon = 10^{-4} \)

These parameters result as a trade off between the objective value and computing time (around a half hour). Indeed the aim of a metaheuristic approach isn’t to find the optimal solution but a satisfying one in a short computing time.

Results are presented in tables I and II.

**TABLE I — Results for the flights to West with and without uncertainties: 523 flights**

<table>
<thead>
<tr>
<th></th>
<th>Without Unc.</th>
<th>With Unc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{c} )</td>
<td>1407</td>
<td>2496</td>
</tr>
<tr>
<td>( \hat{c}^* )</td>
<td>300</td>
<td>604</td>
</tr>
<tr>
<td>( \hat{p} )</td>
<td>78.7%</td>
<td>75.8%</td>
</tr>
<tr>
<td>( c )</td>
<td>312</td>
<td>427</td>
</tr>
<tr>
<td>( c^* )</td>
<td>116</td>
<td>224</td>
</tr>
<tr>
<td>( p )</td>
<td>62.8%</td>
<td>47.5%</td>
</tr>
<tr>
<td>Computing time</td>
<td>1458 s</td>
<td>1493 s</td>
</tr>
</tbody>
</table>
TABLE II — Results for the flights to East with and without uncertainties: 537 flights

<table>
<thead>
<tr>
<th></th>
<th>Without Unc.</th>
<th>With Unc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{c}$</td>
<td>1239</td>
<td>2405</td>
</tr>
<tr>
<td>$\hat{c}^*$</td>
<td>211</td>
<td>469</td>
</tr>
<tr>
<td>$\hat{p}$</td>
<td>83.0%</td>
<td>80.5%</td>
</tr>
<tr>
<td>$c$</td>
<td>289</td>
<td>457</td>
</tr>
<tr>
<td>$c^*$</td>
<td>81</td>
<td>198</td>
</tr>
<tr>
<td>$\rho$</td>
<td>72.0%</td>
<td>56.7%</td>
</tr>
<tr>
<td>Computing time</td>
<td>1816 s</td>
<td>1960 s</td>
</tr>
</tbody>
</table>

In the tables, $\hat{c}$ represents the virtual conflict number, which is used by the algorithm to put more weight to a certain aircraft regulation than another. It is a “virtual” count because a conflict can be counted several times if it occurs on several nodes or links. At the opposite, $c$ represents the number of aircraft pairs involved into a conflict so the “real” conflict count. Both counts are computed for initial flight schedules and after the resolution (represented by the symbol *). Finally, the parameter $\rho$ corresponds to the percentage of resolved conflicts.

The virtual conflict number gives a clear idea of the algorithm performances. In fact we can note that the algorithm reduces this number at least by 70% (perceptible in Figure 8), but it never solves all conflicts because of the short maneuver range in speed change that a subliminal control allows.

To illustrate the effects of the annealing parameters we did simulations with other settings making the algorithm exploring a larger part of the solution space but requiring a longer computing time (around two hours). Then with a coefficient $\alpha$ equal to 0.98, the temperature decrease is slower. Moreover with a number of transitions equal to 400, the algorithm evaluate twice more states $X$ between two temperature changes than during the previous simulations. The table III shows the results for the simulation of each direction considering uncertainties. We can see that a only speed regulation is able to solve more the half of real conflicts considering uncertainties. Moreover, what we have to keep in mind is that if we take into account the initial vertical separation, the number of conflict, real and virtual, would be lower and maybe the algorithm would succeed to resolve them all. Of course a computing time higher than two hours for a resolution of only four hours of traffic is not acceptable but we can think that optimal parameters exist which could bring similar performances to the algorithm for an acceptable computing time. Quite evidently, all this work shows that conflict resolution through speed regulation could offer a significant help for controllers but it will still need their monitoring because the total deconfliction is not guaranteed.
FIGURE 7 — Visualisation of a link conflict resolution (top) and a node conflict one (bottom): an aircraft in green is accelerated and in red it is decelerated. Aircraft in black represent their positions without resolution. Capture from KML file readable by Google Earth

TABLE III — Results for a longer simulation and for each direction under uncertainties

<table>
<thead>
<tr>
<th>Direction</th>
<th>East</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{c}$</td>
<td>2405</td>
<td>2496</td>
</tr>
<tr>
<td>$c^*$</td>
<td>432</td>
<td>507</td>
</tr>
<tr>
<td>$\hat{p}$</td>
<td>82.0%</td>
<td>79.7%</td>
</tr>
<tr>
<td>$c$</td>
<td>457</td>
<td>427</td>
</tr>
<tr>
<td>$c^*$</td>
<td>182</td>
<td>199</td>
</tr>
<tr>
<td>$p$</td>
<td>60.2%</td>
<td>53.4%</td>
</tr>
<tr>
<td>Computing time</td>
<td>7233 s</td>
<td>6615 s</td>
</tr>
</tbody>
</table>

FIGURE 8 — Visualisation of the conflicts (red) for the whole traffic before and after resolution: with uncertainties (bottom) and without uncertainties (top)
VI. Conclusion

We proposed a formulation for deconfliction based on speed regulation, where conflicts should be reduced or ideally avoided without any spatial change in aircraft trajectories. In this modelling, existing conflicts are evaluated and then aircraft True Air Speeds are changed in order to minimise them. Wind uncertainties are included in the modeling. By hypothesis that all aircraft fly at the same flight level, we simplified the modelling but we also increased interactions between aircraft. We solved the problem by simulated annealing with promising results: Around 55% - 80% not considering uncertainty, of the total number of conflicts could be reduced by simply slightly modifying the flight plan speeds. This study can be carried on by the implementation of the third spatial dimension and an other separation maneuver types such as Heading or FL changes, or by delays on departure times.

REFERENCES


COORDINATED CAPACITY AND DEMAND MANAGEMENT IN A REDESIGNED ATM VALUE CHAIN

Strategic Network Capacity Planning under Demand Uncertainty

Stefano Starita, Arne Strauss
Operational Research and Management Sciences
University of Warwick
Coventry, UK
{stefano.starita; arne.strauss}@wbs.ac.uk

Radosav Jovanović, Nikola Ivanov, Goran Pavlović, Obrad Babić
Division of Airports and Air Traffic Safety University of Belgrade - Faculty of Transport and Traffic Engineering Belgrade, Serbia
{r.jovanovic; n.ivanov; g.pavlovic; o.babic}@sf.bg.ac.rs

Frank Fichert
Department of Tourism and Transportation Worms University of Applied Sciences
Worms, Germany
fichert@hs-worms.de

Abstract—We present a model and a numerical example to analyse capacity decisions within a re-designed Air Traffic Management (ATM) value-chain. We assume a new role for the Network Manager (NM), having contractual relations with Air Navigation Service Providers (ANSPs) and Aircraft Operators (AOs). The NM orders en-route airspace capacity from ANSPs at strategic level and defines/adjusts sector opening schemes at pre-tactical level (capacity management). On the demand side, the NM offers trajectory products to AOs, which are defined based on both AOs’ business/operational needs and network performance goals (demand management). In this context, we develop a mathematical model which underlines a part of this joint capacity and demand management process. The model aims at minimizing the sum of cost of capacity provision and cost of delays and re-routings, by managing airspace sector configuration over time and trajectory assignments. We use a realistic numerical study on a small-scale network to illustrate joint capacity and demand management decisions, as well as trade-offs between different performance indicators.

Foreword—This work is envisaged as a part of SESAR 2020 Exploratory Research project “Coordinated capacity ordering and trajectory pricing for better-performing ATM” (COCTA). Opinions expressed in this work reflect the authors’ views only.

I. Introduction

In previous papers, we outlined a new concept of coordinated capacity ordering and trajectory pricing, referred to as COCTA [1], as well as an initial COCTA mathematical model [2]. In this concept, the Network Manager (NM) decides how much capacity units

This project has received funding from the SESAR Joint Undertaking under grant agreement No 699326 under European Union’s Horizon 2020 research and innovation programme.
(we use sector hours) to order from each Air Navigation Service Provider (ANSP). On the demand side, the NM offers different trajectory products to Aircraft Operators (AO). Capacity and demand management (referred to as COCTA mechanism) are jointly performed by the NM to optimize a vector of network performance indicators.

In this paper, we present the redesigned ATM value-chain, roles and the institutional relations between the NM, ANSPs and AOs. We outline the timeline of capacity ordering: long-term (5 years) and strategic (6 months) capacity orders, as well as pre-tactical (7 days) decisions on sector opening schemes (sector configurations). The COCTA demand management elements (different trajectory products, trajectory prices and the underlying airport-pair charging principle) are not in the focus of this paper.

We outline a basic mathematical model that underpins a part of this coordinated capacity and demand management process. The model’s objective is to minimize the overall cost imposed on AOs: cost of capacity provision and cost of delays and re-routings. Although the main goal is to improve cost-efficiency, we are also able to identify trade-offs between several performance indicators. A case study is used for testing and evaluating the model: it is large enough to allow interpretation and sense-checking of the results, and the traffic values are realistic representations of a part of the European network.

The remainder of the paper is structured as follows: Section II illustrates the COCTA concept, before we focus on the mathematical model in Section III. We present numerical results in Section IV and draw conclusions in Section V.

II. COCTA concept

COCTA is the first research project to consider coordinated en-route capacity and air traffic demand management decisions. A brief summary of relevant previous research efforts is provided in [2], and a more detailed capacity and demand management literature review is presented in [3].

A. Redesigned ATM value-chain

COCTA introduces substantial changes in the ATM value-chain [4] by mandating a new role for a network manager. The network manager:

- orders and allocates airspace capacities from ANSPs by applying demand driven capacity management and

- manages demand by defining and offering different trajectory products at differentiated prices to AOs, mindful both of AOs’ business/operational needs and required network performance levels.

These changes in institutional settings are necessary for a paradigm shift, introducing network-centric capacity ordering and allocation and departing from the traditional airspace-use charging to novel airport-pair (route) charging and trajectory pricing.

In this paper we do not focus on the demand side elements of COCTA. However, for the modelling we assume charges based on airport pairs. This assumption ensures that airlines do not have an incentive to deviate from the shortest route between two airports (for details, the reader is referred to [4]).

B. COCTA capacity and demand management process

The COCTA mechanism represents capacity and demand management measures in the COCTA process of optimizing network performance. Within the COCTA research, the
mechanism is primarily designed for strategic (6 months in advance) and pre-tactical stage (7 days in advance), while the tactical stage is considered to a certain extent only. In addition, we also discuss long-term capacity planning (5 years) and ordering to improve the network performance.

1. **Capacity management**

Capacity management is carried out at the network level. Owing to long lead times involved in the capacity provision process [5] the COCTA network capacity planning and management process extends over a 5-year horizon. We assume a long term contract between the NM and ANSPs on an annual capacity budget. This capacity budget is based on long term traffic forecasts and serves as a foundation for an ANSP’s decisions affecting capacity (e.g. staff training and technical equipment). For the sake of brevity, in this paper we focus on the strategic decision on capacity orders which the NM is taking six months in advance.

When airline schedules are published, around six months in advance of a schedule season, the NM has more precise information on O&D pairs and respective times of operations. Based on this knowledge of traffic demand (scheduled flights represent more than 80% of total flights for several years now [6]) the NM can define capacity orders within the capacity budget sketched above. Therefore, about six months in advance, the NM refines its capacity order from ANSPs, aligned with the long-term order. The NM orders capacity from ANSPs, which is measured with sector-hours. The capacity management process continues after this decision, with options to slightly adjust the capacity order, in line with demand information received subsequently. However, in this paper we do not model these steps and more details can be found in [5].

2. **Demand management**

In the redesigned ATM value-chain, we propose a novel approach to demand management. The NM manages demand by defining and offering different trajectory products, at differentiated prices, to AOs. The NM aims to improve network efficiency by optimising the utilisation of the airspace capacity which has been ordered from the ANSPs. Therefore, trajectory products are tailored to improve network performance.

The demand management starts once the initial capacity order is made, i.e. six months before the day of operation. At that moment, the NM should have a good estimate of the cost of capacity to be recovered with airspace charges. This estimate will be used as a baseline to define base airport-pair charges.

III. **Mathematical Model**

A. **Overview**

In this paper, we define a mathematical model for the initial capacity ordering at strategic level and demonstrate the NM’s decision-making using a small-scale, but realistic, example. We analyse principal trade-offs between capacity and demand management actions to improve overall cost-efficiency:

- Ordering (more) capacity, and thereby increasing the cost of capacity provision, to reduce costs of delaying or re-routing flights (uniformly termed displacement costs throughout the document) vs

- Delaying or re-routing flights in order not to increase the costs of capacity provision.

We assume that the NM’s primary aim is to order capacities across the network to maximize cost-efficiency, i.e. to minimize the sum of capacity provision and displacement costs. In addition, we also examine trade-offs between different performance indicators.
B. Assumptions

1. Network, flights and trajectories

We consider a set of flights $F$ flying over a network. Each flight $f$ connects an origin [$o$] to a destination [$d$] airport (OD pair). Trajectories (3D) for each OD pair are chosen from a set $R_{od}$ that contains several alternatives. Although the model can deal with 3D trajectories, in this paper, we consider trajectories only in the horizontal plane (2D - routes). The displacement cost is the additional cost if the route assigned is not an AO’s first choice, i.e., if it is displaced in space and/or time. As outlined above, we assume that AOs prefer flying the shortest routes which are also the cheapest in the COCTA context (assuming zero wind condition). The displacement cost of trajectory $r$ for a flight $f$ is $d_{rf}$. Finally, we use $B$ to denote the route-sector-time incidence matrix ($b_{rst} = 1$ if route $r$ uses sector $s$ at time $t$, 0 otherwise).

2. Sector configurations

We consider several airspaces $a \in A$, with each airspace $a$ composed by a set of elementary sectors $s \in S_a$. An airspace $a$ has a known number of sector configurations at which it can operate. Let $C^a$ be the set of these configurations, indexed by $c$. A configuration $c$ is identified by a partition $P_c$. Elements of a partition are indexed by $p$, to represent how the airspace is split among air traffic controllers. In other words, an element $p$ is a portion of the airspace, identified by a subset of elementary sectors $s \in S^p \subseteq S_a$. In our case study, we only consider horizontal divisions of airspace. However, the formulation of the model introduced is suitable to cope with vertical sectorisation.

Every element $p$ in a partition has a capacity $k^p$ denoting the maximum number of flights allowed to enter a sector, be it elementary or collapsed, per time period (commonly referred to as “entry counts”). A configuration is also defined by the number of sector-hours $h_{ac}$ which it consumes in every time period.

3. Time Scales

Two time scales are considered: a fine-scale used to describe trajectories and a coarse-scale used to model the dynamics of airspace configurations. Parameters $T^F$ and $U$ are the size of the fine-scale and coarse-scale time period, respectively. More specifically, $T^F$ represents the minimum unit used to define trajectories (e.g., 5-10min) and $U$ represents how often a sector configuration can change (e.g., 30-60min). For simplicity, we assume that a coarse-scale time period can be divided into an integer number of fine-scale time periods (i.e., $U\%T^F = 0$).

C. Model Formulation

Under the assumptions summarized in the previous section, we can now formulate the optimization model. The notation used is summarized in the following table:

Sets:

- $O$: Set of origin-destination pairs
- $F, F_{od}$: Respectively, the set of all flights and the set of flights connecting $od$
- $R_{od}$: The set of routes connecting $od$
- $T$: Fine-scale time horizon
- $U$: Coarse-scale time horizon
- $A$: Set of airspaces
- $C^a, S^a$: Set of configurations and elementary sectors for airspace $a$
- $P_c$: Partition of elementary sectors corresponding to a configuration
- $S^p$: Subset of elementary sectors forming a collapsed sector within a configuration
Indices:

- $f$: Flights
- $od$: Origin and destination airports
- $t$: Fine-scale time index
- $u$: Coarse-scale time index
- $r$: Route
- $a$: Airspace
- $c, c^a$: Airspace’s configuration
- $p$: Airspace sector (collapsed or elementary)
- $s$: Elementary sector

Parameters:

- $\rho_a$: Variable cost of providing one sector-time unit for airspace $a$
- $k_p$: Maximum capacity of airspace portion $p$
- $q_a$: Fixed cost of airspace $a$
- $h_a$: Number of sector-time units available at airspace $a$
- $\bar{R}_{ac}$: Number of sector-time units consumed by airspace $a$ working in configuration $c$
- $T$: Length (min) of a fine-scale time unit
- $\bar{U}$: Length (min) of a coarse-scale time unit
- $d'_r$: Displacement cost of route $r$ for flight $f$
- $gd_r$: Ground delay for route $r$
- $to_f$: Flight $f$ scheduled take off time
- $b_{ru}(to_f)$: Is equal to 1 if route $r$ uses sector $s$ at time $u$, assuming take off $to_f$, 0 otherwise
- $l_r$: Length of route $r$ expressed as number of time periods $t$

Variables:

- $z_{acu} = \begin{cases} 
  1 & \text{if airspace configuration is } c \text{ at time } u \\
  0 & \text{otherwise}
\end{cases}$
- $y'_r = \begin{cases} 
  1 & \text{if flight } f \text{ is assigned to route } r \\
  0 & \text{otherwise}
\end{cases}$

The problem of identifying optimal Airspaces Configurations and Demand Management (ACDM) is formulated below as a linear program:

\[
\begin{align*}
\text{min} & \quad \sum_{a \in A} \left( q_a + \rho_a \sum_{u \in U} \sum_{c \in C^a} \bar{R}_{ac} z_{acu} \right) + \sum_{r \in R} \sum_{u \in U} d'_r y'_r \\
\text{s. t.} & \quad \sum_{r \in r_{u,adj}} y'_r = 1 \quad \forall f \in F \quad (2) \\
& \quad \sum_{c \in C^a} \sum_{u \in U} z_{acu} = 1 \quad \forall a \in A, u \in U \quad (3) \\
& \quad \sum_{f \in F} \sum_{r \in r_{u,adj}} \sum_{u \in U} b_{ru}(to_f + gd_r) y'_r \\
& \quad K_p z_{acu} + |F| \sum_{c \in C^a} z_{acu} u \quad (4) \\
& \quad \sum_{u \in U} \sum_{c \in C^a} \bar{R}_{ac} z_{acu} h_u \quad \forall a \in A \quad (5) \\
& \quad z_{acu} \in \{0,1\} \quad \forall a \in A, c \in C^a \quad (6) \\
& \quad y'_r \in \{0,1\} \quad \forall f \in F, r \in R_{c,adj} \quad (7)
\end{align*}
\]
The objective (1) aims to minimize capacity and displacement cost. Constraints (2) ensure that each flight must be assigned to one and only one route. Constraints (3) state that one operating sector configuration must be defined at any time, for each airspace. Inequalities (4) set the capacity limitations across the network. More specifically, if partition \( p \) belongs to configuration \( c \) and \( c \) is chosen as configuration at time \( u \) (i.e., \( z_{acu} = 1 \)), then no more than \( K_p \) aircraft can enter sectors identified by \( p \), in period \( u \). However, if \( c \) is not chosen, then the term \( Fz_{acu} \) guarantees that the constraint is no longer binding. To compute the number of flights entering a sector in period \( u \), we need to consider the actual take off time given by \( t_{of} \) plus the ground delay \( g_d \) (based on assigned trajectory \( r \)). Inequalities (5) are the sector-hours budget constraints for each airspace that accounts for the fixed budget. Finally, (6)-(7) define the limitations for the decision variables.

IV. Numerical Results

We demonstrate fundamental trade-offs in the NM’s decision-making process of initial capacity ordering at strategic level, using a small-scale example. We assume that the NM purchases capacity from ANSPs about six months in advance, with limited (and costly) options for later capacity adaptations. As mentioned earlier, these capacity orders have to be based on traffic forecasts. However, for only approximately 80% of all flights (basically most scheduled flights), information on OD pairs as well as flight times is available. Therefore, the NM has to balance the risk of ordering too much capacity (and thus overspending) with the risk of ordering too little capacity (endangering stable service), using different rules for decision making.

A. Case Study Design

The network considered consists of five ANSPs represented by different colours in Figure 1. Four ANSPs have two elementary sectors each, while the central ANSP has three. The sectors’ 30-minute capacities range between 16 and 19. For the simulation we have to assume costs of capacity provision (see Appendix A).

**FIGURE 1 — Network structure**

There are two main traffic flows: F1 (east and west) and F2 (south and north). There are several sub-flows with indicated shortest routes in Figure 1. We assume that the F1 is a more dominant flow, with approximately two times more flights than the flow F2. Also, eastern flows are more dominant than western, as well as southern compared to northern flows.
We observe two hours of traffic in the given network with up to 150 flights in this time window; traffic closer to the upper bound is particularly challenging for the network.

We consider three aircraft sizes: small (E145), medium (A320) and large (B752).

Since we demonstrate initial capacity ordering six months in advance, we assume around 80% of demand is known to the NM. Therefore, we fix 120 scheduled flights (traditional and low-cost carriers) for which the NM has information on OD airport pairs, timetables and aircraft size (no cancellations of scheduled flights assumed).

Up to 30 flights, i.e. around 20% of demand, is uncertain demand (charter and cargo non-schedule, business aviation, other). We randomly choose between 1 and 30 uncertain flights: left-skewed probability distribution, with expected mean 20. The left skew is purposely introduced to study a challenging problem (i.e., the share of uncertain demand is likely to be significant). A distribution with a positive skew would reduce the impact of uncertain demand, hence making the simulation approach less of interest. One traffic sample with 30 uncertain flights and 120 schedule flights (150 in total) is shown in Figure 2.

Once the number of flights is selected using Monte Carlo simulation, aircraft sizes are assigned to uncertain flights. The average shares of small, medium and large aircraft are 30%, 60% and 10%, respectively. Flights are then assigned to different flows, preserving the share of flights on each flow.

Shortest routes are the cheapest in the COCTA context and, as such, preferred by AOs. If a flight cannot be assigned to the shortest route at the desired time of departure, the NM either delays it (up to 30 minutes) or reroutes it (up to 40NM).

**FIGURE 2 — One traffic sample with 150 flights (30 uncertain flights)**

B. Model testing steps

The NM anticipates scheduled flights as planned plus additional uncertain flights. However, the NM cannot know how many uncertain flights there will be, nor where or when they will appear. To decide on capacity ordering, the NM simulates many different uncertain demand materialisations, as explained above. Based on the results of the simulation, the NM decides on the amount of capacity ordered.

The computational analysis can be divided in two steps: Scenario Identification [SI] and Scenario Test [ST].

The SI step consists of iteratively sampling the uncertain demand and subsequently solving the unconstrained (referring to constraint [5]) ACDM model. The solution will suggest the optimal number of sector hours needed for each ANSP to accommodate the demand. This number will be directly obtained from $z_{ac}$ variables. The procedure is
repeated for a pre-determined number of iterations. Consequently, the output of SI will be a set of sector-hours budgets. With this set, the challenge is to identify a criterion to choose which budget should finally be implemented (i.e. how many sector hours should be ordered from each ANSP). In this analysis, we will consider and compare several decision criteria.

**Sector Hours Identification**

1: Set \( \text{Counter} = 0 \);
2: \textbf{REPEAT}
3: Generate flight demand based on the rules defined in \( X \);
4: Solve unconstrained ACDM;
5: Retrieve optimal solution and store the number of sector hours for each airspace;
6: \( \text{Counter} = \text{Counter} + 1 \);
7: \textbf{UNTIL} \( \text{Counter} = \text{NUM}_{\text{ITERATIONS}} \);

Once a budget has been chosen, a second simulation (ST) is run to test its robustness. Similarly, as in SI, at each step the uncertain demand is sampled and ACDM is solved. The sole difference is that ACDM now has a budget limitation set by vector \( h^* \) (left-hand side of constraint [5]).

**Sector Hours Test**

1: Select budget to test \( h^* \)
2: Set \( \text{Counter} = 0 \);
3: \textbf{REPEAT}
4: Generate flight demand ;
5: Solve ACDM with \( \tilde{h} = h^* \);
6: Store optimal solution;
7: \( \text{Counter} = \text{Counter} + 1 \);
8: \textbf{UNTIL} \( \text{Counter} = \text{NUM}_{\text{ITERATIONS}} \);

In each iteration optimal solutions are stored so that the budget can be evaluated by assessing performance parameters such as displacement cost or number of heavily delayed flights.

**C. Results**

Building upon the results of SI, we have selected and tested eight representative network capacity ordering scenarios, Table 1. All scenarios except MAX-PLUS have materialised at least once in the Simulation step 1, i.e. each of those seven scenarios was cost-optimal for some possible traffic materialisation, assuming, importantly, coordinated (centralised) capacity and demand management in place. To that end, the MAX (as well as MAX2 and MAX3) scenarios could arguably be interpreted as a decision of a conservative (delay-averse) Network Manager. In other words, these scenarios would be selected if the NM aims at having enough capacity to efficiently deal with all the uncertain scenarios it has foreseen. On the other hand, the MIN (as well as MIN2 and MIN3) scenarios could be seen as a decision of an ‘optimistic’ Network Manager. The FREQ scenario could be seen as the most-likely appropriate one. The network capacity budgets range between 11.5 (MIN scenario) and 15 sector-hours (MAX-PLUS scenario). Those eight scenarios account for 89.6% of all outcomes of SI. The FREQ scenario itself is a cost-optimal outcome in more than two thirds of all cases (step 1).

The MAX-PLUS scenario, on the other hand, is intentionally constructed with somewhat more generous capacity budgets compared to the MAX scenario. One might argue that such a scenario could be the outcome of an independent capacity-decision-making of
individual delay-averse ANSPs. Under such an assumption, the MAX-PLUS scenario might represent a valuable benchmark to assess the effects on network performance of various centrally-coordinated capacity provision scenarios.

Concerning the resulting traffic assignment, there are, on average, between 55 (MAX and MAX-PLUS scenario) and 63 (MIN scenario) displaced flights, meaning that, on average, the remaining 75-83 flights are assigned to the shortest routes with no delay (Table II). Average delay per delayed flight expectedly is reduced with increasing capacity budgets: from 9.5 minutes (MIN) to 7.4 minutes [MAX and MAX-PLUS].

<table>
<thead>
<tr>
<th>Scenario outcome</th>
<th>ANSP</th>
<th>Total capacity budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIN</td>
<td>2.5</td>
<td>11.5</td>
</tr>
<tr>
<td>MIN2</td>
<td>2.5</td>
<td>12</td>
</tr>
<tr>
<td>MIN3</td>
<td>2.5</td>
<td>12</td>
</tr>
<tr>
<td>FREQ</td>
<td>2.5</td>
<td>12.5</td>
</tr>
<tr>
<td>MAX2</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>MAX3</td>
<td>3</td>
<td>13.5</td>
</tr>
<tr>
<td>MAX</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>MAX-PLUS</td>
<td>3.5</td>
<td>15</td>
</tr>
</tbody>
</table>

The observed tradeoffs between the amount of capacity provided and displacement costs are intuitively expected across seven capacity-coordinated scenarios (i.e. all but MAX-PLUS), Figure 3. The difference in capacity costs (excluding structural capacity cost) between MIN and MAX scenarios is some 3,100 EUR (14%). That difference is more than offset by the decrease in average displacement cost: from 12,583 EUR in MIN to 7,759 EUR in MAX, with also notably higher variance of displacement costs in scenarios with scarcest capacity budgets, Table II (column 5). The total cost minimum is found in the FREQ scenario (capacity budget of 12.5 sector-hours), which thus represents, on average, the least expensive combination of capacity provision costs and displacement costs.

The observed tradeoffs between the amount of capacity provided and displacement costs are intuitively expected across seven capacity-coordinated scenarios (i.e. all but MAX-PLUS), Figure 3. The difference in capacity costs (excluding structural capacity cost) between MIN and MAX scenarios is some 3,100 EUR (14%). That difference is more than offset by the decrease in average displacement cost: from 12,583 EUR in MIN to 7,759 EUR in MAX, with also notably higher variance of displacement costs in scenarios with scarcest capacity budgets, Table II (column 5). The total cost minimum is found in the FREQ scenario (capacity budget of 12.5 sector-hours), which thus represents, on average, the least expensive combination of capacity provision costs and displacement costs.

### Table I — Capacity Ordering Scenarios Tested (all values in sector hours)

<table>
<thead>
<tr>
<th>Scenario outcome</th>
<th>ANSP</th>
<th>Total capacity budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIN</td>
<td>2.5</td>
<td>11.5</td>
</tr>
<tr>
<td>MIN2</td>
<td>2.5</td>
<td>12</td>
</tr>
<tr>
<td>MIN3</td>
<td>2.5</td>
<td>12</td>
</tr>
<tr>
<td>FREQ</td>
<td>2.5</td>
<td>12.5</td>
</tr>
<tr>
<td>MAX2</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>MAX3</td>
<td>3</td>
<td>13.5</td>
</tr>
<tr>
<td>MAX</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>MAX-PLUS</td>
<td>3.5</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario outcome</th>
<th>Capacity budget (sector hours)</th>
<th>Fixed capacity cost in EUR</th>
<th>Variable capacity cost in EUR (average)</th>
<th>Average displacement cost in EUR (st. dev.)</th>
<th>Total cost in EUR (average)</th>
<th>Average number of displaced flights (st. dev.)</th>
<th>Total delay (minutes)</th>
<th>Average delay (minutes) per delayed flight (average)</th>
<th>Number of flights delayed ≥20min (average)</th>
<th>Relative incidence (% of all sector periods with utilisation ≥85%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIN</td>
<td>11.5</td>
<td>6,080</td>
<td>15,990</td>
<td>12,583</td>
<td>34,653</td>
<td>63.5 [6.6]</td>
<td>526</td>
<td>9.5</td>
<td>6.0</td>
<td>32.4</td>
</tr>
<tr>
<td>MIN2</td>
<td>12</td>
<td>6,385</td>
<td>16,782</td>
<td>11,487</td>
<td>34,654</td>
<td>60.3 [7.2]</td>
<td>507</td>
<td>9.3</td>
<td>4.6</td>
<td>31.1</td>
</tr>
<tr>
<td>MIN3</td>
<td>12</td>
<td>6,330</td>
<td>16,624</td>
<td>9,905</td>
<td>32,859</td>
<td>63.9 [7.4]</td>
<td>372</td>
<td>8.1</td>
<td>2.1</td>
<td>28.1</td>
</tr>
<tr>
<td>FREQ</td>
<td>12.5</td>
<td>6,635</td>
<td>17,541</td>
<td>8,280</td>
<td>32,456</td>
<td>56.4 [7.3]</td>
<td>336</td>
<td>7.8</td>
<td>1.3</td>
<td>27.9</td>
</tr>
<tr>
<td>MAX2</td>
<td>13</td>
<td>6,885</td>
<td>17,769</td>
<td>7,954</td>
<td>32,608</td>
<td>55.6 [6.6]</td>
<td>314</td>
<td>7.6</td>
<td>1.1</td>
<td>27.4</td>
</tr>
<tr>
<td>MAX3</td>
<td>13.5</td>
<td>7,135</td>
<td>17,886</td>
<td>7,771</td>
<td>32,792</td>
<td>54.9 [6.0]</td>
<td>301</td>
<td>7.5</td>
<td>1.0</td>
<td>27.6</td>
</tr>
<tr>
<td>MAX</td>
<td>14</td>
<td>7,385</td>
<td>17,891</td>
<td>7,759</td>
<td>33,036</td>
<td>55.0 [6.0]</td>
<td>300</td>
<td>7.4</td>
<td>1.0</td>
<td>27.6</td>
</tr>
<tr>
<td>MAX-PLUS</td>
<td>15</td>
<td>7,940</td>
<td>17,935</td>
<td>7,713</td>
<td>33,588</td>
<td>54.7 [5.8]</td>
<td>298</td>
<td>7.4</td>
<td>1.0</td>
<td>27.4</td>
</tr>
</tbody>
</table>
191 Demand and capacity building

The MIN scenario is on average 6.8% more costly than FREQ, whereas MAX is 1.8% more costly than the FREQ scenario. The MAX2 scenario (with extra 0.5 hours bought from ANSP R compared to FREQ capacity order) is only 0.4% more costly than FREQ. Importantly, the benchmark scenario (MAX-PLUS) does not perform better than most of the other seven capacity-coordinated scenarios (quite oppositely, only MIN and MIN2 are, on average, more costly than MAX-PLUS). Starting from the MAX scenario capacity order as a baseline, the addition of 0.5 sector-hours from ANSPs R and Q each, resulting in MAX-PLUS scenario, does not improve cost-efficiency, since it yields only a very small reduction in displacement cost, which is more than offset by the associated increase in capacity costs. The MAX-PLUS scenario is, on average, 3.5% more costly than the FREQ scenario, and 1.7% more costly than MAX scenario. This seems to highlight that a central coordination approach is generally more advantageous than having every ANSP deciding independently, with a delay-averse approach employed.

Environmental performance, measured via extra CO₂ emitted (compared to shortest routes), is solely driven by re-routings (with re-routing larger aircraft being more harmful, ceteris paribus). The scenarios with lowest capacity orders perform better than the others in this respect. Among the four most generous-capacity scenarios, FREQ is the best performer concerning CO₂ emissions. MAX-PLUS on average yields the same amount of CO₂ emissions as MAX and MAX-3 scenarios.

Concerning the right end of the distribution of delayed flights, which can be interpreted also as a rough proxy for equity (fairness), the average number of flights delayed by >20 minutes rapidly decreases with capacity increase, Figure 3. It should nevertheless be noted that, owing to the cost-minimisation objective, practically all long re-routings are applied on small and medium-sized aircraft, while the average number of long re-routings applied on large aircraft is close to zero.
On a related matter, it should be noted that large aircraft by far most frequently get assigned the optimal (shortest) trajectory, with this frequency ranging between 79% in MIN scenario and 88% in FREQ, MAX, MAX2 and MAX3 scenarios. This means that only between one in five and one in nine large aircraft are expected to get displaced, on average. The percentages of shortest route assignments are notably lower for medium-sized aircraft (ranging between 53% and 64% in different scenarios), and in particular for small aircraft (40-46%). Finally, reflecting on performance of the MAX-PLUS scenario in this respect, adding further capacity compared to MAX scenario, does not yield any tangible benefits.

Another interesting aspect concerns the robustness of different capacity ordering scenarios, in terms of ability to absorb [withstand] certain deviations from assumed traffic parameters while obeying capacity constraints. In this regard, more generous capacity budgets expectedly perform somewhat better than scarcer ones, measured by relative incidence of sector-periods with capacity utilisation ≥85%, Table II. This indicator peaks at 32.4% in the MIN scenario, meaning that nearly a third of all sector-periods are expected to experience quite high utilisation of their declared capacities.

Finally, an interesting albeit not surprising situation concerns the comparison of the MIN2 and MIN3 scenarios, which have the same total capacity budget (12 sector-hours each), but those are slightly differently distributed across ANSPs, see Table I above. More specifically, starting from the MIN scenario as a baseline, the decision whether extra 0.5 sector-hours are ordered from ANSP Q (leading to Scenario MIN2) or from ANSP U (leading to Scenario MIN3) quite notably affects the resulting network performance. Whereas the additional capacity cost is similar in both cases, the effects on displacement costs and environmental performance are remarkably different. More specifically, enabling extra capacity in ANSP Q does not have a positive net effect on total cost (i.e. the reduction in displacement cost is not sufficient to offset the increased capacity cost), but at the same time it leads to the by far most environmentally efficient outcome across all scenarios tested. Conversely, ordering extra 0.5 sector-hours from ANSP U notably decreases both the displacement cost (average value as well as the standard deviation) and the total cost, see Table II, but at the same time results in the worst-CO₂ score across all scenarios tested.

A possible comparison baseline, corresponding to the present state of affairs in Europe, should assume airspace-based charges and consequently some longer routes chosen by AOs even when sufficient capacities are available to support shortest-route options [7]. Concerning the capacity provision matter itself, one might argue that the present insufficient coordination between ANSPs in a strategic timeframe could effectively be closest to the MAX-PLUS scenario tested [strategic dimensioning of capacities against local traffic peaks]. This is more likely to be true concerning structural costs and maximum capacity provision costs (see Appendix), than concerning sector hour provision costs (since structural capacities, albeit charged for, are often not delivered on the day of operations [5]). Thus, while maximum capacity provision is nowadays arguably paid for [by AOs], the full operational benefits thereof are not necessarily extracted. The latter implies that the typically delivered capacity levels [and thus also associated displacement costs] are arguably closer to MIN2, MIN3 or FREQ scenario, while the charges more likely reflect the MAX-PLUS, MAX or MAX-3 scenario examined. This further implies that the cost-efficiency benefits estimated in our small case study are most likely on a conservative side.
V. Conclusion and Outlook

We have developed a systematic approach to illustrate the impact of trading off costs of capacity provision versus cost of displacement in the context of a small case study. Compared to previous papers, one major addition is the introduction of a realistic portion of uncertain demand, allowing to analyse the NM’s decision making on capacity ordering.

The small numerical example offers insight into the effects of timely (well in advance) coordinated capacity provision (network-centric approach), and centralised demand management, with (still) effectively no active route charging approach in force. The results obtained confirm the existence of intuitively expected performance tradeoffs associated with different airspace capacity levels provided across the network.

The methodological approach employed enables to estimate the likely effects of incremental changes in capacity provided in different network segments. For instance, ordering an extra sector-hour from ANSP [sector] A vs. ordering it from ANSP B vs. do-nothing option, as in the above-discussed case of MIN2 vs. MIN3 vs. MIN scenarios.

It should be mentioned that experiments with flatter demand pattern over time were also run, in order to understand and quantify the dependency of results upon the extent of traffic “peakness”. The results indicate a possibly tangible impact of traffic profile per se, first of all concerning the incidence (frequency) of demand management actions needed, and the associated costs thereof. Due to space constraints in this paper, for more details we refer the interested reader to ref. [8].

Future work will, inter alia, address the interdependencies between higher strategic displacements and airport capacity restrictions [slots].

Finally, we have been operating in a single-objective framework so far, while only monitoring the impact on key performance indicators other than total cost. In forthcoming work, we intend to test incorporation of other objectives as well to provide more comprehensive results.

REFERENCES

[5] COCTA consortium, “Initial mechanism design,” 2017. Available at: https://cocta.hs-worms.de/fileadmin/media/SESAR/699326-COCTA-D4.1-Initial_mechanism_design-00_02_00_FINAL.pdf
[8] COCTA consortium, “Prototype models and small academic examples,” 2017. Available at: https://cocta.hs-worms.de/fileadmin/media/SESAR/699326-COCTA-D5.1-Prototype_models_and_small_academic_examples-00_02_00_FINAL.pdf
VI. Appendix a: Capacity and displacement cost

Since the capacity decision is made six months in advance, a large share of total cost is considered to be fixed cost, called structural capacity costs. The variable part of capacity provision costs is called sector hour provision cost, Table III. However, since the ANSP cannot cease its operation (i.e. has to operate at minimum capacity at any time) the minimum cost for a time period of two hours is the sum of the structural capacity costs plus the product of minimum capacity and the sector hour capacity provision cost. For the numerical example we assumed different costs for neighbouring ANSPs, a situation which can be found in many parts of the European airspace. For more details, refer to ref. [8]. Regarding the displacement cost per aircraft type, we rely on findings in [9].

### TABLE III — Cost of capacity provision

<table>
<thead>
<tr>
<th></th>
<th>Q</th>
<th>T</th>
<th>R</th>
<th>U</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum sectors simultaneously open</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Minimum capacity [sector-hours per 2 hours]</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$q_a$ Structural capacity cost [EUR per 2 hours]</td>
<td>5,000</td>
<td>4,000</td>
<td>4,000</td>
<td>4,000</td>
<td>4,000</td>
</tr>
<tr>
<td>$\rho_a$ Sector hour provision cost [EUR per active ( \frac{1}{2} ) sector hour]</td>
<td>920</td>
<td>570</td>
<td>750</td>
<td>650</td>
<td>460</td>
</tr>
</tbody>
</table>
Abstract—The User Driven Prioritisation Process (UDPP) is a concept under
development in SESAR with as objective achieving additional flexibility for Airspace
Users (AUs), i.e., the ability of the ATM system to accommodate AUs’ changing
business priorities. More flexibility could result in better costefficient delay
management during congested situations, with substantial reductions of operational
cost impacts for AUs. Equity (in the sense that one AU’s prioritisation does not
negatively impact another AU) is the main constraint for UDPP. This paper contributes
to explore the limits of flexibility beyond the current UDPP validated features. A User
Delay Optimisation Model (UDOM) is presented to analyse the hypothetical case in
which an AU has high flexibility to minimise its own global delay costs, having full
freedom to transfer delay among its flights and to exchange flight sequence positions
with other AUs. After imposing a constraint of equity (total AU’s delay must remain the
same), it is shown that: a) there is an optimal level of delay for each of the AU’s flights;
and b) such equity condition increases flexibility in the system.

I. Introduction

In order to maintain safety in the Air Traffic Management (ATM) system, the European
Network Air Traffic Flow Management (ATFM) Function at Airports or En-Route imposes
delays or other measures on certain flights before departure [1], [2]. It is well known that
ATFM delay causes operational irregularities with important costs to the Airspace Users
(AUs), airports and passengers [4][5], ATFM delays being one of many irregularities
reducing the operational efficiency, but one over which airlines have little influence.

Profitability in air transport industry is very sensitive to cost variations (profit margins
might be as low as 1-2%) [6], thus AUs would like further flexibility, i.e., the ability of the
ATM system to accommodate AUs’ changing business priorities, to reduce the ‘impact
of delay’ (cost of delay) during irregular operations.

Delay is used today as a key performance indicator (KPI) of ATM capacity (capacity to
maintain safety in operations), and thus most of the KPIs steering the ATFM Demand
and Capacity Balancing (DCB) Function are based on average delay per flight, while DCB
targets are strongly oriented towards a ‘No-Delay paradigm’ [2], [3]. As a consequence, in the event of a demand-capacity imbalance (a.k.a., hotspot), the Flow Management Position (FMP) in charge will most likely find a solution that decreases the overall system delay first, and whenever possible also reduce the impact of delay on AUs.

However, the impact of delay on AUs’ operations, which is highly important information only known by the AUs, cannot be fully taken into account by DCB. If AUs’ priorities could be considered during the DCB decision-making processes, this will have a large positive impact on the efficiency and predictability of the ATM operations. Airspace Users’ participation in ATM and airport collaborative processes is therefore essential to minimise the impacts of deteriorated operations on all such stakeholders, thus giving strong arguments for the application of de-centralised decision-making (i.e., user-driven approach) as potential solution to achieve efficiency in the ATFM slot/delay allocation [7], [8].

SESAR envisioned the development of the User Driven Prioritisation Process (UDPP) to achieve additional flexibility for AUs to adapt their operations in a more cost-efficient manner [9]. UDPP concept is today under development and new features are being progressively incorporated aiming to fulfil different operational requirements and implementation constraints. Some of these features have already been proposed and validated with different levels of maturity, such as Enhanced Slot Swapping validated in 2015 and deployed in May 2017. Other less mature features are explained in [9], [10].

The aim of this paper is to explore the limits of flexibility beyond the current UDPP validated features, in particular exploring the hypothetical case in which high flexibility is given to an AU to minimise its own global delay costs, i.e., the AU has full freedom to transfer its total baseline delay (i.e., initial ATFM delay) among its flights and to exchange freely flight sequence positions with other AUs only being subject to one particular equity constraint: AU’s total baseline delay cannot be reduced.

The potential implications of introducing high flexibility subject to equity will be discussed via the theoretical analysis of the User Delay Optimisation Model (UDOM), a simplified mathematical framework developed in the context of this research to capture the complex relationships between time, cost of operations and the flexible and equitable allocation of slots and delay. Two different degrees of equitable flexibility will be explored: a) the equity condition must be fulfilled at each single hotspot; and, b) equity requirements can be fulfilled after many hotspots in a long-term period (e.g., one year).

The remainder of the paper has been structured as follows. Section II provides a background and state of the art of UDPP; Section III describes the UDOM framework and results; Section IV discusses the implications of such results for UDPP; Section V presents the conclusions and future works.

II. UDPP Background and Definitions

A. Current concept of operations and reason to change

In today’s operations, a few hours before a potential demand-capacity imbalance is foreseen with a certain level of confidence, the Network Manager activates a regulation scenario and issues ‘ATFM slots’, which will apply a tactical time-based separation between flights to ease the safe and smooth management of air traffic flows and sector/airport capacities during tactical and flight execution operations [1], [2]. Those ATFM slots are then allocated to the flights involved in the regulation, thus changing their times of departure with respect to the original slots scheduled for those flights, and thus causing delay on flights as a consequence.

The ATFM slots are not allocated on an arbitrary basis. Instead the process typically follows a transparent set of rules and policies previously agreed and accepted by all the relevant ATM stakeholders, including the AUs. The most common policy used today to allocate delay –when no other more constraining rule or operational policy applies– is
the First Planned First Served (FPFS), which sorts the flights by the estimated time of arrival at or over (ETA/ETO) the constrained airport or sector, according to the information present in the filed flight plans and assigns the slots in such order [2].

FPFS is widely accepted by AUs because it preserves the original sequence of flights (considered fair), and it is well accepted today in ATFM operations because it minimises the total delay in a regulation [11], [12]. FPFS policy does not take into account that delay is allocated differently to the flights and that each flight may have different impact of delay.

Figure 1 shows the cost model that is being developed in the context of UDPP together with the AUs participation. Each flight has its own particular complex cost structure only known to the AU. The cost structure of a flight is typically not linear, due to the presence of different milestones and time constraints for each flight, such as crew out-of-hours constraints, maintenance slot requirements [such as a ramp check], passenger missed-connection costs, high-yield passenger business-retention [‘soft’] costs, or a missed airport curfew, etc. [Reference values for these and other variables affecting the cost of delay of flights and AUs can be found in [5]]. If a flight is delayed so that these important milestones or constraints cannot be fulfilled, then large negative impacts on AUs operational costs are typically the consequence. To mitigate such impacts, the AUs would like whenever possible more flexibility to prioritise their flights to redistribute delay on the basis of the consequences on operations and costs.

**Figure 1 — Typical cost structure model per flight**

AUs are very heterogeneous in their size, form and business strategies, and thus, they often have very different operational needs, in particular regarding the flights subject to ATFM regulations. But in general it has been recognised by AUs that flights often have some tolerance to delay [i.e., margins], because although a minute of delay always has a cost, this cost can often be considered marginal in practice if delay is not trespassing the more constraining operational margins.

Figure 2 shows three flights of the same AU that are impacted differently by delay, since each flight has a different position in a sequence as well as different cost structures. Note that each flight has very different cost structure shape, either in the size of their delay margins and/or in the magnitude of the impact of delay. In the example, flight FL001 has little delay and little impact of delay, FL002 has mid delay but relatively large impact, and FL003 has the largest delay but relatively small impact in comparison with FL002. Note that the impact of delay for a single flight [e.g., FL002] might also include the costs associated to the potential knock-on/cascade effect caused by a certain amount of delay allocated to that flight.

**Figure 2 — Different cost structures for each flight**
Figure 3 shows the global cost of delay for the AU of the example taking into consideration his three flights. The initial situation in the baseline sequence (e.g., FPFS sequence) is shown in the left part of the figure. The right part of the figure shows the benefits of giving flexibility to AU to transfer delay between its flights. For instance, by exchanging the positions of FL001 and FL002 (UDPP Slot Swapping) the delay D1 initially allocated to FL001 is transferred to FL002, and delay D2 to FL001. A large cost reduction might be possible for the AU by just changing that position.

**FIGURE 3 — Contribution of flexibility to global cost optimisation**

The UDPP mechanism of slot swapping has already been validated in terms of impact on equity and acceptability by the AUs and FMPs. Nevertheless, the AUs are not always in the ideal situation of having low priority flights (with enough margins and/or relatively low economic value) in positions nearby their most impacted (high priority) flights so they can exchange their positions between them. Indeed, in a recent study performed internally in EUROCONTROL (still not published), based on the analysis of all the airport regulations in 20 consecutive AIRAC cycles, it has been found that in 85% of the regulations in which AUs are involved, typically only a few flights (equal or less than 3) are affected, which strongly limits the flexibility provided by basic UDPP concepts such as slot swapping. In this situation in which the AUs have a small number of affected flights it is said that the AU is an LVUC (Low Volume User in Constraint) [10]. In addition, note that some AUs often operate just a few flights and thus they might have little flexibility or even never be able to prioritise their important flights, which might be inequitable from the access KPA point of view (e.g., business aviation is specially vulnerable to that problem). Therefore, there is a need to explore new features in UDPP to enable more flexibility for all the AUs, in particular LVUCs.

New UDPP features are being under research in the context of SESAR2020, including the possibility of exchanging slots and delay among different AUs, either in the same hotspot or in several hotspots over time. To shortly introduce the idea, in Figure 4 a new example is presented in which the output of the previous example (Figure 3) is the baseline situation. Note that the situation of the AU could notably improve if the flight FL001 could be advanced some positions in the sequence, thus reducing its delay. A new UDPP feature being investigated today is the possibility for flight FL001 to take a better position that was initially allocated to another AU (the position in front would be enough in this example). In exchange, the AU should compensate the delay reduction on that flight by accepting more delay in other flights (e.g., delay FL003 one or more positions), not necessarily in the same hotspot, which might be a good solution for LVUCs.

Such advanced concepts will increase a lot the flexibility (e.g., in Figure 4 the AU will benefit from a non-negligible cost reduction), but it becomes more complex to put in place a set of UDPP Rules for which equity can be demonstrated, i.e., ensure that the actions of one AU will not impact negatively the operational priorities of other AUs (equity is the main UDPP constraint).

---

1 Roughly, one year and a half from Jan 2016 to July 2017 (AIRACs from 1601 to 1707 taken from EUROCONTROL’s DRR).
Figure 5 shows a conceptual map of different degrees of flexibility being studied today in the UDPP context via the development of different models and features. The development of mechanisms that allow the AUs to exchange delay between themselves in a transparent and equitable manner is key to enhance flexibility and cost reduction opportunities. However, the introduction of highly flexible advanced mechanisms is constrained by the difficulties on designing, testing and validating, including the generation of proofs of equity. For instance, the extension of flexibility over time (exchange of delay between different hotspots) may require the introduction of a system based on ‘credits’ (a virtual currency with no monetary value) to account the amount of delay exchanged between flights and between AUs over time. This introduces additional challenges in terms of implementation, operational acceptability and access to include all AUs. Therefore, to make steps forward towards a more flexible ATM while minimising the associated risks, it is important that all relevant stakeholders (and specially AUs) understand the potential benefits of the different degrees of flexibility and equity potentially provided by each proposed UDPP mechanism/feature as well as its particular limitations.

Next sections will explore, through a simplified analytical approach, the effects that forcing equity in the UDPP mechanism may have on the AUs dominant strategies when they are able to transfer delay freely between their own flights, and in particular it will be shown how equity may contribute to increase, rather than to reduce, the levels of flexibility for AUs.
III. Description and Analysis of the User Delay Optimisation Model (UDOM)

Table 1 describes the model parameters and variables used in the UDOM model, which will be further discussed in the following sub-sections.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>$[0, +\infty)$</td>
<td>Delay of a flight ($d_0$, means $d=0$)</td>
</tr>
<tr>
<td>$U_0$</td>
<td>$[0, +\infty)$</td>
<td>Max utility of a flight when its $d=0$</td>
</tr>
<tr>
<td>$N$</td>
<td>$[0, +\infty)$</td>
<td>Number of flights operated by an AU in the reference period</td>
</tr>
<tr>
<td>$\varepsilon_i$</td>
<td>$(-\infty, +\infty)$</td>
<td>Elasticity of the utility function of flight $i$. Used to model in a simplified way (continuous model) different operational flight margins</td>
</tr>
<tr>
<td>$\rho_i$</td>
<td>$[0,1]$</td>
<td>Probability of a flight $i$ for being affected/delayed by a hotspot</td>
</tr>
<tr>
<td>$\delta_i$</td>
<td>$[0, +\infty)$</td>
<td>Average delay expected for flight $i$ in the route operated (i.e., typical delay from hotspots on that route)</td>
</tr>
<tr>
<td>$\delta_i^0$</td>
<td>$[0, +\infty)$</td>
<td>Baseline (random) delay for flight $i$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>$(-\infty, +\infty)$</td>
<td>Delay shift, to increase or reduce the delay of flight $i$</td>
</tr>
<tr>
<td>$d_i^*$</td>
<td>$[0, +\infty)$</td>
<td>Optimal delay for flight $i$ in the actual hotspot</td>
</tr>
<tr>
<td>$d_i^\circ$</td>
<td>$(-\infty, +\infty)$</td>
<td>Shifted delay, i.e., delay difference between the optimal delay and the baseline delay, for flight $i$</td>
</tr>
</tbody>
</table>

A. Definitions and assumptions

One of the main assumptions in UDOM is that the AUs taking part in the system can be modelled as utility maximising, i.e., the major objective of each AU is to maximise its utility function. Therefore, AUs are assumed to have a utility function, which is depending on several variables, such as the delay cost structure of each flight.

Utility is an important concept in economics and game theory, because it represents satisfaction experienced by the consumer of a good. In the context of this document, the concept of ‘utility’ will be understood as the value perceived by a particular AU if a given slot is allocated to a particular flight operated. Without loss of generality, in this document it is assumed that utility is directly related to economic profits obtained by the AUs for operating their flights, however the concept of utility may also include any type of operational constraints known by the flight dispatcher, and any indirect economic or non-economic type of benefits or costs.

Figure 6 shows a simplified representation of a utility function. In reality, utility functions are unknown and may be non-linear and non-convex (like for instance the cost functions). However, in this document it is assumed for the sake of simplicity that every single flight has a maximum utility when the delay, $d$, is zero, at a certain slot, and the utility is then progressively decreasing as far as the delay is increasing. Utility will be negative if the cost of delay has become greater than the maximum economic value expected for that flight if operated on time.

A utility function for a single flight, $U(t)$, can be represented analytically as a quadratic function. If negative delay is not considered [simplification], the utility as a function of the delay, $d$, assigned to a flight can be expressed as:

$$U(d) = \frac{\varepsilon}{2} d^2 + U_0, \quad \forall d \geq 0$$

$$\frac{\partial U}{\partial d} = \mu = \varepsilon d < 0$$

$$\frac{\partial^2 U}{\partial d^2} = \varepsilon < 0$$
Demand and capacity building

where ε is the elasticity of the utility function, μ, the marginal utility and U₀, the maximum utility in case of no delay allocated to that flight. Note that since the elasticity is negative, any delay incurred by a flight will cause a reduction on the utility perceived by the AU for this flight. Different revenue and cost structures can be modelled by changing the parameters, thus they can be adapted to different types of AUs’ type of activity (low cost, HUB, business aviation or others).

FIGURE 6 — Utility function of a flight

Under the highly dynamic and uncertain ATM operational environment each flight operated is subject to a certain probability ρᵢ of being involved in a hotspot and thus being delayed. In the long term (e.g., one year) the average delay expected for each flight can be quantified and expressed as δᵢ. Therefore, the expected utility for an AU that operates N flights can be expressed as {dᵢ means d=0, i.e., on-time}:

$$\bar{U} = \sum_{i=1}^{N} U_i(\delta_i)(1-\rho) + \sum_{i=1}^{N} U_i(\delta)\rho$$

(2)

In the absence of uncertainty (i.e., in the hypothetical case in which all flights could be operated in their scheduled time), the maximum utility is determined only by the sum of utility functions of each individual flight operated by an AU (assumed constant daily utility). However, the actual utility is fluctuant and equal or lower to the maximum utility, thus the average long-term utility (i.e., expected utility) will be lower than the maximum in the absence of uncertainty. The higher the expected average delay for each operated flight and the higher the probability of being delayed, the lower the average expected utility will be. See Figure 7.

FIGURE 7 — Expected utility with and without uncertainty

Participation in UDPP is voluntary. Thus, AUs will only participate if they do not receive a negative payoff, i.e., a lower expected utility than with no participation. In this section it is shown how AUs will be able to improve the expected return (increase the expected utility) of their operations in the presence of ATFM regulations that will affect their scheduled times, either if they decide to optimise in the short-term (i.e., only optimising the delay allocation between the flights involved in a particular hotspot) or if they decide to have a long-term strategy (i.e., optimising the average expected long-term utility by managing and allocating the delay between flights involved in hotspots occurring in different places and times). Such flexibility will create natural incentives to the AUs to participate, because they will typically be able to maximise their flight utility by managing and re-allocating the ATFM baseline delay of their own flights.
Since a short-term optimisation strategy can be considered as a particular case of the long-term optimisation strategy, the former will be explained after the second.

### B. Definition and condition of the equity constraint

In the ideal high flexibility conditions under consideration, an AU would be able to exchange delay between his flights, even when the AU is a LVUC and has only one or a few flights in a given hotpot. Therefore, in case of a hotspot, the AU is willing to increase the delay with a delay shift, $\tau$, for the flight that brings less utility in order to be able to reduce the delay in the same quantity for the flight that brings higher utility in the same or in future hotspots. To avoid potential system abuses, the model incorporates an equity constraint that forces the user to have no debts (nor surplus) at the end of the reference period (AUs total baseline delay cannot be reduced):

$$\sum_{i=0}^{N} \tau_i = 0 \quad [3]$$

### C. UDOM with equity imposed at the end of a long-term reference period (multihotspot flexibility)

To simplify the analysis of UDOM let us consider first a simple case in which an AU has two different flights, $f_1$ and $f_2$, with equal probability $\rho$ of being delayed and also equal expected delay $\delta$ (this could be the case, for instance, for two flights scheduled to the same destination airport during the same period). Then, the equation of expected utility according to (2) can be expressed as:

$$U_\text{d} = U_{f1}(d_0) + U_{f2}(d_0)[(1-\rho) + U_{f1}(\delta) + U_{f2}(\delta)](\rho) \quad [4]$$

The optimisation problem that the AU faces, subject to the equity constraint, is:

$$\max_{\tau_{f1},\tau_{f2}} U = \left[U_{f1}(d_0) + U_{f2}(d_0)\right](1-\rho) + \left[U_{f1}(\delta + \tau_{f1}) + U_{f2}(\delta + \tau_{f2})\right](\rho) \quad [5]$$

$$\text{st} \{ \tau_{f1} + \tau_{f2} = 0 \}$$

or equivalently,

$$\max_{\tau_{f1}} U = \left[U_{f1}(\delta + \tau_{f1}) + U_{f2}(\delta)\right](1-\rho) + \left[U_{f1}(\delta) + U_{f2}(\delta + \tau_{f2})\right](\rho) \quad [6]$$

Note that the value of $\delta + \tau$ for each flight will indicate to the AU which the optimal delay is for each of its flights subject to a hotspot with regard to the long-term average expected utility. To calculate the long-term optimum delay for each flight the AU must take into consideration all his flight operations expected for the reference period, together with the expected number of flights that might be regulated in the period as well as the average delay expected for each flight (statistical characterisation could be based on historical operational records).

The optimal delay shift, $\tau^*_i$, can be found by equalling the first derivative of [6] to zero. For instance, for the flight $f_1$:

$$\frac{\partial U}{\partial \tau_{f1}} = [\epsilon_{f1}(\delta + \tau_{f1}) - \epsilon_{f2}(\delta - \tau_{f2})] \rho = 0 \quad [7]$$

$$\tau_{f1}^* = \frac{\delta \epsilon_{f2} - \epsilon_{f1}}{\epsilon_{f2} + \epsilon_{f1}} \quad [8]$$

The (long-term) optimum delay to apply to flight $i$ in case of a hotspot is given by:

$$\delta_i = \delta + \tau_{f1}^* \quad [9]$$
And the shifted delay $d_{i}^\ast$ to be applied when a flight $i$ is affected by a random delay $\hat{d}_i$ (that follows a distribution with mean $\delta$) can be calculated with:

$$d_{i}^\ast = \hat{d}_i + r_i$$  \hspace{1cm} (10)

1. **Illustrative example 1: Multihotspot flexibility**

Let us consider the following scenario for illustration: $\epsilon_{f1} = -2$; $\epsilon_{f2} = -10$; $\delta = 15\text{min}$; $\rho = 0.2$. Let us also consider a maximum utility per each of the flights equal to $U_0 = 500$. Therefore, the maximum total utility under zero uncertainty (i.e., $\rho = 0$) is given by (4), $U_{\text{max}} = 500+ 500 = 1000$, while in the presence of uncertainty (i.e., $\rho = 0.2$) the expected utility is:

$$\bar{U} = U_{\text{max}} (0.8) + \left(U_{f1}(15) + U_{f2}(15)\right)(0.2) = 1000(0.8) + \left(500 - \frac{2}{2}15^2 + 500 - \frac{10}{2}15^2 \right) = 730$$  \hspace{1cm} (11)

According to (8): $r_{f1} = 10$; $r_{f2} = -10$. This means that for a given random delay $\hat{d}_i$ that affects flight $f1$ or flight $f2$ (as a consequence of an ATFM delay during a hotspot), the AU will try to apply an extra or reduced amount of delay to the flight until the (long-term) optimum delay given by (9) is reached. This has as effect that in the long-term the (optimised) expected utility will be, according to (6):

$$\bar{U} = U_{\text{max}} (0.8) + \left(U_{f1}(15 + r_{f1}) + U_{f2}(15 - r_{f2})\right)(0.2) = 1000(0.8) + \left(500 - \frac{2}{2}(15+10)^2 + 500 - \frac{10}{2}(15-10)^2 \right) = 850$$  \hspace{1cm} (12)

Figure 8 shows the comparison among the long-term utilities, i.e., without uncertainty ($U=1000$), with uncertainty and FFPS policy ($U=730$), and with uncertainty and the UDOM² sequence positions/delay allocation ($U=850$).

Analysing (8) it could be argued that any AU has economical interest to participate in this –ideal– UDPP mechanism if the difference between the elasticity of each flight utilities is different from zero ($\epsilon_{f2} - \epsilon_{f1} \neq 0$). Otherwise the AU will be indifferent (with UDOM the expected utility achieved would be the same as without UDOM).

**FIGURE 8 — Long-term utilities for the scenario 1**

![Graph showing long-term utilities for scenario 1](image)

D. **Multiple hotspots with different probabilities to happen**

The model depicted in (5) can be extended to consider the cases in which two flights have different probabilities to be involved in a hotspot, i.e., $\rho_{f1}$, $\rho_{f2}$, and when it happens they are delayed with different average delay, i.e., $\delta_{f1}$, $\delta_{f2}$ (e.g., one flight has destination to Heathrow and the other to Madrid):

$$\begin{align*}
\max_{\rho_{f1},\rho_{f2}} U &= U_{f1}(d_{f1}(1-\rho_{f1}) + U_{f2}(d_{f2}(1-\rho_{f2})) \\
&+ U_{f1}(\delta_{f1} + r_{f1})\rho_{f1} + U_{f2}(\delta_{f2} + r_{f2})\rho_{f2} \\
\text{s.t.} &\{r_{f1} + r_{f2} = 0\}
\end{align*}$$  \hspace{1cm} (13)

² Under ideal flexibility conditions it can be assumed that all the exchange proposals are possible (i.e., no 'market incompleteness').
For this model, the optimal delay shift for flight $f_1$, $t^*_1$, can be found with:

$$
t^*_1 = \frac{\delta_2 \rho_2 \epsilon_2 - \delta_1 \rho_1 \epsilon_1}{\rho_2 \epsilon_2 + \rho_1 \epsilon_1}
$$

(14)

E. Generalisation of the UDOM to $N$ flights and multiple hotspots

The optimisation model can also be generalised for $N$ flights of a same AU, each of them characterised by a different elasticity $\epsilon_i$ and affected by delays with different probabilities $\rho_i$ and with a particular and different expected delay $\delta_i$:

$$
\max D = \sum_{i=1}^{N} U(\epsilon_i)(1-\rho_i) + \sum_{i=1}^{N} U(\delta_i + \tau)(\rho_i)
$$

(15)

subject to:

$$
\sum_{i=1}^{N} \tau_i = 0
$$

After some mathematical development (e.g., using multipliers of Lagrange), the optimal delay shift for each flight $i$ can be expressed by:

$$
\tau_i = \frac{\sum_{i=1}^{N} \delta_i \rho_i \epsilon_i}{\sum_{i=1}^{N} \rho_i \epsilon_i}
$$

(16)

Next example will be with three flights.

F. UDOM with equity imposed at the end of each hotspot (flexibility constrained by short-term reference periods)

In the particular case that high flexibility is only allowed to transfer delay among flights in the same hotspot (i.e., the AU must have the same total baseline delay at the end of the hotspot), the above formulae can be adapted to find the optimal solution for the AUs in this particular situation (note that optimising in the short-term may provide higher utility in the short-term but less in the long-term; however, long-term optimisation might be more heuristic, or perhaps impracticable, due to difficulties in assigning realistic costs to future flights). To optimise the delay allocation of the flights involved in the same hotspot, the AU just needs to substitute in equation (16) the value of the average delays, $\delta_i$, by the actual (random) delay, $\tilde{\delta}_i$, while the probabilities for each flight to be involved in the hotspot should be parameterised to $\rho_i = 1$ (the flights are actually involved in a hotspot). The resulting optimal delay shift $\tau_i$ will determine the optimal delay for each flight in such a hotspot (see equation (10)) and therefore will indicate the demand of slots of such particular AU (note that the equity condition is fulfilled by the AU at the end of the short-term reference period, i.e., in a single hotspot, something that may not happen when the reference period is longer term, i.e., multiple hotspots).

1. Illustrative example 3: Single-hotspot flexibility

Consider an AU with three flights $f_1, f_2, f_3$ involved in a hotspot. The actual delay allocated by FPFS is $\tilde{\delta}_1 = 5$, $\tilde{\delta}_2 = 12$ and $\tilde{\delta}_3 = 20$ respectively, and the elasticity of the utility functions is approximated by $\epsilon_1 = -2$, $\epsilon_2 = -10$ and $\epsilon_3 = -9$. Consider the same maximum utility $U_0 = 500$ for all the flights (in case that they were not delayed). The total utility of the sequence corresponding with the Baseline Delay would be implemented (e.g., FPFS) can be calculated as:

$$
U_{BD} = U_{f_1}(5) + U_{f_2}(12) + U_{f_3}(20) = 475 - 220 - 1300 = -1045
$$

(17)

Using equation (16) it can be found that the optimal delay shifts are: $\tau^*_1 = 21$, $\tau^*_2 = -7$ and $\tau^*_3 = -14$. Applying these shifts, the new ESFP optimised utility for that AU is:
This illustrates that it is possible to move from large losses to significant profits (max utility without delay was 1500).

IV. Discussion

A. UDOM findings and contributions to UDPP

The UDOM analysis has shown that if high flexibility is given to an AU to re-allocate his flights’ delay, with freedom to take and give delay from/to other AUs’ flights, but constrained by equity forcing the AU to give back the delay that has been taken from the network (either in the short or the long term), then the AU shall be able to minimise his costs by transferring delay among his flights. This finding is very important, because it shows that, while equity can be enforced, all AUs shall have economic incentives to participate in UDPP.

The UDOM has also shown that, in the event of an ATFM hotspot in which a flight is delayed, the AU operating that flight may sometimes prefer to allocate more delay to that flight, even if this flight is of high relative importance\(^3\). This is something that may seem contradictory compared to the current ‘no-delay’ paradigm, but that is fully justified by the need of the AU to give back the delay taken from other AUs during the prioritisation of its flights (the AU must give positions in order to take positions). Such finding, i.e., the need of an AU to accept more delay in some flights to optimise his own utility, is also of high importance, since such self-interested participation of an AU, constrained by equity rules, is indeed what contributes to generate more flexibility for the others (beyond the flexibility already provided by the flexible utilisation of their own slots). When an AU voluntarily accepts more delay in a flight he is indeed offering positions in the sequence that can be taken by others to reduce their delay. Future UDPP mechanisms will therefore consist of a set of simple UDPP principles and rules for (almost) effortless cooperation between AUs driven by the enhancement of their own flexibility, while concentrating on the optimisation of their own operations.

B. UDOM limitations and UDPP Challenges

UDOM has been useful to facilitate the understanding of AUs’ dominant strategies in a high flexible situation; however, the UDOM solutions are still far from being implementable in real operations. Hereafter some limitations are discussed (grouped in some well-known categories [13], [8]).

Externalities and Market Incompleteness: the decision-making processes of AUs are mutually dependent and subject to complex interactions. AUs trying to allocate their optimal delays may sometimes result in more than one flight pointing to the same position in the sequence, thus leading to the unavailability of the slot desired for some flights. Such type of negative externality (which could also be understood as a kind of market incompleteness) requires further research, e.g., to know i) how often AUs can actually reach their preferred solutions, ii) how often many UDOM de-centralised solutions could be merged into feasible and equitable sequence solutions, iii) how often other constraints, for instance availability of airport resources, is limiting flexibility of AUs, etc.

Bounded Rationality: in decision-making, individual rationality is limited by the available information, AUs’ cognitive limitations and the finite time they have to make a decision. ATM environment is highly uncertain and dynamic, which poses a real challenge for the

\[^3\] E.g., an ‘important’ flight could receive 5 minutes of extra delay, if its baseline delay is 7 minutes and it has a delay tolerance of 15 minutes.
flight dispatchers that have to make complex operational decisions with high impact on costs. In addition, the assignation of realistic costs to delays is often non-trivial even for the AU himself. Ideal conditions of UDOM should be updated to shed light on such complexities of real operational environments.

Asymmetric Information and Incentive compatibility: the economic incentives of flexibility are a strong argument in favour of long-term cooperation [see 14], [15], and [16]. However, information is imperfect and it is asymmetrically available to AUs, which may lead to some AUs to be suspicious about the potential non-cooperative behaviour of others. Monitoring and protecting against potential abuses under high flexibility conditions, including data security and confidentiality (e.g., AUs learning other AU cost priorities from UDPP activity) require further research and validation.

V. Conclusions and Future Work

A hypothetical situation has been studied in this paper with the help of the analytical model UDOM, in which an AU can optimise his operations under high flexibility conditions subject to one equity constraint: within a given reference period, the AU is allowed to re-allocate his ATFM baseline delay among his flights without reducing it. Such study contributes to illustrate the potential benefits and challenges of increasing flexibility with respect to current UDPP validated levels as well as to pave the way for future UDPP features.

Our analysis shows that, under UDOM ideal conditions, there is an optimal level of delay—typically greater than zero—for each flight affected by an ATFM slot, that minimises the total cost of flights operated by the AU within the period. Imposing equity constraints encourages AUs to request additional delay for some of their flights in order to reduce delay in other flights. This may be an important source of flexibility for other flights located later in the sequence [flights of the same AU or others’], because it enables a potential exchange of flight sequence positions [delay transfer] between flights with different delay needs.

This is an important finding that has two corollaries. Firstly, flexibility can be generated by and for UDPP participants by exploiting the available flight delay tolerances, either from own or from others’ flights participating. Secondly, the larger the number of AUs participating in UDPP, the higher the levels of flexibility achievable for all participants (assuming that some flights tolerate extra delay at relatively low cost impact).

New UDPP features that allow the exchange of slots between AUs (to increase flexibility) will therefore consist of a set of simple UDPP principles and rules to smoothly coordinate between AUs [collaborative decision-making], to increase flexibility and equity on their own behalf [possibly benefiting also passengers and airports], while AUs can basically focus on optimising their own operations. It does not matter if AUs participate in UDPP only because of self-interest, because when their self-interested operational decisions are constrained with equity rules, their participation can nevertheless contribute to enhance the KPAs of flexibility, operational efficiency and cost-effectiveness in the overall ATM Network.

New UDPP features exploring the high degree of flexibility shown in this paper are difficult to develop and validate. Future research must firstly find a valid set of rules to solve those cases in which several AUs compete for the same slots; strong evidence will be required to prove that the new feature solutions always converge to feasible and equitable sequences; access and equity for all AUs must be guaranteed, including the incorporation of LVUCs needs; finally, realistic operational conditions, with highly uncertain and dynamic ATM constraints, must also be addressed in future UDPP research.
Acknowledgment

The authors of this paper would like to acknowledge and thank Andrew Cook (University of Westminster), Lorenzo Castelli (Università degli Studi di Triestel), Andrada Brujor, Andrea Ranieri, and Ruben Martinez (Advanced Logistics Group), for their important contributions to the UDPP concept that have largely influenced the understanding and development of the concepts presented in this paper, and to Kris Delcourte (EUROCONTROL) for his kind review.

REFERENCES

Abstract — We present the model developed within the Vista project, studying the future evolution of trade-offs between Key Performance Indicators. The model has a very broad scope and aims to simulate the changes that business and regulatory forces have at a strategic, pre-tactical and tactical level. The relevant factors that will affect the air transportation system are presented, as well as the scenarios to be simulated. The overall architecture of the model is described and a more detailed presentation of the economic component of the model is given. Some preliminary results of this part of the model illustrate its main mechanisms and capabilities.

I. Introduction and objectives

The air transportation system is a continuously evolving complex socio-economic entity. In order to monitor its changes, SESAR and other bodies regularly define Key Performance Indicators (KPIs) grouped in Key Performance Areas (KPAs) and the associated targets to drive it in the desired direction [1]. In particular, projects are supposed to estimate the impact of their implementation in the system through the use of these indicators.

However, the impact of a single change in the system, let alone multiple changes, cannot be easily forecast because of the high degree of entanglement of the different components of the system, including the different stakeholders – notably passengers, airlines, ANSPs and airports. The degree to which the indicators depend on each other thus arises directly from the interactions and the complex behaviours of the actors in the system. Possible trade-offs and synergies might arise between indicators, actors, and changes in the system.

This calls for a holistic view of the system rather than the independent assessment of its sub-parts. Hence, the primary objective of Vista is to quantify the current and future (2035, 2050) relationships between a currently non-reconciled set of performance targets in Europe by using an integrated model of the European air transportation system on which ‘what-if’ scenarios can be tested.

Vista has commenced this task by making a list of future business and regulatory changes and how they may likely impact on the stakeholders. In order to take into account the complex feedback between the actors, a model with three layers has been
designed, mimicking the three temporal stages: strategic, pre-tactical and tactical phases. The three layers have a fine granularity in terms of scope, down to the individual passenger for the tactical portion. They also feature complex behaviours from the various stakeholders, whilst bearing in mind the need to keep the model simple enough to be able to calibrate it with real data.

This paper aims at presenting the general ideas behind the model and some more specific details on the economic part, trying to highlight the challenges of building a holistic ATM model with such a broad temporal and spatial scope. It is organised as follows. Section II presents the operational environment modelled in Vista: stakeholders, factors and indicators. This section describes the different stakeholders modelled; how business and regulatory factors, taken into account to predict the evolution of the KPIs, have been selected; and which indicators will be estimated by the model in the future. Section III presents the model itself, first with a general description of its architecture and then with a more detailed description of the economic model at the heart of the strategic layer. Section IV presents some preliminary results obtained with the economic model. Finally, we draw some conclusions in Section V, together with some plans for the future of the model.

II. Operational environment

In this section, we highlight the main operational components of the air transportation system in order to include them in a comprehensive model. Moreover, we look at how they are likely to change in the two time horizons set by Vista – 2035 and 2050 – defining the main drivers of these changes as ‘factors’. Finally, we briefly look at the observables in the system, how to measure them what to expect from the model.

A. Stakeholders

Five stakeholders are represented in the model: ANSPs, airports, airlines, passengers, and the environment. It is the ambition of Vista to provide a unique view of each perspective and how these are likely to evolve. We discuss each one in turn.

ANSPs are heavily regulated and have traditionally been providing the full scope of air navigation services, including CNS (communication, navigation, surveillance), AIS (aeronautical information services) and, in some cases, aeronautical meteorology. This monolithic approach is gradually changing. Next to pressure from the regulatory side on unbundling of services, technological innovations such as virtual centres and remote towers pave the way for different ANSP business models. In some states, competition on, for example, tower services, is enforced by the local regulator, or the ANSP itself may decide to outsource services to increase cost efficiency. In addition, almost all ANSPs have become engaged in one or more strategic alliances and industrial partnerships [2]. For example, an ANSP can at the same time join an operational alliance on free route airspace, establish a joint venture with a training provider and team up with ANSPs that share the same ATM system manufacturer.

For large airports, the current business model heavily relies on non-aeronautical revenues (parking, shopping, etc.) [3]. Most of their other revenues are from airlines using them as hubs. Congestion is a major issue for most and they need to implement different strategies to increase their capacity, including soft management procedures or heavy changes in infrastructure [4]. Small airports rely proportionally more on their aeronautical revenues, and try to attract low-cost, point-to-point traffic. Many of them also play the role of feeders for hubs. The evolution of airports relies heavily on the business models of airlines and the increase in traffic in the future. Capacity extensions and better ATM tools (e.g. extended arrival management) are thus to be expected at several airports. A spectrum of private and public ownership exists, but nearly all are heavily regulated, in particular regarding aeronautical charges [5].
Airlines are probably the most market-driven stakeholders. Since it is quite easy for them to reassign – or ground – aircraft, they are able to respond to external stimuli quite quickly. Low-cost carriers (LCCs) generally have lower yields compared with ‘traditional’ operators. LCC expansion has mainly been based on point-to-point (P2P) strategies, aiming at higher utilisation by using a homogeneous fleet, and lower costs by using secondary airports, in particular [6]. However, more recently, some LCCs have shifted to the ‘legacy’ model to some extent. For instance, Ryanair has started to operate at primary airports and easyJet has agreements to feed Norwegian and WestJet long-haul flights. The more ‘traditional’ carriers have been forced to lower costs, sometimes trying to gain market share with an ‘in-house LCC’. In future, this apparent convergence will depend, in particular, on the price of fuel.

For the passenger, price, travel time, comfort and convenience constitute some of the factors influencing their choice [7]. The literature often defines archetypal profiles for passengers, usually taking into account socio-economics and travel purpose (often simply ‘business’ or ‘leisure’). Some profiles have been defined during the project DATASET2050 [8], based on several sets of data, and will be loosely adopted in the Vista model, since different types of passenger demonstrate different behaviours when it comes to price, convenience, etc.

The last ‘stakeholder’ is the environment, ensuring that Vista includes the impact of emissions in the trade-off assessments with other KPIs. Whilst noise is not included in the model, it is planned to include a measure of CO₂ emissions, which are linear with the amount of fuel burned, and an estimate of NOₓ emissions (although these strictly depend on the background atmospheric conditions, including temperature and humidity, and altitude [9]).

B. Factors

The evolution of the above stakeholders depends on external factors and internal dynamical effects. As a consequence, Vista identified regulatory and business factors and how they will influence the system in the 2035 and 2050 framework. Business factors define economic, technological and operational changes. Regulatory factors encompass regulations and policy instruments that have a direct impact on air transport operations, or that enable the implementation of business factors. A total of 85 references have been reviewed including ICAO, European and national regulations, SESAR documentation and research publications. We have also consulted with stakeholders.

For each factor, possible values and their expected impact on the system have been identified. In some cases, their quantitative impacts are based on literature reviews and goals defined by the SESAR program. In other cases, their impact affects how some operations are carried out. Where possible, a discretisation between a ‘low’, ‘medium’ and ‘high’ effect of the factor is considered. These values are related to the same baseline, for the different timeframes considered in Vista. Indeed, for some of them, we took as reference the late ‘time-based operations’, ‘trajectory-based operations’, and ‘performance-based operations’ targets as they were defined by SESAR in the past [10], [11]. When possible, we mapped them to the possible values of the factors as follows: low is trajectory-based operations; medium is performance-based operations; and, high represents an enhancement of performance-based operations. For example, the impact of development of foreground factor ‘traffic synchronisation tools’ is based on SESAR defined targets which will represent an increase in airport capacity by 1% in time-based operations and 1.07% for trajectory-based operations, a fuel efficiency of -0.3% in time-based operations and an increase in airspace capacity at TMAs of 5% in time-based operations and 6.74% in trajectory-based operations [12]. For other factors, such as ‘passenger provision schemes’, the possible values are linked to operational changes (e.g. modification of the threshold of entitlement to compensation in case of delay) and its impact will be modelled by changes on the behaviour of the stakeholders either directly or indirectly (e.g. these changes might represent a higher cost of delay, which might in turn impact the willingness to recover delay). As the value of the factors are related to a baseline which is not linked with the temporal evolution, different values
for the factors can be reached in 2035 and 2050 depending on the scenario considered, see below and Table I.

**TABLE I — Background scenario definition. The last column includes two examples of the values of factors in each scenario. ‘BED1’ represents the global economic development of the EU-EFTA zone, while ‘BTS4’ is the possibility to have enhanced route structures.**

<table>
<thead>
<tr>
<th>Period</th>
<th>Name</th>
<th>Example of factor values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>2035</td>
<td>L35: Low economic</td>
<td>BED1: Low</td>
</tr>
<tr>
<td></td>
<td>M35: High economic</td>
<td>BED1: Medium</td>
</tr>
<tr>
<td></td>
<td>H35: High economic</td>
<td>BED1: Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low technology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High technology</td>
</tr>
<tr>
<td>2050</td>
<td>L50: Low economic</td>
<td>BED1: Medium</td>
</tr>
<tr>
<td></td>
<td>M50: High economic</td>
<td>BED1: High</td>
</tr>
<tr>
<td></td>
<td>H50: High economic</td>
<td>BED1: High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low technology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High technology</td>
</tr>
</tbody>
</table>

1. **Regulatory factors**

A total of 22 regulatory factors have been identified. They are grouped into three categories: regulations affecting the gate-to-gate phase of the flight (including SESAR development and integration, performance-based and ANSP requirements regulations), regulations affecting airport operations (grouped by legislation with impact on airport demand, airport processes and airport access and egress), and regulations affecting other areas in air transport (such as passenger provision schemes, e.g. Regulation 261). As previously mentioned, most of regulatory factors are enablers of operational and technological changes. For example, regulation allowing the operation of UAS is needed for the deployment of these systems, but the regulation itself, without the business development, does not represent an impact on the system. On the contrary, other regulations have a direct effect on the behaviour of stakeholders and the system, this includes, for example, changes to Regulation 261 regarding passenger compensation, or the introduction of new emission trading systems.

2. **Business factors**

37 business factors have been classified across categories: factors affecting the gate-to-gate phase of operations (including SESAR operational changes), airport processes and accessibility, and factors affecting demand and other economic variables (such as economic development in Europe and fuel prices).

3. **Scenarios**

A scenario in Vista is modelled by combining a temporal frame (current, 2035 or 2050) and individual values for the regulatory and business factors. The high number of factors and their possible values needs careful management at the analysis stage. For some factors, their individual impact will be assessed. For others, either a small impact on the indicators measured in Vista or a high consensus on their evolution is defined or their coupling with other factors is very high, thus not allowing the modification of their values independently in the model. The factors that will be analysed in more detail are classified as foreground factors, the rest as background factors.

Background factors are grouped in a meaningful manner to create background scenarios to which different options for the foreground factors can be applied. Table I shows the different background scenarios defined. These scenarios are created to model 2035 and
2050 timeframes with and without high economic growth in Europe and considering whether technology is developed to accommodate the economic evolution. This should allow us to analyse the impact of a shortfall of technological development along with the individual impact of foreground factors. The values ‘low’ and ‘high’ for economy and technology then need to be mapped to the ‘low’, ‘medium’, and ‘high’ values of each of the background factor. This has been already been done in Vista and some examples of values for each scenarios are given in Table I.

A total of 14 foreground factors are identified, e.g. changes to regulations defining passenger provision schemes. These foreground factors are grouped into four higher-level categories: environmental and mitigation policies, regional infrastructure development, passenger focus and Single European Sky evolution.

C. Key performance indicators and trade-offs

Different indicators have been identified for each stakeholder. They are selected considering their relevance to stakeholders and the possibility of capture them in the model with the degree of reliability and precision required. Specific consultation with stakeholders has been conducted and will be expanded on further activities to ensure that the most relevant metrics are considered. The metrics currently considered focus on the on time (gate-to-gate and door-to-door times, delays experienced or generated per stakeholder, missed connections), economic (delay hard/soft costs, value of time, revenue and costs) and environmental performance (CO₂, NOₓ).

The high uncertainty on the modelling of 2035 and 2050 timeframes means that the objective of Vista is not to precisely compute the value of the indicators on a specific scenario, but to understand the main tendencies and the trade-offs between them under the different scenarios, timeframes and factors. In this manner Vista will provide insight on the relationship between indicators and factors. The trade-off methods that will be used and results will be reported in future publications.

III. Model

A. General architecture

Since the breadth of the project is large, the model in Vista is composed of different layers with timeframes aligned with key aspects of future target setting (2035 and 2050). This is depicted in Figure 1.
Overseeing the other layers, the ‘environment’ is designed to be the host of all the static variables feeding the other parts of the model. It comprises the values taken by all the factors, but also historical data to feed to model. It communicates data to each layer.

The first operational layer, the ‘strategic layer’, is designed to capture high-level, long-term decisions by the stakeholders. These decisions are based on a changing environment, comprising socio-economic variables (e.g. demand, fuel prices) and taking into account simple economic feedback. The strategic layer has the objective of providing outputs from the main flows between cities, down to the microscopic level of individual schedules and passenger flows. The main features of the economic model are described in the next section.

The ‘pre-tactical’ layer aims at transforming the consequences of the strategic decisions (the schedules, passengers flows and capacities) into realistic day-to-day operations. As explained previously, there are different archetypes of passengers in Vista, each of them having different characteristics which translate into probabilities of making specific itineraries. This layer is thus tasked with converting the schedules and passengers flows produced by the strategic layer and converting them into passenger itineraries. It also produces flight plans that could be used for the actual flight during the tactical phase, as well as other possible flight plans available during disruption, based on historical data and traffic patterns. The pre-tactical layer will also estimate ATFM regulations and delays.

The third layer is the ‘tactical’ layer, which is designed to simulate a day of operations. As input, it takes the passenger itineraries produced by the pre-tactical layer, as well as the flight plans and possible airspace disruptions. It simulates the entire day of operations by tracking, microscopically, each passenger and each flight. It generates ad hoc delays and disturbances, based on the information provided by the pre-tactical layer. The model is not strongly agent-driven and only minimal decisions are taken by the airlines during this phase. This matches the idea that on the day of operation the options are very much reduced for them. The tactical layer is based on the mobility model called ‘Mercury’ and developed during previous projects, including POEM, ComplexityCosts, and DATASET2050.
Finally, the last layer displayed in Figure 1 is the learning loop. Since the Vista model is actually a succession of models, each of them feeding the one downstream, some discrepancies could appear between the layers. For instance, the evaluation of the cost of one minute of delay for the airline during the strategic phase could be wrong and turn out to be significantly different during the tactical phase. In fact, these discrepancies are desirable, to some extent, because they exist in the real system. The network operations centre – which takes tactical decisions – is typically distinct from the marketing department – which takes some strategic decisions – in most airlines. Of course, there is some communication between them, but this is imperfect. This is an important feature of this complex socio-economic system that the actors are not hyper-rational with perfect information but have rather a bounded rationality with imperfect information. However, it is also important that the information within the strategic layer is consistent with the tactical one, to avoid unrealistic discrepancies leading to unrealistic decisions as they will be adjusted in the mid-term. One possibility to achieve this is to check the output of the tactical run and compare it with the expectations that the agents formed during the strategic phase. This is the idea behind the learning loop, which should be able to compute different KPIs important to the stakeholders and thus adjust their behaviours.

B. Economic model

In this section, we describe the economic model already developed in Vista, for which the results are shown in Section IV. We briefly show its functionality and the principles behind its design. As shown in Figure 1, the economic model in Vista is part of the strategic layer and constitutes its main component. It is the first block of a chain of models and thus provides information to all the blocks downstream. It should reflect the main, high-level changes occurring in the ATM system in the future. Its objective is essentially to take into account the macro-economic factors to forecast the main changes of flows in Europe. As a consequence, its output should include the:

- main traffic flows in Europe;
- typical market shares of different airline types;
- capacities of ANSPs and airports;
- average prices for passenger itineraries.

In order to do that, the model should simulate various mechanisms within the system and take into account the inputs from the environment. In particular, it should take into account or reflect the:

- main changes in demand, in terms of:
  - volume;
  - passenger heterogeneity (types);
- major business changes:
  - point-to-point vs hub-based operations (airlines);
  - competition vs cooperation (ANSP);
  - privatisation vs nationalisation (ANSP and airports);
  - etc.

1 Of any socio-economic system in fact.
capacity restrictions and congestion, in particular:
- congestion at airports;
- ATCO resource constraints.

major changes in costs/prices of commodities:
- fuel;
- airspace/airport charges;
- new technology developments.

All these mechanisms are directly or indirectly influenced by several business or regulatory factors, and a list of the factors influencing the economic model has been made to ensure that no important ones have been omitted. In order to take all this into account, and because of the heterogeneity of the system in terms of flows and types of actors, it has been decided to use a deterministic agent-based model embedded in the network in order to find the economic equilibrium between the different actors and how it is impacted by the changes in the environment.

1. General flow of the ABM

The agent-based model currently features three types of agents, the:

- airport (one agent per airport);
- airline (one agent per airline);
- passengers (one agent per OD pair, including all the possible itineraries).

Each agent has its own objective, with a specific cost function. The simulation is turn-based, and during a turn all the agents form expectations and decide to act accordingly in order to reach their objectives. A turn proceeds as follows:

- airlines estimate the prices of each itinerary, based on past prices;
- airlines estimate the delays at airports, based on past delays;
- airlines choose their operated capacity for each airport pair based on their cost function, the estimated delays, and the estimated prices;
- airports estimate their traffic;
- airports decide whether to expand their capacity, based on expected traffic and their own cost function;
- passengers choose between the different itineraries available for a given OD pair;
- the price of each itinerary is updated based on the discrepancy between demand and supply;
- delays are updated based on real traffic;
- airports and airlines compute their final profit.

The airports also have a lag in the construction of their capacity, contrary to the airlines. Indeed, there are several turns between the decision to expand their capacities and the step in which the capacity is available to airlines.
2. **Cost functions**

The agents have different objectives and different expectations. First, the airlines estimate the price for each itinerary for a given turn. This is done by using an exponentially-weighted average of the past changes in prices. The same technique is used with the delay generated by the airports. Once the airline 'knows' the price and the delay, it computes the optimal (seat) capacity to provide for each airport pair, based on the following cost function, for a given airport pair:

\[
c_a = c_0 + c_1 S + c_2 S^2 = c_0 + (c_1 + \chi \Delta \delta_{t0} + \chi \Delta \delta_{tD}) S + c_2 S^2. \tag{1}
\]

The cost structure of the airline has been chosen to reflect the long-term choices that the airlines (are supposed to) make in the strategic layer. Apart from the constant term, it includes a linear term which is the cost of operating a capacity \(S\) (for instance in number of passengers). This term can be slightly non-linear with the capacity due to inefficiencies but we keep it linear in the model. The coefficient \(c_1\) includes the cost of crew, cost of fuel, cost of delay, etc. The latter is modelled as a linear law, following the findings of [13], in which the cost of delay is found to follow a quadratic law for airlines, with a relatively weak quadratic term for lower delays\(^2\). Hence, \(\chi\) represents the cost of one minute of delay for the airline, \(\Delta \delta_{t0}\) and \(\Delta \delta_{tD}\) being the additional delay generated by the origin and destination airports, with respect to the initial (calibrated) situation.

The second term is related to the cost of capital, which is non-linear with the capacity produced. Indeed, it is important to realise that in the model the airlines do not address the question of the optimal capacity given the capital (goods), but rather the optimal level of capital given the expected costs and revenues. As a consequence, the airline adjusts its capacity \(S\) based in fact on the underlying choice of the capital (goods) ‘\(K\)’, here representing in particular the aircraft. The exact cost of (additional) capital is certainly a complex function, which depends on the size of the airline, its financial situation, the state of the finance system, the state of the aircraft production/leasing sector, etc. However, this function needs to be monotonically increasing and concave in order to have diminishing returns, which is why \(\alpha > 1\).

The profit of the airline for one airport pair is simply \(r_a = pS - c_a\), where \(p\) is the price of the ticket for the airport pair AB. The optimal capacity is thus simply given by:

\[
S = \left(\frac{p - c_1}{c_2}\right)^{1/2}.
\]

To set the capacity provided on an airport pair in this turn, the airline uses the estimations of \(p\) and \(c\) that it performed previously and uses this equation. Note that \(c_1\) in the future will include other terms, for instance ANSP charges.

The airport is currently essentially a delay manager. Given traffic \(T\), the airport produces some delay because of congestion, following the equation:

\[
\delta t = \delta t_0 + \frac{T}{C}. \tag{2}
\]

where \(\delta t_0\) is a constant and \(C\) is defined as the capacity of the airport. This linear phenomenological equation is often used in the literature [15], [16], albeit usually without a constant term. It is justifiable with queuing theory, assuming a maximum number of movements per unit of time and a random queue for the flights [17]. Note that other functional forms are sometimes used in the literature, for instance with a divergence when reaching the capacity [18]. In practice, exponential and linear laws both fit well the data [14]. When a linear law is fitted with data, the intercept is found to be significant for nearly every big airport in Europe, and is sometimes negative.

Note that here we include every flight operated at this airport in the generation of delay, either departing or arriving. It does not preclude other types of delays to be added to the delay of the flight, and thus represents only the part of delay purely due to congestion at the airport (terminal and runway). The delay generated by the airport is dependent

---

\(^2\) But the full cost of delay is to be included in the model, including the statistical effects studied in [14]
on the traffic at this airport, but this traffic is in turn dependent on the delay because airlines are sensitive to it, as shown in equation 1. This is a typical example of (negative) economic feedback, which is resolved in the model by a convergence of the expected levels of traffic and delays with the actual ones.

Just as the airline tries to predict the delay at the airport, the airport tries to estimate the traffic, since it does not have direct access to the supply function of the airlines’ agents. Once the traffic is estimated, the airport computes its expected profit in two hypothetical cases, its capacity:

1. remains the same, or,
2. is increased by a fixed increment.

To do this, we use for its operating cost a linear law with respect to its capacity. A similar law has been used in [14]. Additional data would be required to compute the actual production function of the airport and test this linearity. Once the airport knows the two values of the profit in cases 1 and 2, the airport decides to build additional capacity if the profit in case 2 is higher than a given threshold with respect to case 1. The airport then spends a fixed number of steps with the current capacity, after which the capacity is increased suddenly by a given amount.

Note that airports do not yet change their charges, as it is currently implemented in the model. Since airports have very diverse regulations, some react sharply to markets and others are slower to react. As a consequence, this version of the model considers that all airports have constant charges per passenger. We also considered that the revenues of the airport were linear with the traffic (in fact, the number of passengers). In reality, there is an aeronautical component, proportional to the number of flights (and to the number of passengers also for some airports), and a non-aeronautical component. The latter is more complex, and comprises parking charges, concession rents, etc. Most of these are directly proportional to the number of passengers or can be safely assumed to be so in the long run (for concessions, for instance).

Finally, the passengers are represented by agents which are quite passive. They do not forecast any value and they have an implicit utility function which drives them to choose one itinerary over the other. Their choice is driven by different variables, some of them depending on other agents [such as prices], others depending on the model environment [such as passenger income levels]. Given an OD pair, the demand on a particular itinerary \( k \) operated is given by:

\[
D_k = D_k^0 (1 - \alpha \Delta p_k + \beta \Delta i_k) C(p_k, \{p_l\}_l). \tag{3}
\]

This equation is composed of two terms. The first is a ‘volume’ term, sensitive to the difference in price \( \Delta p_k \) of the itinerary with respect to a baseline [the initial situation] and the difference in income \( \Delta i_k \) of the passengers on this itinerary with respect to the same baseline. \( \alpha \) and \( \beta \) represent, respectively, the price elasticity and the income elasticity of the passengers. There is a huge literature (see for instance [19]–[21]) dedicated to the computation of these two types of elasticities, which we will use during the calibration phase of the model. The second term \( C(p_k, \{p_l\}_l) \) is a term of competition, which is a function of the prices of all the itineraries possible for this OD pair. It is directly inspired by choice models where, given a discrete choice with different advantages and disadvantages, passengers have a given probability to choose one option or the other. This function of choice needs to be decreasing in its first argument [the price of itinerary \( k \)] and increasing in its other arguments [the prices of the itineraries \( l \neq k \)]. There are several choices possible for this function, among which the multinomial logit function is popular, see for instance [22]. For numerical reasons, the logit function is sometimes problematic, so we choose instead a linear function:

\[
C(p_k, \{p_l\}_l) = 1 - \frac{1}{s} \left( \Delta p_k - \sum_{l \neq k} \Delta p_l \right) . \tag{4}
\]
with the additional constraints \( 0 \leq C \leq 1 \). Hence the competition is sensitive to the difference between the price of itinerary \( k \) and the average price of the other itineraries. The parameter \( s \) is the (inverse of the) intensity of choice. When it is large, the passengers are not very sensitive to the difference of prices between different itineraries (and in particular between airlines serving the same OD pair). In the limit where \( s \to 0 \), passengers all choose the less costly solution, hence having perfect competition between itineraries (and thus airlines).

Note that in reality other factors should enter the composition of the competition term. In particular, it is known that the passengers are sensitive to the quality of service. This is, however, difficult to calibrate, as there are no available, substantial data on this. Another parameter is the total travel time, since passengers are usually more attracted by shorter travel times, e.g. selecting a direct option rather than a flight with a connection. This is related to the complex issue of the value of time of passengers, for which there is a substantial literature. We thus plan to include this factor in the next version of the model.

**IV. First strategic (economic) results**

In this section we explore the possibilities of the Vista model, and more particularly its strategic component. The results shown below have been obtained on a very simplified setting in order to show the main mechanisms within the model and how the interaction between the agents leads to some non-trivial behaviours. The set-up is illustrated in Figure 2. We define four airports – labelled from 0 to 3 – and two airlines – A and B. One airline, A, is notionally a P2P low-cost carrier, whereas the other, B, is a traditional, hub-based airline. Company A operates two airport pairs: from 1 to 3 – branch \( \delta \) – and from 2 to 3 – branch \( \varepsilon \). Company B operates three airport pairs: 1 to 0, 2 to 0, and 0 to 3, respectively branches \( \alpha \), \( \beta \), and \( \gamma \). Each branch represent notionally a certain number of flights operating during a given period, e.g. a day.

**FIGURE 2 — Representation of the configuration used to produce the results.**

Passengers travelling from 1 to 3 thus have two possibilities: taking a direct flight with company A, or making a connection at 0 with company B. The same applies for itineraries from 2 to 3. Other passengers begin their journey at 1 and finish at 0, as well as from 2 to 0 and 0 to 3, for which they have to take a flight with company B. We calibrate the model so that company A has a lower cost than company B, but the branches \( \alpha \), \( \beta \), \( \delta \), and \( \varepsilon \) have a relatively low demand with respect to the branch \( \gamma \). In other words, company B generates most of its revenue from one high-yield branch, with the two others feeding it, whereas A competes with direct flights from 1 to 3 and 0 to 3.

We wish to observe three effects in this system. Firstly, how the increase of demand on one branch affects the other branches. Secondly, how airlines respond to an increase in the fuel price. Lastly, how a capacity increase somewhere in the system affects the whole system. In order to do this, we run a single simulation. During steps 5-15, we increase slowly the demand on branch \( \gamma \). Then around step 90, airport 3 increases its capacity. Finally, we simulate the increase of fuel price by increasing the parameter \( c_0 \) by 20% in step 170. Note that in reality, an increase in fuel price would have a smaller impact.
on the operational cost of company B compared to company A. However, even with the same increase, both companies already react differently, as we see in the following.

We show in Figure 3 the evolution of the traffic on the different branches (top panel) and the evolution of the profit of the airports (bottom panel). Focusing first on the top panel and dismissing the transient effects, one can see different effects. Firstly, the increase of demand on the branch $\gamma$ during the first 15 steps impacts all the airlines. More specifically, the traffic on this branch increases, whereas the traffic on branches $\alpha$ and $\beta$, belonging to the same company, decreases, and the traffic on branches $\delta$ and $\epsilon$ increases. This is triggered by a simple effect: since the demand on $\gamma$ increases, the price on this branch increases. Hence, passengers willing to go from 1 to 3 with company B now have a higher total price for their ticket because of this branch. As a consequence, some of them switch to the competition, i.e. to the direct flights on branches $\delta$ and $\epsilon$.

Around step 90, airport 3 increases its capacity. The direct consequence for companies having flights landing at 3 is that their cost is decreased, because the delay at this airport decreases. One could naively think that all will benefit from this increase of capacity, or at least not lose anything. However, this is not what happens, since both branches of company A indeed have higher traffic, whereas all branches from company B see traffic decrease. Since company A is cost-driven, the increase of capacity allows it to decrease its prices from 0 to 3 and 1 to 3 to a greater extent than company B, and thus capture all of the additional traffic, and more. This mechanism also has an impact on the airports. If airports 1 and 2 see their profit increase because the traffic on branches $\delta$ and $\epsilon$ increase, airport 0 ends up with a slightly smaller profit. Whereas one could have expected a higher profit for any airport connected with another airport undergoing a capacity extension, the hub of company B actually suffers from this situation.

**FIGURE 3 — Evolution of the traffic per branch (top panel) and the profit per airport (bottom panel) during a simulation. The simulation increases the demand on branch $\gamma$ between step 5 and step 15, airport 3 increases its capacity around step 90, and the cost of all airlines are increased at step 170 by 20%.**
Finally, around step 170, the operational costs of the airlines are increased as described above. As seen in Figure 3, the different branches react differently to this increase. The most affected are branches \( \delta \) and \( \varepsilon \), which lose a sizeable share of their traffic. Branches \( \alpha \) and \( \beta \) are also affected, but \( \gamma \), on the other hand, sees its traffic increase. This happens for the same reasons as previously outlined during the capacity increase. Indeed, since this branch is notably less cost-driven than the others (because of its high-yield market), it captures some of the lost traffic on the other branches, where the prices have risen significantly.

The model is thus able to capture complex economic feedback in the system, due to the adaptation of the agents to changing conditions. This non-trivial feedback is triggered by the heterogeneity of the agents, which have different roles and different objectives, and powerful network effects.

V. Conclusions and future work

In this paper, we have presented the main foundations on which the Vista model is being built and its objectives. We have highlighted the need for a holistic approach because of the entanglement of the different sub-parts of the system, taking into account the heterogeneity of agents, their numerous interactions, and the influence of external factors. We have segregated these into background and foreground factors. These categories help Vista to focus on a few scenarios instead of having to test each factor independently. We have explained the sequential architecture of the full model, featuring three layers mappable on the strategic, pre-tactical, and tactical phases of ATM. The layers are independent simulations and may be used in other simulation engines in future. The tactical layer is already implemented and has been used in several other projects, in which it has demonstrated its capabilities.

More recently developed is the strategic layer, including the economic model that we have described in more detail. Required to be computationally light but including economic feedback, it has been designed as a network-based deterministic agent-based model in order to take into account heterogeneous behaviours in the system and network feedback. The first results from this model have been presented. Although run on a non-calibrated and highly simplified scenario, it demonstrated its capability to produce complex behaviours arising from the competing, heterogeneous agents in the system. In particular, the sensitivity to cost for airlines combined with the competition with other airlines triggers non-trivial responses when the system is disturbed.

Future work includes the full integration of the different layers into one single engine, which can produce a typical day of (future) operations, taking into account all the factors influencing these layers. We also need to improve the model by adding certain features, such as including some pro-active (instead of essentially passive) ANSPs. We are entering the final stages of stakeholder consultation, through a series of presentations and workshops. Finally, the model needs to be properly calibrated and tested against published traffic and passenger forecasts. This will allow us to move to the ultimate objective of examining the trade-offs between the stakeholder KPIs, demonstrating whether future alignment improves or deteriorates, and the underlining drivers of such behaviour.

REFERENCES


Project B.04.01, “Updated Step 1 Validation Targets - Aligned with Dataset 13,” 2014.


AN AGENT-BASED AUCTION MODEL FOR THE ANALYSIS OF THE INTRODUCTION OF COMPETITION IN ATM

J. Torres, D. Toribio, R. Marcos, O. G. Cantú Ros and R. Herranz
Nommon Solutions & Technologies
Madrid, Spain
javier.torres@nommon.es

Abstract—The provision of air traffic services has for a long time been a national monopoly. The introduction of competition in the ATM sector has been proposed as a means to incentivize the adoption of new technology and more efficient strategies. In this paper, we analyze a possible mechanism for the introduction of competition in the ATM market consisting in the tendering of licenses to operate en-route air navigation services within certain geographical areas. The license tendering process is simulated by means of an agent-based model. The model is used to investigate the potential impact of the proposed institutional design and how the outcomes of the process are influenced by different parameters of the tenders, such as the frequency of the auctions and the order in which the different areas are auctioned.

I. Introduction

The Single European Sky (SES) initiative aims to restructure the European airspace, create additional capacity and increase the overall efficiency of the ATM system, so that the European ATMs system can cope with sustained air traffic growth under safe, cost-efficient and environmentally friendly conditions. The European Commission has set ambitious goals for the SES to be reached by 2020, including a 3-fold increase in airspace capacity and a cost reduction of at least 50% in the provision of ATM services.

In this context, the question of how to provide the appropriate organizational structures, institutions and incentives for new operational concepts and technologies to yield the expected results stands high on the policy agenda. The introduction of competition has been proposed as a means to provide the right incentives for the realization of the high-level objectives of the SES, through the speed up of the innovation cycle and the fostering of more efficient operations. On the other hand, competition does not prevent every market failure (e.g., negative externalities) and, depending on market conditions, liberalization can also have undesired outcomes, such as the emergence of oligopolies or monopolies. The SESAR 2020 Exploratory Research project COMPAIR (http://www.compair-project.eu/), in which the present work is framed, investigates how to introduce competitive incentives in ATM so as to best contribute to achieving the European policy objectives for aviation.

Economic research on ATM is not abundant. Most economic studies which have analyzed economic drivers for ATM performance have focused on price as the main instrument for change [(1)]. Only a few studies have investigated different institutional approaches that have an impact on the industrial structure of the ATM sector, such as the study by Baumgartner and Finger [2] and the SESAR WPE projects ACCHANGE [3] and ACCESS [4] that studied how institutional change could affect ATM performance. In line with these studies, COMPAIR has analyzed four different institutional designs.
One of the institutional designs proposed by COMPAIR is the tendering of licenses to provide air traffic services within certain geographical areas. Auctions and tendering processes have been widely studied using game theoretical approaches [5]. However, due to the strong assumptions behind game theory, such as agents’ rationality, these models may fail to capture the complexity of the interactions between the system constituents. Additionally, the equilibrium seeking nature of game theory models limits their usefulness to study the dynamics of this type of institutional setting, in which the assignment of the different geographical areas to the distinct ANSPs is not necessarily expected to reach equilibrium.

In recent years, agent-based modelling (ABM) [6] has been recognized as a powerful tool for simulating and analyzing complex bidding environments. In this paper, we present the agent-based model developed by the COMPAIR project to simulate the tendering of ATC licenses with the aim of providing insights into what type of auction design would produce the most efficient outcomes. The rest of the paper is organized as follows: Section II describes the general logic of the simulation model and the main modelling assumptions; Section III describes the case study used to investigate the proposed tendering mechanism and the scenarios analyzed; Section IV discusses the main results of the simulations; and Section V concludes and discusses future research directions.

II. Description Of The Model

A. Overall Description

The model simulates the tendering of licenses to operate en-route air traffic services in specific geographical areas and for a certain period of time by employing the agent-based modelling paradigms. It comprises three main elements:

1. Geographical context, which provides the environment for the agents to operate in.
2. Agents. Three types of agents are considered: (i) the regulator, (ii) the ANSPs, and (iii) the airlines.
3. Exogenous variables, which represent arbitrary external conditions that affect the model but are not affected by it. The exogenous variables considered in the model are fuel prices and passenger demand.

The simulation consists of two stages:

- The first stage simulates the tendering process, where the ANSPs compete for the control of different geographical areas. In this stage, only the regulator and the ANSPs participate. Each ANSP submits a certain unit rate per service unit ($p*€/km, where $p$ is the weight factor of the aircrafts) that will be the maximum unit rate applicable in that area during the license period if that ANSP wins the tender. Contract conditions include the minimal capacity the ANSPs have to provide during the license period and the maximum market share an ANSP can handle in order to avoid monopolistic behaviors.

- The second stage simulates how agents evolve between auctions. In this stage also airlines participate. They react to the ANSPs decisions by choosing different routes according with the air navigation charges in each geographical zone. Charges are adjusted every given period of time until the license period is over.

Once the license period expires, the tendering process is repeated, which can lead to contract renewal for the incumbent provider or to a new provider. The simulation finishes when the temporal horizon is reached.
B. Modelling Assumptions

The main modeling assumptions are the following:

- ATCOs may monitor not only flights in their current charging zone but also flights in any of the charging zones controlled by the ANSP they are working at.

- At the beginning of the simulation, ATCOs working at a specific charging zone ("legacy ATCOs") will always work at the ANSP controlling their original area and maintain their labor agreement throughout the simulation (until retirement).

- New ATCOs (non-legacy ATCOs), who are hired throughout the simulation, will have the same cost regardless of their nationality and will be employed by the same ANSP during all the simulation, unless they are dismissed. The rationale behind the unitary cost is that a consequence of hiring ATCOs from any country in the EU and allowing them to work remotely is that they will have the same cost, either because they may be all hired from the same country or because the liberalization and free competition between ANSPs of all countries will lead to costs’ homogenization.

- When hiring new ATCOs, there is an initial extra cost due to training.

- When dismissing new ATCOs, there is an extra cost due to dismissal costs.

- Under same technology conditions, different ATCOs are assumed to be equally efficient regardless their experience, and ANSP they work at. The difference of productivity between ANSPs is a parameter of each ANSP simulating its level of technology adoption, and not an ATCO’s parameter.

- If an ANSP’s capital becomes negative, the ANSP goes into bankruptcy.

- New ANSPs are not allowed to enter the market.

- An average plane size, load factor and operational cost per kilometer are considered for all flights regardless of the origin-destination pair.

C. Geographical Context

The geographical context provides the environment for the agents to operate in. It is composed by: (i) a set of charging zones that the ANSPs compete to control; (ii) a group of airports representing the main destinations within the charging zones; and (iii) a collection of routes per origin-destination pair defining the possible paths the airlines can fly.

D. Agents

1. Agents’ description

   a. Regulator

      The role of the regulator is to provide and store the public data created throughout the simulation [e.g., air navigation charges for each charging zone], announce the auction parameters and select the winners of the auctions.

   b. ANSPs

      The ANSP agents are the main agents of the simulation. They make decisions to achieve their objectives according to their internal parameters, their competitors and the environment. They are modeled as profit-maximizers, but objective functions could be easily implemented, such as revenue maximization or cost minimization.
The parameters that define an ANSP are: (i) charging zones they control; (ii) human resources (number of ATCOs); (iii) financial capital. The capital available by ANSPs to invest either in hiring ATCOs, improving their technology level or to pay the cost of dismissing staff; (iv) bidding strategy. It defines the learning method ANSPs will employ to characterize their competitors’ behavior and calculate their bids and; (v) technology level.

c. Airlines

The airline agents, which represent the different airlines that fly daily over the European sky, are assumed to be cost minimizers. Their objective is to meet the total expected demand at the minimum possible cost. Operating costs other than fuel cost and fees are modeled as an internal parameter of the agent.

2. Agents’ interaction rules

The sequence of agents’ decisions and actions follows the schemes included in Figure 1, Figure 2 and Figure 3.

FIGURE 1 — Agents’ behaviour rules

![Diagram of Agents' Behaviour Rules]

a. Auctioning Process

The auctioning process is depicted in Figure 2.

FIGURE 2 — Auctioning process. Agents’ interactions

![Diagram of Auctioning Process]
The regulator announces the auction parameters, which include the minimum capacity that the winning ANSPs shall provide in each area, calculated based on the OD demand forecast and assuming that the distribution of flights per route in each OD pair will be the same as the distribution of the last periods, and allocates the auction areas to the winning bidders.

The ANSPs submit a bid corresponding to the maximum charge that would be applied to the auctioned zone.

To submit the bid the ANSPs take the following actions:

1. Calculate their total resulting market share in case of winning the auction and evaluate if this accomplishes the condition of the maximum market share allowed.

2. Determine the minimum profitability they want to achieve. This lies between a minimum and a maximum value set to 7% and 12% of the total cost of controlling the network respectively, and grows proportionally with an adaptive factor, \( \alpha \), calculated as:

\[
\alpha = \frac{M \text{ share}_{\text{ANSP}}}{\text{max } M \text{ share}} + \frac{\text{demand}_{\text{ANSP}}}{\text{max } \text{demand}_{\text{ANSP}}} \epsilon [0,2], \text{ and} = \frac{\text{min } \text{profitability}_{\text{ANSP}}}{\text{min value}} + \alpha \cdot \frac{\text{max } \text{value} - \text{min } \text{value}}{2}
\]

With \( M \text{ share}_{\text{ANSP}} \) the ANSP’s current market share, \( \text{max } M \text{ share} \) the maximum allowed market share, \( \text{demand}_{\text{ANSP}} \) the expected demand for the zones currently managed by the ANSP, and \( \text{max } \text{demand}_{\text{ANSP}} \) the maximum demand the ANSP can control with the current resources.

3. Estimate in an iterative process the best bid charge by multiplying the current charge by a bid factor, ranging from 0.5 to 1.5 in steps of 0.001. For each bid factor they: (a) estimate the resources needed according to their technology level and the expected number of flights, calculated based on the passenger demand forecast and the average plane size and occupancy rate; (b) estimate the total profit, as the difference between the expected income and cost, and the profitability, dividing the expected profit by the expected cost; (c) obtain the probability of beating their competitors. This is calculated with one of the following learning methods: Friedman [7] and Gates [8] which characterize the behavior of all their competitors and estimate the probability of winning the auction accordingly, and Fine [9] which only characterizes the pattern of the winning bids of previous auctions, (d) calculate the auction expected profit, defined as the product of the expected profit by the probability of winning the auction.

4. Finally submits the bid that maximizes the auction expected profit.

Once the regulator has allocated the areas to the winning ANSPs, they decide the amount of capital to invest during the following license period in order to upgrade their technology level, which is used as the main driver of the productivity of the ANSPs. This amount corresponds to a percentage (an 80% in this case) of the expected profit of the starting license period, regardless the characteristics and size of the controlled areas. The monetary impact of technology upgrade has been obtained from the figures of the Master Plan 2012 [10].

b. Evolutive Process

The sequence of agents’ decisions and actions follows the scheme included in Figure 3.
The ANSPs examine different combination of charges within the areas they control and, for each combination of charges, take the following actions: (i) estimate the resources needed according to the demand forecast, the charge of their competitors and the distance that each route flies over each charging zone; and (ii) calculate the expected profit of the combination of areas they control during the following time step. Based on this information, they select the combination of charges that maximizes their expected profit.

The airlines’ goal is minimizing its costs while meeting passenger demand. Once the ANSPs publish the charges of each charging zone, airlines select the route of each flight according to the cost of the route with a probability

\[
P(r = R) = \frac{e^{utility_R}}{\sum_{r} e^{utility_r}},
\]

with \(r\) running over all possible routes for a given pair,

\[utility_R = K_{OD} / cost_R\]

and \(K_{OD}\) a constant with a different value for each OD pair.

### E. Exogenous Variables

The exogenous variables considered in the model are the fuel price and the passenger demand.

Passenger demand defines a number of passengers at each simulation step and for each origin-destination pair. Forecasted passenger demand is known by all the agents. The regulator uses it to establish the minimum capacity that the ANSPs have to provide in each zone. The ANSPs employ it to establish the unit charge for each time step. Finally, the airlines use this information to set the number of flights per origin-destination pair. Actual demand is calculated by the model as a deviation from the forecasted data, by adding a stochastic noise at each step. As a result, actual demand may differ from the forecast, and the forecast values for the following simulation steps are modified accordingly (see Figure 4).

### FIGURE 4 — Example of scenario, actual and forecasted value

Similar to passenger demand, there is a forecasted fuel price profile known by the airlines and the ANSPs. As for travel demand, the actual fuel price is calculated as a
deviation from the forecast, by adding a stochastic noise to the forecasted value at each simulation step and adapting the forecast values for the following simulation steps.

The purpose of including a stochastic component is to test the ability of the different agents to adapt to changing circumstances in the presence of uncertainty.

III. Data Sources And Case Study

A. General Parameters

The proposed case study simulates the liberalization of the ATM market in Western Europe in 2015. The model has been initialized with the ANSPs’ and airlines’ data of 2014 year ended, summarized in Table 1 and Table 2 respectively.

**TABLE I — ANSPs’ en-Route Data**

<table>
<thead>
<tr>
<th>ANSP</th>
<th>Staff cost (M €)</th>
<th>Non-staff operating cost (M €)</th>
<th>Other cost (M €)</th>
<th>IFR flight-km (000 km)</th>
<th>Average charge per km (€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgocontrol</td>
<td>98.8</td>
<td>15.1</td>
<td>20.7</td>
<td>173,363</td>
<td>0.96</td>
</tr>
<tr>
<td>DFS</td>
<td>629.6</td>
<td>81.0</td>
<td>185.8</td>
<td>1,103,672</td>
<td>0.73</td>
</tr>
<tr>
<td>DSNA</td>
<td>641.8</td>
<td>190.0</td>
<td>127.0</td>
<td>1,542,050</td>
<td>0.78</td>
</tr>
<tr>
<td>ENAIRE</td>
<td>393.8</td>
<td>67.8</td>
<td>142.5</td>
<td>882,223</td>
<td>0.79</td>
</tr>
<tr>
<td>ENAV</td>
<td>295.5</td>
<td>113.7</td>
<td>146.9</td>
<td>711,039</td>
<td>0.83</td>
</tr>
<tr>
<td>IAA</td>
<td>52.9</td>
<td>20.7</td>
<td>14.8</td>
<td>214,828</td>
<td>0.55</td>
</tr>
<tr>
<td>LVNL</td>
<td>126.8</td>
<td>22.4</td>
<td>11.2</td>
<td>209,564</td>
<td>0.58</td>
</tr>
<tr>
<td>NATS</td>
<td>319.6</td>
<td>87.2</td>
<td>188.7</td>
<td>798,501</td>
<td>0.98</td>
</tr>
<tr>
<td>NAV</td>
<td>72.7</td>
<td>8.7</td>
<td>7.9</td>
<td>240,379</td>
<td>0.49</td>
</tr>
<tr>
<td>NAVIAIR</td>
<td>48.6</td>
<td>12.1</td>
<td>19.3</td>
<td>138,344</td>
<td>0.66</td>
</tr>
<tr>
<td>Skyguide</td>
<td>141.1</td>
<td>13.8</td>
<td>33.8</td>
<td>208,425</td>
<td>0.70</td>
</tr>
</tbody>
</table>

**TABLE II — Airline Data**

<table>
<thead>
<tr>
<th>Airline</th>
<th>CASK total (€ cent)</th>
<th>CASK fees (€ cent)</th>
<th>CASK fuel (€ cent)</th>
<th>CASK other (€ cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EasyJet</td>
<td>5.91</td>
<td>0.46</td>
<td>1.87</td>
<td>3.58</td>
</tr>
<tr>
<td>Air France</td>
<td>6.93</td>
<td>0.53</td>
<td>1.9</td>
<td>4.50</td>
</tr>
<tr>
<td>Lufthansa</td>
<td>8.8</td>
<td>1.47</td>
<td>1.89</td>
<td>5.44</td>
</tr>
<tr>
<td>British Airways</td>
<td>7.49</td>
<td>0.55</td>
<td>2.45</td>
<td>4.49</td>
</tr>
</tbody>
</table>

The network analyzed, presented in Figure 5, includes eleven charging zones, eleven ANSPs and a set of possible routes between the thirteen airports considered in the simulation.

**FIGURE 5 — Geographical scope and network of the case study**
All airlines are modeled as an average airline by means of a single agent that meets all the demand. The data used to model the airline have been obtained from the annual financial reports of the main European airlines ([11], [12], [13], [14]). These data are presented in Table 2.

The data of the ANSPs have been obtained from the 2014 ATM Cost-Effectiveness Benchmarking Report [15]. These data have also been employed to calibrate the number of flights per origin-destination pair since the network is a simplification of reality and the number of flights has to be adapted to this network. Given the distribution of flights per OD pair, and the distance that each route flies over the charging zones, the number of flights per OD pair has been adjusted to obtain the actual demand of 2014 in every country.

The ANSPs simulated are Belgocontrol (Belgium), DFS (Germany), DSNA (France), ENAIRE (Spain), ENAV (Italy), IAA (Ireland), LVNL (Netherlands), NATS (United Kingdom), NAV (Portugal), NAVIAIR (Denmark) and Skyguide (Switzerland), which provide air traffic services to around 60.5% of the total IFR flight-km in Europe. The figures of the Maastricht Upper Airspace Control Centre (MUAC) have been split into these countries and allocated to the corresponding ANSPs. The parameters of each ANSP are summarized in Table 1.

Passenger demand forecast has been obtained from EUROCONTROL’s report “Challenges of Growth 2013” [16]. We have employed the data of the most likely scenario, the so-called “Regulated growth”, which considers that the demand will grow 1.8% annually. Since the data provided for demand growth is aggregated at a regional level, demand growth had to be assumed homogenous among the EU countries. According to this scenario, the demand in 2050 would double the current demand.

B. Scenarios

To analyze the influence of different auction parameters, several scenarios are built:

1. Market Share

The market share is calculated as the flight-km controlled by an ANSP divided by the total number of flight-km in the network. A maximum allowed market share is set to avoid the appearance of monopolistic or oligopolistic behaviors.

We analyze two different values of the maximum allowed market share: 40% and 60%. These values ensure the existence of at least 3 and 2 ANSPs, respectively.

2. Auctioning Order

In the simulation, the charging zones are auctioned individually and sequentially in the same time step. Thus, the order in which they are auctioned has an influence on the results (e.g., due to the limit imposed on the market share, it may occur that an ANSP cannot bid for some area if it has been previously allocated other areas).

We analyze the outcome of auctioning the areas in different orders according to the size of each national market:

- Ascending: Denmark, Belgium, Switzerland, Netherlands, Ireland, Portugal, Italy, United Kingdom, Spain, Germany, France.
- Descending: France, Germany, Spain, United Kingdom, Italy, Portugal, Ireland, Netherlands, Switzerland, Belgium, Denmark.
- Mixed: Denmark, France, Belgium, Germany, Switzerland, Spain, Netherlands, United Kingdom, Ireland, Italy, Portugal.
3. **License duration**

The license duration determines the frequency of the auctions. The larger the license duration, the fewer auctions will take place within the simulation. If only few auctions occur, ANSPs do not have enough data to properly analyze the bidding behavior of their competitors and adapt their own behavior.

Two different values of the license duration are analyzed: 5 and 10 years.

Since the aim of this paper is to analyze the impact of the different auction parameters and not to compare the effectiveness of the bidding strategies, all ANSPs employed the same one (Gates model) for all scenarios described before.

**IV. Analysis of Results**

**A. Market Share**

The maximum market share parameter has a very significant influence on the outcome of the tendering, especially on the distribution of charging zones being controlled by each ANSP.

As expected, for a maximum market share of 40%, we find more market competition between ANSPs than with a maximum market share of 60% (see Figure 6). In the first case, two big ANSPs control almost 40% of the market each and two or three ANSPs control minor areas (Figure 6.a). On the contrary, when the market share is set to 60%, there is a dominant ANSP whose market share tends to increase in every tendering process controlling more than 50% of the market at the end of the period of study. Moreover, in this scenario the whole market is controlled by fewer ANSPs (Figure 6.b).

*FIGURE 6 — ANSPs’ market share: a) maximum market share set to 40%; b) maximum market share set to 60%*
The maximum market share does not seem to affect the trend followed by the evolution of the charges and the total number of ATCOs in the network (Figure 7 and Figure 8). In the case of a market share of 40%, the average charge obtained in 2050 is 38 €cents/km, 10% greater than in the case of a maximum market share of 60%, and the total number of ATCOs is 15% higher. This is due to the investment in technology made by ANSPs. In the 60%-scenario the total profit is divided by a fewer number of ANSPs, hence they have more money to invest in technology and increase their efficiency to a greater extent than in the 40%-scenario.

**FIGURE 7 — Average network charge**

![Average network charge](image1)

**FIGURE 8 — Total number of ATCOs**

![Total number of ATCOs](image2)

In spite of these advantages, a maximum market share of 60% could lead to an oligopoly in which the market is dominated by two ANSPs which control over 90% of the market, with a tendency to increase this percentage. The emergence of oligopolies is an undesired outcome of the liberalization of any market. Thus, a maximum market share over 50%, although presenting some minor benefits in the short-term period, could lead to oligopolistic behaviors in the long-term. When limiting the market share to 40%, the market is consolidated into four ANSPs out of eleven, which seems a more appropriate number of players to ensure real competition.

**B. Auctioning Order**

The auctioning order influences locally the charging prices resulting from the tendering but it has a minor impact on the global outcome.

Figure 9 presents the resulting charges obtained in each country for different auctioning orders in a scenario with a maximum market share of 40%. It may seem that, in the “descending” order (Figure 9.a), the total fees the airlines will have to pay are greater than in the other scenarios. However, the average network charge paid by airlines (considering the flight demand over each country) is quite similar for the three options (Figure 9.d). The reason is that in the “descending” scenario the biggest countries are auctioned first and the ANSPs behave more aggressively offering lower charges. Finally, when the smaller areas are auctioned, the dominant ANSPs have ensured a high market share for the following license period, in some cases close to the maximum market share, and they are not allowed to participate or they are not interested in tendering.
for these areas unless they could obtain a great profit. Then, the less efficient ANSPs have a chance to be allocated one of the small countries, offering a higher charge. The same effect occurs with the latest zones to be auctioned in the “ascending” (Figure 9.b) and the “mixed” order (Figure 9.c), but to a minor extent. In the three scenarios, it is observed that the last zone to be auctioned gets the highest charges (Denmark in the “descending” scenario, France in the “ascending” scenario and Portugal in the “mix” scenario), with differences in charges specially marked in the “descending” one.

Comparing Figure 9.a, Figure 9.b and Figure 9.c, we can conclude that the mixed ordering produces more homogeneous charges between the different countries.

**FIGURE 9 — Influence of the auctioning order in the charges for a maximum market share of 40%**
C. License Duration

The last parameter we evaluate is the frequency of auctions. Two scenarios have been evaluated: (i) a 5-year license duration; and (ii) a 10-year license duration. For both scenarios, the maximum allowed market share was set to 40% and the auctioning order to “mixed”. The results are depicted in Figure 10 and Figure 11.

**FIGURE 10 — Influence of the licenses duration in the charges for a maximum market share of 40% and “Mix” auctioning order**

**FIGURE 11 — Influence of the licenses duration in the ANSPs’ market share for a maximum market set of 40% and “Mix” auctioning order**
The charges in both scenarios tend to decrease at the same rate. Also, the rates in 2050 are almost the same for both scenarios (Figure 10). Since the charges fall at the same rate, the average bid factor of the winning bids lowers as the frequency of auctions decreases.

There is a considerable difference in the resulting market share of the ANSPs for the two scenarios. In the 10-year scenario (Figure 11.a) the market remains stable from 2035 to 2050. Five ANSPs control the whole market, having 4 of them a market share over 15%, which suggests a very competitive scenario. In the 5-year scenario, the ownership of the charging zones switches after every tendering process (Figure 11.b). Two dominant ANSPs control the 40% of the market each, the maximum they are allowed to. The remaining zones are shared by two minor ANSPs. These results would suggest that a license duration of 10 years would lead to a more stable and competitive market.

D. General Outcome

An important outcome of all the scenarios tested is that, in general, the ANSPs which control the biggest charging zones at the beginning of the simulation (the ANSPs with the highest market share on the first period) perform better in the long term, since they have more resources to invest at the beginning of the simulation. On the contrary, the smallest ANSPs usually disappear between the second and the fifth auction as they are not competitive enough against the dominant ANSPs.

It is also noticeable that when there is a dominant ANSP that controls a big part of the market, due to its investment capacity and the economies of scale, e.g., reallocating ATCOs to different charging zones according to the labor requirements, both the total number of ATCOs and the average charge are a bit lower than when the market is controlled by more ANSPs. However, what seems a clear benefit in the short/medium-term may lead to the emergence of an oligopoly in the long-term.

V. Summary And Future Research

In this paper, we have presented an agent-based model designed to investigate the impact of a hypothetical tendering of licenses to operate air traffic services within Europe. The model simulates the behavior of a group of ANSPs which compete for the control of different charging zones to maximize their profit and a set of airlines that aim to meet the passenger demand while minimizing their costs. The ANSPs have been endowed with learning and adaptive behaviors, i.e. ANSPs use historical data to devise a strategy and they have the capability to adapt their strategy to respond to new conditions, aimed to calculate the bids according to their actual status and the previous bids of their competitors.

We have illustrated the potential of the proposed approach to analyze the dynamics and the final outcome of the process by exploring the influence of different auctioning parameters, namely the frequency of auctions, the maximum market share established by the regulator, and the order in which the charging zones are auctioned. The results allow us to derive useful insights about the criteria to be taken into account for such type of institutional framework.

Several model enhancements are currently being implemented and will be used for future studies:

- Different investment strategies could be implemented so that the ANSPs select the amount to invest on technology depending on their status and the environment conditions.
More complex and realistic scenarios will be modeled, such as scenarios considering uncertainty in the exogenous variables. This will allow us to study the adaptability of the ANSPs to changing and unexpected conditions with different degrees of volatility, and to measure the ability of different institutional designs to provide the required level of resilience and adaptability.

More airline agents empowered with learning capabilities will be included, in order to have a more realistic representation of airline behavior. This will allow us to take into account the cost of congestion and the daily distribution of flights.

ANSP behaviors other than profit maximization will be implemented, e.g. to explore the potential impact of anticompetitive practices.

The possibility of ties and merges between ANSPs will be explored.

The simulation scenario will be extended to the whole ECAC area.

Finally, simulations will be conducted to compare the outcome obtained with different type of auctions. In particular, we will compare the single-unit auction described in this paper with a combinatorial auction in which all the areas are tendered at the same time and ANSPs bid for different combination of charging zones, in order to investigate the trade-offs between the economies of scale offered by the combinatorial auction versus the presumably more effective learning process enabled by a sequence of single-unit auctions.

Acknowledgment

The project leading to these results has received funding from the SESAR Joint Undertaking under grant agreement No 699249 under European Union’s Horizon 2020 research and innovation program.

REFERENCES

OWNERSHIP FORM & AIR NAVIGATION SERVICE PROVIDER PERFORMANCE

Nicole Adler, Adit Kivel
Hebrew University of Jerusalem, Israel
Jerusalem, Israel

Eef Delhaye, Stef Proost
Transport & Mobility Leuven & KU Leuven
Leuven, Belgium

Abstract—The ownership form of Air Navigation Service Providers varies across countries ranging from state agencies, to semi-private firms with for-profit or not-for-profit mandates. This research focusses on the link between the performance of ANSPs and their ownership form. A theoretical economic model suggests that effort to achieve efficiency will be higher in the case of public companies with a board of stakeholders composed of airspace users and in the case of private companies in which stakeholders are also shareholders. A stochastic frontier analysis estimation of the production and cost functions of 37 European air navigation service providers over nine years suggests that the public-private ownership form achieves statistically significantly higher efficiency levels compared to a governmental corporation which in turn is an improvement over a state agency.

I. Introduction

Air traffic control provision is one of the last elements of the aviation supply chain to be considered for liberalization. In the United States, where the Federal Aviation Administration serves the entire market as a single government agency, there has been a long discussion as to whether there is a need to commercialize or privatize the service ([24][25]). In Europe, the fragmentation of service provision, the home bias of each member state for the national provider, the monopolistic nature of some of the air traffic control services, the network component of most services and the split incentives which require the service providers to invest in new technology without enjoying the direct benefits neither encourage cost nor productive efficiency in Europe ([1] and [6]).

With respect to other industries, [2] analyze the combined impact of ownership form, economic regulation and competition on airport performance using data envelopment analysis. The empirical results suggest that in the absence of competition, public airports operated less cost efficiently than fully private airports. In a competitive setting, public and fully private airports operate equally efficiently, however private airports set higher aeronautical charges. In an industry in which there is no competition given the current geographical monopoly status of the ANSPs, it is unclear whether a public or private ownership form would stimulate innovation and create a more productive sector [3]. On the one hand, private firms with access to financial markets may have greater interest in cost efficiency. On the other hand, public firms may reduce the level of information uncertainty; information which is required in regulating such firms. [23] focus on the choice of public versus private provision of goods and services as a function of transaction costs. One of their conclusions is that neither public nor private provision can fully resolve incentive problems that arise from imperfect information. [15] develop a model in which a provider chooses to invest in improving the quality or reducing the
costs of a specific service. The results of the model suggest that the case for privatization is stronger when quality-reducing cost reductions can be controlled through contract or competition, when quality improvements are important, and when patronage and powerful unions are a problem. Hence, there would seem to be a basis for arguing that there is a relationship between performance and ownership form.

In this research, we develop in section II an economic model in order to analyze the ANSP market and the potential impact of moving from a government agency to a more commercialized setting. Next, in section III, this model is tested empirically for the European air navigation providers by estimating econometrically both production and cost functions and their relationship with ownership form. Section IV draws conclusions.

II. Economic model

In this section we develop an economic model to understand the possible links between performance, regulation and ownership form. For this analysis we extend the theoretical model presented in [10] explaining the efficiency efforts of a regulated monopoly as a function of the objective of the monopolist and the regulatory framework in place. We assume that the objective of an ANSP is likely to draw from three underlying interests, namely maximization of consumer surplus (CS) of the airlines (and indirectly passengers) with weight parameter $\gamma^1_{ANSP}$, maximization of profits ($\pi^\text{ANSP}$) with weight parameter $\gamma^2_{ANSP}$ and national interest (NI) with weight parameter $\gamma^3_{ANSP}$. The national interest represents two factors: first the benefits of the union of ANSP personnel under the form of higher wages and more relaxed working conditions and second the national manufacturers of air traffic control equipment. This leads to the ANSP mixed goal function of firm $i$ presented in [1].

$$\text{Goal}^\text{ANSP} = \gamma^1_{ANSP} \text{CS} + \gamma^2_{ANSP} \pi^\text{ANSP} + \gamma^3_{ANSP} \text{NI}$$

(1)

In contrast to Blondiau et al. (2016), the weights now also depend on the ownership form of the ANSP. Multiple assumptions are possible including (1) a public company ANSP public could strive for socially optimal decisions such that the sum of consumer and producer surplus are maximized, $\gamma^1_{ANSP} = \gamma^2_{ANSP} = 0$; (2) a public company may attach a higher value to NI as a result of lobbying or fraud $\gamma^3_{ANSP} > 0$; or (3) a private company ANSP private could be influenced by the type of shareholders. Depending on the shareholder composition, a higher weight may be placed on consumer surplus $\gamma^1_{ANSP} > \gamma^2_{ANSP}$ (e.g. when airlines are represented on the board) or on profit $\gamma^1_{ANSP} < \gamma^2_{ANSP}$ (e.g. when pension funds are shareholders). The same argument may also hold true for public companies in which the consumers are represented on the board.

We assume that the production costs to provide air navigation services can be broken down into three components; a fixed ANSP cost per flight-km controlled $a$, an imperfectly observable cost component $\theta$ that varies as a function of the complexity of the airspace managed and differences in operational practices and an imperfectly observable cost reduction potential $e$ or efficiency expressed in average costs per flight-km. This leads to the ANSP cost per flight-km $c(e)$ controlled expressed in [2].

$$c(e) = a + \theta - e$$

(2)

The ANSP operating costs are expressed in [3] in which $D$ represents the total number of standardized flights.

$$OC_{ANSP} = D \cdot c(e) = D \cdot (a + \theta - e)$$

(3)

For the management and personnel of the ANSP, effort $e$ is costly in terms of stress and longer hours but such costs are not represented in the accounting system. We represent
this subjective cost as a quadratic function, $SC$, defined in (4), which means that exerting more effort becomes increasingly costly. We further assume that the costs of effort are higher for relatively larger ANSPs, hence we include the demand parameter $D$ to represent the scale of operations.

$$SC(e) = D \cdot \frac{e^2}{2} \tag{4}$$

The ANSP also receives an income, which depends on the regulated charge permitted. Current SES II regulation is influenced by both price-cap ($p_{cap}$) and cost-plus ($p_{cost+}$) regulatory approaches. Under cost-plus regulation, the ANSP charges are equal to the total accounting cost divided by traffic served plus a cost mark-up on capital which allows ANSPs to make a small profit. Under a price-cap, charges are determined by expected costs and demand. Cost efficiency incentives are very different in the two systems. In a pure cost-plus system, all costs are covered so incentives to make large efforts to reduce costs are low. In a price-cap system, any average cost realization below the price cap becomes a profit. Hence we use the general form for price-cap and cost-plus regulation as shown in (5). The charge depends on the weights given to the two types of regulation. The level of effort also plays a role. We use a static formulation here where the realization of cost for an individual ANSP does not affect the price-cap of that ANSP in the future years. Otherwise there will be strategic behavior by each ANSP and the price-cap will be less efficient because too much effort by one ANSP will have a negative ratchet effect on the price-cap of that ANSP. The price-cap is changed over time but it is a function of the aggregate performance of the ANSP’s in Europe and the change is not individualized per ANSP.

$$p_{\text{charge}}(e) = (1-B)p_{\text{cap}} + Bp_{\text{cost}} + E(tot\text{ cost}) \cdot B + \frac{tot\text{ cost}}{D} = A + Bc(e) \tag{5}$$

In the second line of (5), A stands for the first term that is constant and exogenous because it is the cost and demand expected by the regulator that is used for the price cap, while only the second term $[Bc(e)]$ is influenced by the ANSP.

For this analysis, we use two additional assumptions. First, we assume that A and B are given, this means that the price cap and the mix of price cap and cost plus regulation is given. Second, we assume that the national interest groups prefer the status quo as they were well served in the period before the change in European regulation. Assuming national interest was historically the main ANSP incentive, we have set the importance of national interest proportional to the costs of efficiency effort. This reflects the idea that adding consumer surplus incentives and profit incentives on top of the national interest will require additional efficiency efforts. $\phi$ is introduced to interpret $\gamma_{ANSP}$ as a share of the actual costs in (6).

$$\gamma_{ANSP} \cdot NI = -\gamma_{ANSP} \cdot SC(e) = -\gamma_{ANSP} \cdot D \cdot \frac{\phi \cdot e^2}{2} \tag{6}$$

Applying the two assumptions, we derive the efficiency effort $e$ that is optimal from the point of view of the ANSPs, assuming fixed demand $D$, by differentiating the objective function 7 derived from equation 1 with respect to efficiency efforts $e$ and applying equations 5 and 6:

$$\text{Goal}^{ANSP} = \gamma_{ANSP} D(\max(p_{\text{max}} - p_{\text{charge}}) + \gamma_{ANSP} \left[ D(\max(p_{\text{charge}} - c(e)) - SC(e)) - \gamma_{ANSP} SC(e) \right]$$

where the change in consumer surplus equals the difference between the maximum price (the price cap $p_{\text{cap}}$) and the price actually set $p_{\text{charge}}$.

Consequently, (8) estimates optimal ANSP efficiency effort as follows.
Based on equation 8, we find that effort is increasing in the weight attached to consumer surplus ($\gamma_{ANSP}^1$). Airlines, the consumers of ANSPs, have a strong interest in lower costs. Hence if the ANSP places a higher weight on its’ consumers, it will have a stronger interest in reducing costs and exerting efficiency efforts. Effort is decreasing in the weight attached to national interest ($\gamma_{ANSP}^3$). If there is a strong home bias, for example towards local intermediate good suppliers, or if there is a strong labor union lobby, the ANSP is less interested in reducing its cost. This is to the benefit of the home suppliers and labor lobby. The influence of the weight attached to profit on effort depends on how close the price regulation resembles a price cap. The effort is highest in the case of a pure price cap, but decreases when cost plus is applied too (representing a higher $B$ value in [5]). We now return to the role of ownership. If state agencies care more about national interest (high ($\gamma_{ANSP}^3$) coefficient) then the effort level in 8 will be lower than when a government corporation has consumers on the board: a high $\gamma_{ANSP}^1$ weight will increase the cost reduction effort. If the private firm is controlled by private shareholders, its main interests are profits (high $\gamma_{ANSP}^2$) and if the price-cap is weak, the firm will invest effort in achieving efficiency but not necessarily low prices: a monopolist prefers a high price when demand is not elastic. A government corporation with airlines on its board may be as productive and cost efficient as a private firm but this will be translated into lower prices and higher consumer surplus rather than high profits. This means that reality may be more complex than the simple public/private classification of ANSP’s: the type of price regulation as well as the ownership structure matter for the efficiency incentives. As price regulation is the same for all ANSPs, it is of interest to check if performance is indeed a function of ownership form. We focus on this question in the next section.

III. Econometric estimation of the cost and production functions of ANSPs

In this section, we conduct an econometric study in which we analyze ANSP data mainly drawn from the Performance Review Unit’s air traffic management cost-effectiveness (ACE) reports. The inputs consist of labor, capital and non-staff operating inputs, the outputs consist of total flight hours controlled en-route and IFR airport movements.

We build on earlier literature in the econometric cost-efficiency benchmarking of ATM in Europe including [26] with earlier contributions by [17] and [18]. We extend the previous studies in a number of ways. First, we have collated the newest performance data that has become available since the previous studies but removed the oldest data because of changes in the data collection procedures, thus the dataset spans the years 2006 to 2014 inclusively. Second, we estimate two cost and two production functions, per en-route and per terminal control. Previous studies estimated a joint cost function for en-route and terminal provision, known as gate-to-gate provision, utilizing an aggregate output measure referred to as ‘composite flight hours’. However, the aggregation of en-route flight hours and terminal movements is somewhat artificial and relatively crude. The goal is to reduce potential bias due to variation in boundaries between en-route and terminal activities among ANSPs. However, the composite flight hour measure may also suffer from bias as it rests on the accuracy of aggregate costs at the European level. Previous studies (e.g. [22]) have documented that significant bias may also exist in the composite flight hour measure due to the existence of cross-subsidization between en-route and terminal control activities. Consequently, unlike previous econometric benchmarking studies, we estimate the activities separately. This does come at the cost of a less reliable cost break down with respect to the two activities. Furthermore, we estimate both productivity and cost functions whereas only the latter has been published to date. The economic theory underlying the estimation of a cost function relies on the
assumption that firms minimize costs subject to the available technologies. However, this may be less relevant for ANSPs because, despite a large majority being corporatized public entities, they are also statutory monopolies and up until 2009 were operating under a full cost recovery regime. The price cap incentive regulation in place since 2010 is set at the European level and appears to have political issues in setting strong price caps, suggesting that the impact has been weak [6]. Therefore, it could be argued that most ANSPs face relatively weak incentives to ensure an efficient use of inputs during the period considered in this analysis.

This section is structured as follows. In section A, we present the methodological modelling approach relevant to analyze the air traffic control market. In section B, we discuss the dataset and the approach taken to construct the variables for the cost and production functions. Finally, in section C, we present the results of the estimations.

A. Stochastic frontier analysis

The model published in [4] analyzes panel-data, which accounts for potential heteroscedasticity and includes explanatory variables in the inefficiency distribution. The production model in [4] defines inefficiency as in equation (9) and output as in (10). In these equations $y_{it}, x_{it}$ represent the output and the exogenous explanatory variables for ANSP $i$ in year $t$. The inefficiency term $u_{it}$ is half normal distributed and positive with mean $z'\delta$. The error term is $v_{it}$:

$$u_{it} = N\left(z'\delta, \sigma_{u}^2\right) \tag{9}$$

$$E\left(\ln y_{it}\right) = \beta_0 + \sum_{n} \beta_n \ln x_{itn} + E\left(v_{it}\right) - E\left(u_{it}\right) \tag{10}$$

$$= \beta_0 + \sum_{n} \beta_n \ln x_{itn} - \left\{ z'\delta + \frac{z'\delta}{\sigma_{u}} \Phi \left( \frac{z'\delta}{\sigma_{u}} \right) \right\} \tag{11}$$

where $\phi()$ and $\Phi()$ are the density and cumulative distribution functions of the standard normal variable respectively. We apply the same model to estimate a Cobb-Douglas cost function, which represents a log-linear relationship between cost, input prices, output level and exogenous drivers$^3$. The relationship can be written as specified in (11).

$$\ln \left( \frac{E_{it}}{w_{it}} \right) = \beta_0 + \beta_y \ln y_{it} + \sum_{n, k} \beta_n \ln \left( \frac{w_{itn}}{w_{itk}} \right) + v_{it} + u_{it} \tag{11}$$

where costs $E_{it}$ are logarithmically transformed. The explanatory variables $W_{itn}$ are normalized and logarithmically transformed factor prices $k$ per unit $i$ per year $t$ and the output level is $y_{it}$. The explanatory variables should be uncorrelated with the error term as they are determined exogenously to the production and cost relationships. The error term is decomposed into a noise term $v_{it}$ and an inefficiency term $u_{it}$. The noise term is usually assumed to be random with zero mean, whereas the inefficiency is strictly non-negative and assumed to follow a half-normal, truncated-normal or exponential distribution.

$^3$ The advantage of the Cobb-Douglas specification is its duality property and simplicity. Furthermore, since all the models proved to be statistically significant, there was no need to move to the more flexible translog function. The functions are also useful for defining the air traffic control function in ongoing work modeling air navigation service provision within a game theoretic framework.
In order to estimate the en-route air traffic control production function we solve (12) and (13) simultaneously.

\[
\ln(\text{IFR flight hours}_i) = \beta_0 + \beta_1 \ln(\text{ATCO}_i) + \\
\beta_2 \ln(\text{sectors}_i) + \beta_3 \ln(\text{seasonality}_i) + \\
\beta_4 \ln(\text{complexity}_i) + V_{it} - U_{it}
\]

(12)

\[
U_{it} = \delta_1 \ln(\text{complexity})_i + \delta_2 \text{ ownership[corp]}_i + \\
\delta_3 \text{ ownership[agency]}_i + \tau_{it}
\]

(13)

where \(i\) refers to the \(i^{th}\) ATC provider; \(t\) represents the year of the observation; \(\ln\) represents a natural logarithm; \(V_{it}\) represents identical and independent error terms with a normal distribution \(N(0, \sigma^2)\); \(U_{it}\) represents the inefficiency term in the form of a truncated normal distribution with mean \([z, \delta]\) as in (9) and is a function of environmental variables (complexity and ownership form); \(\tau_{it}\) is a random variable defined by the truncation of the normal distribution (with a mean of zero and constant variance). \(U_{it}\) is expressed without an intercept which means that there is no constant element of inefficiency that is identical for all units at all times given the level of heterogeneity. The estimates for the terminal data will be similar but with the appropriate variables as displayed in table 2.

B. European ANSP dataset

We derive most of the data from the air traffic management cost-effectiveness benchmarking reports, which contain information on ANSP costs and revenues each year, reported separately for en-route and terminal control. They also report the output measures including instrumental flight rules (IFR) controlled in kilometers and in hours en-route and movements around airports. Detailed input components include annual employment costs for air traffic controllers (ATCO) and support staff, the hours worked in air control centers, towers and approach centers and the net book value of fixed assets on the balance sheet. Airspace characteristics reported per ANSP include the maximum number of en-route sectors, traffic density, seasonality (equal to traffic levels in the peak month divided by average monthly traffic), size of airspace in square kilometers and traffic complexity. The complexity index represents an aggregate of structural complexity (derived from vertical, horizontal and speed interactions) and adjusted density. Indicators related to institutional settings include the form of ownership with a distinction between a state agency [AGENCY], a government-owned corporation [CORP], or a public-private joint venture which is the default in equation 13. Relevant economic indicators include the purchasing power parity index, intermediate goods and energy price index, exchange rates and inflation rates.

Data quality is an important element of the statistical analysis. Many of the numbers were collected manually from annual reports which increases the probability of errors. In addition, there may be inconsistencies in the numbers reported for one ANSP over time. In a few instances, this is caused by a change in the construction of the indicator. We conducted checks on the evolution of all relevant indicators per ANSP and applied corrections where necessary based on the imputation technique, with linear interpolation of values for one variable based on the evolution over time for another variable\(^5\). We found errors in the reports and have corrected them accordingly. We note that from 2006 to 2008 and in 2010, the number of flight kilometers published in the reports is defined as ‘distance’ whereas other years utilize flight km. The ‘distance’ variable was incorrect for MUAC, Germany, Belgium and the Netherlands due to double counting. We note that the IFR airport movements reported for Greece in 2014 is three times higher than in 2013 which could represent an error. Finally, new variables were added to the reports from 2010, including seasonality. We assume that the 2010 values remained consistent

\(^4\) In general, most European ANSPs fall under government-owned corporation.

\(^5\) For example, evolution of "staff cost in en-route control" for Finavia is imputed using interpolation based on the evolution of "total cost in en-route control" for Finavia.
in the earlier years. In addition, we assume that the maximum number of sectors remains constant. We also dealt with missing data through imputation based on linear interpolation of values for the same variable in neighbouring ANSPs (or countries)\(^6\). After performing these checks, we obtain a representative panel dataset of 37 ANSPs covering nine years (2006-2014), with no drastic jumps or structural breaks over the years. The panel is close to being balanced although ARMATS (Armenia) is missing for the years 2006 to 2008. The dataset is available from the authors for purposes of replicability.

From the dataset, we construct a number of indicators that are applied in the SFA as listed in Tables 1 and 2.

**TABLE I — Variables in stochastic frontier cost function**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>total cost/cost of operation index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent Inputs</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>total IFR flight hours controlled (en-route) and total IFR airport movements (terminal)</td>
</tr>
<tr>
<td><strong>Labor</strong></td>
<td>(total staff cost/ATCO hours)/ cost of operation index</td>
</tr>
<tr>
<td><strong>Capital</strong></td>
<td>((depreciation cost + cost of capital)/NBV/capital goods price index)/cost of operation index</td>
</tr>
<tr>
<td><strong>Environmental variables</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Airspace characteristics</strong></td>
<td>seasonality, complexity</td>
</tr>
<tr>
<td><strong>Ownership form</strong></td>
<td>governmental agency, corporation, public-private firm</td>
</tr>
</tbody>
</table>

where the cost of operation index = \( \frac{\text{intermediate goods and energy price index}}{\text{PPP}} \), \( \text{PPP} = \frac{\text{purchasing power parity}}{\text{exchange rate}} \) and NBV=net book value.

In order to ensure comparability, monetary indicators are standardized using purchasing power parity and a cost of operation index. Standardization ensures that the econometric cost function is homogeneous and in alignment with the underlying economic theory on production and cost functions [13].

**TABLE II — Variables in stochastic frontier production function**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>En-route</th>
<th>Terminal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total IFR flight hours controlled</td>
<td>total IFR airport movements</td>
</tr>
<tr>
<td><strong>Independent Inputs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Labor</strong></td>
<td>ATCO hours in air control centers</td>
<td>ATCO hours in approach centers and towers</td>
</tr>
<tr>
<td><strong>Capital</strong></td>
<td>maximum number of en-route sectors</td>
<td>( \text{NBV/capital goods price index} \times \text{PPP} )</td>
</tr>
<tr>
<td><strong>Environmental Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Airspace characteristics</strong></td>
<td>seasonality, complexity</td>
<td></td>
</tr>
<tr>
<td><strong>Ownership form</strong></td>
<td>governmental agency, corporation, public-private firm</td>
<td></td>
</tr>
</tbody>
</table>

Finally, we apply a logarithmic transformation to all continuous variables because of the log-linear characteristic of the Cobb-Douglas models.

**C. Estimation of stochastic frontier cost and production functions**

We implement the estimation in STATA, using the tailor-made SFPANEL package [7]. We tested a number of alternative specifications including SFA with time decay in the inefficiency term [5], SFA with exogenous drivers affecting the distribution of the

\(^{\text{6}}\) For example, we impute missing values on “cost of capital” for Croatia, based on observations in Serbia and in Slovenia.
inefficiency term \( \delta \) and the true fixed effects model with time-variation in the inefficiency term and unit-specific intercepts \( \beta \). We only present the results of \( \delta \) specification as this model provided the most promising estimations, although none were materially different. We estimate all models with robust standard errors to account for possible heterogeneity in the noise error term despite the increase in estimated standard errors and reduction in the statistical significance of the results obtained.

In Table 3 we present the results of the stochastic production and cost functions for en-route operations and in Table 4 we present the equivalent for terminal operations. Each of the SFA production and cost estimates in Tables 3 and 4 include two models. The first model does not limit the average distribution of the inefficiency. When such a model was not able to explain the inefficiency (\( \sigma_u \) was not significant), we include explanatory variables to describe the mean of the distribution of the inefficiency. The \( \sigma_u \) and \( \lambda \) in Models 1 are usually insignificant hence the complexity and ownership variables are clearly an important element in explaining ANSP inefficiency levels (except for the analysis of the terminal production function in which \( \sigma_u \) of model 1 is significant).

**TABLE III — En-route frontier cost and production functions estimates**

<table>
<thead>
<tr>
<th>Elasticities</th>
<th>( \beta_1 ) x1 (Total IFR flight hours controlled)</th>
<th>( \beta_2 ) x2 (Labor cost)</th>
<th>( \beta_3 ) x3 (Capital cost)</th>
<th>( \beta_4 ) Z1 (Seasonality)</th>
<th>( \beta_5 ) Z2 (Complexity)</th>
<th>( \beta_6 ) Z3 (Ownership agency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.919 ** 0.016</td>
<td>0.385 ** 0.035</td>
<td>0.216 ** 0.021</td>
<td>1.379 ** 0.192</td>
<td>0.700 ** 0.153</td>
<td>0.080</td>
</tr>
<tr>
<td>SE</td>
<td>0.905 ** 0.018</td>
<td>0.417 ** 0.041</td>
<td>0.218 ** 0.022</td>
<td>1.686 ** 0.214</td>
<td>0.153</td>
<td>2.463</td>
</tr>
<tr>
<td>Enroute, cost</td>
<td>( \sigma_u )</td>
<td>( \sigma_v )</td>
<td>( \lambda )</td>
<td>Log Likelihood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate</td>
<td>0.296 ** 0.025</td>
<td>0.181 ** 0.022</td>
<td>1.633 ** 0.041</td>
<td>-97.510</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>0.296 ** 0.025</td>
<td>0.181 ** 0.022</td>
<td>1.633 ** 0.041</td>
<td>-57.280</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enroute, production</th>
<th>( \sigma_u )</th>
<th>( \sigma_v )</th>
<th>( \lambda )</th>
<th>Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>3.723</td>
<td>0.271 ** 0.029</td>
<td>13.745</td>
<td>-150.271</td>
</tr>
<tr>
<td>SE</td>
<td>25.244</td>
<td>0.142 ** 0.019</td>
<td>25.237</td>
<td>-59.249</td>
</tr>
</tbody>
</table>

A **/*** next to coefficient indicates significance at the 5%/1% level.

A positive efficiency score parameter estimate shows that the variable has a negative effect on efficiency.

All variables in the Cobb-Douglas functions proved highly significant across all models. With respect to output, it is clear that there are small economies of scale ranging from 10 to 15%. In the cost analyses, labor is significantly more important than capital which represents their proportions in the total cost functions. The environmental variables are also highly significant and with the expected signs. Seasonality and complexity both increase costs as expected. However, complexity both increases costs but also reduces inefficiency. We assume that additional complexity would appear to require a consistent and professional management that is better able to utilize labor resources. Furthermore, it would appear that the public private partnership model creates substantial incentives, since the government ownership form variables decrease efficiency levels. This seems to suggest that under government ownership a relatively high weight is placed on national interest, such as local suppliers and labor unions. This is confirmed by analysis focusing specifically on the role and preferences of unions (see [9]). The agency variable represents ANSPs that in general belong to the Department of Transport or Civil Aviation Authority and are the most directly connected to the government.
Based on the results of Models 2 of the en-route analyses, Fig. 1a and 1b present average production and cost efficiencies for the 37 countries over the nine years of analysis, and Fig. 2a and 2b present the average production and cost efficiencies per ANSP.

**FIGURE 1A — Average production efficiency for en-route ANSPs from 2006 to 2014**

Figure 1a suggests that the efficiency estimates gradually improve from 0.4 to 0.55 with a dip in 2009 due to the financial crisis which reduced air traffic movements substantially. Efficiency scores in the cost analysis of Figure 1b are also slightly higher, ranging from 0.52 to 0.65. Figures 1a and 1b therefore indicate that cost efficiency trends over time are positive although still lie at around 40% inefficiency on average by 2014. This means that the average ANSP is 60% less production efficient than the best performing ANSP and 45 to 40% less cost-efficient than the best performing ANSP. On the other hand, the averages mask large, statistically different estimates across the ANSPs, as presented in Fig. 2a and 2b.

**FIGURE 1B — Average cost efficiency estimates for en-route ANSPs from 2006-2014**

**FIGURE 2A — Average production efficiency estimates per en-route ANSP**

**FIGURE 2B — Average cost efficiency estimates per en-route ANSP**
When comparing efficiency levels across ANSPs, as presented in Figures 2a and 2b, we see that the efficiency levels of ten of the ANSPs lie above 0.7 with MUAC, NATS and SkyGuide at the top. Eighteen of the smallest ANSPs scores lead the bottom of the rank with efficiency estimates below 0.4. As noted above, the cost analysis scores are slightly higher so that only seven countries are below 0.4.

In Table 4, we present the SFA cost and production estimates for the terminal activities of the ANSPs. We note that terminal activities are reported at the country level hence aggregate air traffic control procedures at large hub airports and small, regional spokes may lead to less reliable comparisons.

<table>
<thead>
<tr>
<th>Terminals</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>SE</td>
<td>Estimate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>( x_1 ) (IFR airport movements)</td>
<td>0.841 ** 0.020</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>( x_2 ) (Labor cost)</td>
<td>0.454 ** 0.033</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>( x_3 ) (Capital cost)</td>
<td>0.072 ** 0.022</td>
</tr>
<tr>
<td>Environmental variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>( Z_1 ) (Seasonality)</td>
<td>2.337 ** 0.210</td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td>( Z_2 ) (Complexity)</td>
<td>0.194 * 0.080</td>
</tr>
<tr>
<td>Exogenous inefficiency determinants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>( Z_{1a} ) (Ownership gov/corp)</td>
<td>-0.548 ** 0.077</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>( Z_{2a} ) (Ownership agency)</td>
<td>1.280 ** 0.072</td>
</tr>
<tr>
<td>( \delta_3 )</td>
<td>( Z_{3a} ) (Ownership agency)</td>
<td>1.372 ** 0.171</td>
</tr>
<tr>
<td>( \sigma_u )</td>
<td>1.180</td>
<td>1.521</td>
</tr>
<tr>
<td>( \sigma_v )</td>
<td>0.246 ** 0.035</td>
<td>0.082 ** 0.024</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>4.401 ** 1.498</td>
<td>5.068 ** 0.037</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-135.581</td>
<td>-101.612</td>
</tr>
</tbody>
</table>

A ** next to coefficient indicates significance at the 5%/1% level.

The terminal cost function shows that all variables are statistically significant with the expected signs. The second model proved the most relevant with both complexity and ownership form explaining the levels of inefficiency. Again, small economies of scale are estimated at 12 to 13%. Increased complexity improves efficiency levels, which may indicate supplementary economies of scale caused by the additional workload required to handle the complexity. Ownership form also impacts terminal ANSP activities with the agency approach causing slightly higher levels of cost inefficiency compared to the government corporation which in turn adds substantial cost inefficiency above and beyond the public-private form. However, terminal production would not appear to be impacted by the ownership form and model 1 is sufficient.

Figures 3a and 3b present changes in terminal efficiencies over time. Terminal control providers suffered substantially in 2009 as a result of the financial crisis and subsequent reduction in air traffic movements. The largest impacts are clearly shown with respect to the production function which suggests that the ANSPs had difficulty recovering until 2014. Average cost efficiency levels were also impacted in 2009 but gradually improved. However, we also note that average cost efficiency estimates peak at around 0.59 by 2014 and although the trend is positive, the levels of inefficiency are rather substantial.
Whilst the average production efficiency estimates lie around 0.8 in 2014, this masks large heterogeneity between the providers (not shown for the sake of brevity). Cost efficiency estimates range from 0.12 for the Armenian ANSP to 0.92 in Switzerland and Germany. The efficiency estimates show a mix across the continent with Slovenia and Croatia performing relatively better than some of the Western European countries, including Sweden and France.

IV. Conclusions and Future Directions

In this research we focus on the effect of ownership form and airspace characteristics on ANSP performance in Europe. Based on a simple economic model, we learn that effort and efficiency will likely be higher in the case of public companies with a board of stakeholders and in the case of a private company where stakeholders are also shareholders, as is the case with MUAC, NATS and Skyguide. Strong national interest encouraging technology purchases from local suppliers or powerful labor unions, on the other hand, decrease efficiency.

We also estimate econometrically the cost and production functions of 37 European ANSPs over a nine year timeframe. The coefficients are significant and present the expected signs. We note that input prices for labor costs (wages) seems to carry a greater importance in comparison to capital costs. This observation may be explained by the higher share of labor costs at the ANSP total cost level. With respect to the cost function and economies of scale, we find that a 10% increase in traffic given airspace size corresponds, on average, to a cost decrease of around 12%. Structural differences in air traffic characteristics between ANSPs are important in explaining productivity and efficiency performance differences. Seasonality and traffic complexity seem to be particularly relevant. The results of the models also show that complexity explains inefficiency levels but perhaps in an unexpected direction. Given the significant and negative value of the parameter, this suggests that the managers of ANSPs handling higher levels of complexity are more efficient.

We find, consistently, a negative time trend in levels of inefficiency suggesting that, on average, the Single European Skies initiative has been encouraging improvements in cost and productive efficiency over time although much work remains. The significance of the ownership variables in most of the results clearly shows that the choice is
fundamental and impacts the production process directly and the level of inefficiencies too. We find that private-public partnerships achieve significantly higher productivity and cost efficiency. This suggests that governmental agencies and corporations attach a much higher weight to national interests than to the airspace users.

Future directions include expanding the dataset to cover the United States [at the level of the air route traffic control centers], Canada, Australia and New Zealand in order to further develop the analysis and better understand the impact of fine-grained differences in ownership form and the potential for economies of scale.

Acknowledgment

This research was undertaken as part of the COMPAIR project. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 699249. The authors would like to thank the participants of the Air Transport Research Society Conference held in Antwerp in July 2017 for their helpful comments.

REFERENCES

A TRAJECTORY CLUSTERING FRAMEWORK TO ANALYSE AIR TRAFFIC FLOWS

Luis Basora and Jérôme Morio
Information Processing and Systems Department ONERA,
The French Aerospace Lab - 31055
Toulouse, France
Email: luis.basora@onera.fr, jerome.morio@onera.fr

Corentin Mailhot
ENAC
DGAC French Civil Aviation
Toulouse, France
Email: corentin.mailhot@aviation-civile.gouv.fr

Abstract—This paper describes a framework to automatically identify air traffic flows from a set of trajectories by using a clustering algorithm. The framework offers two methods to cluster trajectories, each one using a different distance/similarity measure between trajectories. Results and performance characteristics of both methods are compared by applying them to real trajectories over a French Area Control Center. The framework can output statistics and figures for flow analysis and its use is facilitated by the relatively low number of parameters to be provided by the user. Its aim is to help support the SESAR vision of flow-centric operations by being integrated into Air Traffic Management tools, e.g. for airspace design/management or for analysis of traffic patterns in a free route environment.

I. Introduction

In Air Traffic Management (ATM), massive amounts of data are available to feed data-driven models for real-time decision support or to identify behaviours or patterns relevant to the operational performance of the system in post-event analysis. In particular, sources such as radars, Automatic Dependent Surveillance-Broadcast (ADS-B) or trajectory prediction models generate samples of data points representing trajectories that can then be clustered to identify traffic flows. Trajectory clustering algorithms could be useful in the context of the flow-centric operations vision of SESAR, as described in the European ATM Master Plan [1], by being integrated into tools supporting airspace design/management, complexity management, etc. This is particularly true in a free route environment, where the capacity to understand traffic flows is even more necessary as fixed routes will no longer structure the traffic.

Clustering is a widely used data analysis technique in the statistics/machine learning field. It is about grouping entities with similar characteristics together, so the notion of similarity/distance is essential to the problem. A number of clustering algorithms have been reported in the literature, such as k-means [2], BIRCH [3], DBSCAN [4] and OPTICS [5], all of which are oriented towards the clustering of point data. Even though trajectories have a functional nature (curves), these algorithms can still be applied since trajectory data is available in the form of samples of data points.

For instance, in a recent paper [6], DBSCAN shows promising results when applied to the characterisation of traffic flows based on recorded radar tracks in the terminal airspace of the New York Metro region. This algorithm has the capability to handle noise/outlier data and does not require the number of clusters to be provided as an input parameter.
Previously, in [7], another framework is described based on DBSCAN and k-means to analyse the patterns of traffic over the Northern California terminal area. In both of these studies, however, no distance/similarity measures between trajectories is provided, i.e. the clustering is based solely on the density of the individual trajectory points.

Alternatively, in [8] and [9], two different methods are presented to cluster trajectory segments rather than trajectory points. The former (TRACLUS) implements a variant of DBSCAN to cluster the segments but has never been successfully applied to air traffic as far as we know. The latter is designed to identify air traffic flows in the National Airspace System (NAS), but it is limited to 2D and based on the development of specific algorithms for incremental clustering requiring a non-obvious parameter setting by the user.

Another interesting study that examines the problem of clustering air traffic trajectories is reported in [10] where a spectral clustering approach is used to consider the temporal characteristics of the air traffic flows in the US. However, the procedure to take into account the shape of the trajectories is not evident and the curvature of the clustering centroids might not always be very realistic from an operational point of view. Other approaches such as in [11] relies on a graph structure like a road network, which is not adapted to ATM because of the direct routes often allocated by ATC for tactical reasons or when applied to a free route environment.

More recently, in order to overcome the limitations of some of these techniques, an approach based on entropy minimisation and Lie group modelling has been proposed in [12]. Unfortunately, only partial results are available at the moment of writing this paper and the implementation of the algorithms is fairly complex.

In this paper, we propose a new framework for air traffic flow analysis based on an improved version of DBSCAN called Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [13] which is able to manage clusters of different densities with a single input parameter. Two methods based on two different distance functions between trajectories, Euclidean Distance (ED) and Symmetrized Segment-Path Distance (SSPD) [14] can be selected by the user. These distances offer a trade-off between accuracy and runtime. In addition to the clustering capability, the framework provides several components to filter and interpolate trajectories, compute basic flow statistics and export clusters in Keyhole Markup Language (KML). The framework can be useful to analyse traffic patterns in a wide range of operational scenarios both in en-route and approach. In addition, its implementation and integration into ATM applications is facilitated by the public availability in several languages of some of the algorithms like HDBSCAN [15].

This paper is organised as follows. Section II details the methodology and the different components of the framework. Section III presents the results and the comparison of the two methods by using DDR2 [16] trajectory data provided by Eurocontrol over the Reims Area Control Center (ACC). Finally, section IV summarizes and identifies some ideas for future work and potential applications.

II. The framework

The framework is designed to satisfy the following requirements:

1. It shall be possible to filter trajectories geographically, by airspace, by altitude and in time.
2. It shall be possible to cluster in 2 or 3 dimensions, optionally taking the heading into account as well.
3. The clustering algorithm shall allow for variable trade-off between computing time and quality of the resulting clusters so as to be useful in scenarios with either large or small sets of trajectories.
4. The clustering algorithm shall work within an area where traffic density distribution is not uniform.

5. The clustering algorithm shall work with noise data and be able to identify outliers.

6. The clustering algorithm should require a reduced number of input parameters with clear ATM operational significance.

7. For each cluster it shall be possible to compute the centroid in addition to the set of associated trajectories.

8. For each cluster it shall be possible to compute statistics such as the flow rate, flight distribution per origin/destination, average distance and heading of the cluster trajectories.

Figure 1 shows the framework architecture with the two clustering methods in the center and the pre-processing and post-processing steps. The user can choose the method and the associated parameters. The clustering method based on ED works faster than SSPD at the possible expense of a lower quality clustering result. The specific steps of the framework depend on the selected distance and are described in the following sub-sections.

The framework is implemented in Python 3 with the following Python libraries: scikit-learn 0.18.2, RDP 0.8, HDBSCAN 0.8.8, Pyproj 1.9.5.1, Planar 0.4. The HDBSCAN library is highly optimised, but otherwise no parallelism, multi-threading or code optimisation has been used so far in areas such the SSPD implementation where considerable gains of performance can be expected.

**FIGURE 1 — Trajectory Clustering Framework process for both the ED and SSPD-based distance methods. The framework automatically identifies the flows from a set of trajectories by applying the HDBSCAN clustering algorithm and generates statistics for flow analysis.**
A. Preprocessing

Trajectory datasets contain 3D position samples (latitude, longitude, altitude) for each trajectory and possibly other information like speed and heading. Latitude/longitude coordinates are projected in order to facilitate distance computation. Speed is not used by the clustering algorithms and heading is optional in the ED-based clustering method to ensure flows in opposite directions are separated into different clusters. If heading information is required but not part of the dataset, it is calculated for each trajectory segment from geographic coordinates \([\text{lat}_1, \text{lon}_1, \text{lat}_2, \text{lon}_2]\) and \(\Delta \text{lon} = \text{lon}_2 - \text{lon}_1\) by:

\[
h = \{\text{arctan}[a] + 360\} \% 360
\]

where

\[
a = \frac{\cos(\text{lat}_2) \sin(\Delta \text{lon})}{\cos(\text{lat}_1) \sin(\text{lat}_2) - \sin(\text{lat}_1) \cos(\text{lat}_2) \cos(\Delta \text{lon})}
\]

It is important to note that position samples can be irregularly distributed along the trajectory as shown in Figure 2. Many data points can be removed as they are redundant, whereas other positions need to be interpolated.

**FIGURE 2 — Trajectory discretisation example**

We use the Ramer-Douglas-Peucker (RDP) [17] [18] algorithm to remove redundant trajectory information, but only in the case of SSPD where considerable runtime gains can be expected. Further details on the application of this algorithm are given in section II-B2.

Regarding the interpolation method, there are several methods available from simple linear interpolation to more sophisticated polynomial or spline interpolation. In the framework, the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) [19] [20] method is used as it better takes into account the operational reality of air traffic trajectories. As shown in the example of the vertical profile of a flight in Figure 3, linear interpolation is too abrupt, whereas the order 3 spline ensures the continuity but with oscillations (overshoot). With PCHIP, however, we get a smoother curve much more representative of a true flight trajectory.

**FIGURE 3 — Comparison of interpolation methods**
Finally, as for the filtering capability, in addition to a an altitude range or a time window, the framework allows the user to chose either a bounding box around the geographic area to be studied or a set of airspaces (e.g. ACC or sector). In the latter case, the convex hull is computed for the set of airspaces.

B. Clustering

Once the preprocessing phase is complete and trajectories have been properly prepared, we can apply the HDBSCAN algorithm to them to identify the traffic flows. Compared to DBSCAN, HDBSCAN presents several improvements. First, it only requires the user to define the minimum cluster size, i.e. the minimum number of points (trajectories in our case) to form a cluster. Secondly, HDBSCAN works better than DBSCAN for data with varying density, which is usually the case for air traffic trajectories. Thirdly, HDBSCAN prevents the "bridge effect" between two clusters because of a single or a few data points in the middle of the two clusters (potentially gluing them together) by considering these points as noise.

The choice of an appropriate distance/similarity measure between trajectories is as important as the choice of the clustering algorithm. There are quite a few functions to compute trajectory distances reported in the literature, each one offering a different trade-off between computation time, accuracy and sensitivity to noise data, from the simple ED to the more sophisticated and complex ones like Fréchet and Hausdorff. As our framework is to be used in a wide range of use case scenarios, it cannot be based on a single metric. Also, we need a similarity measure that accounts for both the physical distance and the global shape between trajectories, i.e. time or other trajectory characteristics will not be considered.

We decided to choose the ED function since it is computationally fast and so potentially useful for scenarios involving a large number of trajectories, such as the ones covering wide geographic areas or large time windows. For scenarios where accuracy is more important and execution time less of an issue, SSPD distance should be used instead.

HDBSCAN labels each trajectory with a cluster number (outliers are labelled with a –1) and measures the strength of the cluster membership for each trajectory in the cluster with a probability. In this framework, a cluster is defined by three attributes: a cluster number, the list of trajectories belonging to this cluster (with probability equal to 1), and a representative trajectory or centroid (this last attribute is computed after the execution of HDBSCAN).

In the following subsections, the specificities of the application of HDBSCAN with each distance are detailed.

1. **ED-based Clustering**

The first step in this method is to build a matrix where each row represents a whole trajectory built from the sequence of trajectory points. This requires all the trajectories to have exactly the same number of points, which is indeed the case as we interpolate the trajectories according to the number of points chosen by the user. This is done at least once in the pre-processing phase before filtering, and again if the trajectories are clipped as a result of a filter being applied.

For instance, in the case of a 3D clustering with heading, \( N \) trajectories and \( n \) points per trajectory, the matrix would be defined as:

\[
\begin{bmatrix}
  x_{11} & y_{11} & z_{11} & h_{11} & x_{12} & \ldots & x_{n1} & y_{n1} & z_{n1} & h_{n1} & x_{n2} & \ldots \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\
  x_{1N} & y_{1N} & z_{1N} & h_{1N} & x_{2N} & \ldots & x_{nN} & y_{nN} & z_{nN} & h_{nN} & x_{n2N} & \ldots \\
\end{bmatrix}
\]

where \( x, y, z \) is the projected position, \( z \) the altitude [in meters] and \( h \) the heading computed as in (1). In order for these elements to have the same weight, we standardise each value in the matrix by subtracting the mean and dividing by the variance.

Once these steps are concluded, we can apply the HDBSCAN algorithm with the following two parameters: the minimum number of trajectories per cluster (user parameter) and the clustering matrix built as in (2). ED is computed from the clustering matrix by the HDBSCAN itself as it is part of the standard distances already implemented in the algorithm.

The last step is to calculate the centroid of each cluster as the mean of the trajectories of the cluster.

2. **SSPD-based Clustering**

From a mathematical point of view, SSPD should be considered as a similarity measure rather than a distance or metric, because although it is symmetric, the triangle inequality property is not satisfied. Compared to the Fréchet distance, SSPD does not consider the speed of the trajectory since it is purely geometrical. On the other hand, if two trajectories are similar during the en-route phase because they share the same air routes and diverge only in the final phase of the flight at the terminal area, distance would be over-estimated by both Hausdorff and Fréchet. This is not the case with SSPD, because it better takes into account the global difference in the shape of both trajectories. Also, it presents a good trade-off between the simpler and faster ED and the more complex and time-consuming Hausdorff and Fréchet distances. However, with the SSPD distance it may be difficult to separate trajectories that are geographically close, similar in shape and length but having opposite directions. This may be less of an issue when the clustering is performed in 3D, as air routes for instance are designed to vertically separate flows in opposite directions.

After the pre-processing phase where trajectories are projected, interpolated and finally filtered for the area of interest, there is no need to interpolate the filtered trajectories as opposed to in the ED case. This is because SSPD does not require the trajectories to have the same number of points. However, the SSPD distance being much slower to compute than the ED, the RDP algorithm is applied first to remove redundant trajectory points.

The RDP algorithm can be particularly effective when applied to the en-route phase of the trajectory where aircraft follow long great circle segments linking the flight plan waypoints. Nevertheless, even during the cruise phase, aircraft do turn to follow jet routes and take advantage of favourable winds, avoid hotspot areas, etc. Therefore, we need to use this algorithm carefully enough so as not to exclude too many (if any) turning points. RDP accepts a threshold parameter allowing for the simplification, to a greater or lesser degree, of the trajectory, which can be specified by the user as a framework parameter. With a threshold of 200 meters we notice already an important reduction of redundant points for some trajectories like in Figure 4.

**FIGURE 4 — Application of Ramer-Douglas-Peucker (RDP) algorithm example: points reduced from 100 to 15.**

Once RDP has been applied to the trajectories and the distance matrix is computed, we can run HDBSCAN to perform the clustering. Afterwards, we finally obtain the cluster centroids by choosing the trajectory in the cluster that minimises the distance with the other trajectories in the cluster. Therefore, with this method, centroids are true trajectories and not virtual trajectories like with ED.
C. Post-processing

The objective of this phase is to generate from the clustering results all the necessary outputs to enable the user to analyse the structure of the traffic represented by the clusters. In particular, the framework can generate:

1. 3D and 2D plots of the centroids indicating visually the flow intensity and direction.
2. Histograms to show the distribution of trajectories per cluster.
3. A set of statistics per cluster:
   - average length/altitude/heading,
   - flow rate,
   - flight distribution per origin/destination pair, origin and destination.
4. KML file to display in detail the clusters as well as other related ATM structures such as the airspaces and the air routes.

III. Results

We assess the performance of both clustering methods by applying them to one day of traffic (26 June 2015) over the Reims ACC area. The dataset provided by Eurocontrol in DDR2 format contains 9442 trajectories. A bounding box with coordinates [(46.7°, 1.328889°), (51.116667°, 8.218611°)] is defined around the Reims area to filter the trajectories geographically and different evaluations are performed on several altitude intervals. The coordinates are projected with Lambert-93 which is the official projection in Metropolitan France.

The objective is to identify the flows over Reims ACC and for that we need to specify the minimum number of trajectories (minimum cluster size) for a flow to be considered as such. This is the main clustering parameter to be defined by the user, and its value is highly application-dependent. Our purpose being purely the assessment of the algorithms, we set it up so that major flows can be identified.

The computer used in the evaluations is an Intel®Xeon Quad-Core 2.80GHz processor with 6GB of memory.

A. ED-based clustering results

This first experiment covers the complete volume of traffic in Reims ACC and is performed with the parameters specified in the first column of Table I.

<table>
<thead>
<tr>
<th>Experimental settings</th>
<th>ED</th>
<th>SSPD</th>
<th>ED vs SSPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Min. Alt (ft)</td>
<td>None</td>
<td>31500</td>
<td>30000</td>
</tr>
<tr>
<td>Max. Alt (ft)</td>
<td>None</td>
<td>34500</td>
<td>50000</td>
</tr>
<tr>
<td>Interpolation Points</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Min. Length (NM)</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Min. Cluster Size</td>
<td>50</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>

We want to identify flows of at least 50 trajectories. Even though increasing the number of points after interpolation may improve the clustering accuracy, we have found 100 points
to be a reasonable value. We also filter out any trajectory shorter than 20 NM, which reduces the number of trajectories to 9307.

PCHIP interpolation has the longest computation time, taking 99 seconds against 36 seconds for HDBSCAN. In total, outputs are generated in about 3 minutes, which is acceptable for a scenario with more than 9000 trajectories.

Figure 5 shows the 2D and 3D views of the 38 cluster centroids, where the thickness of the centroid lines is proportional to the intensity of the flow as indicated in the flight distribution per cluster figure. The algorithm has identified both en-route and terminal flows, the latter ones mainly around Paris Charles-de-Gaulle or Paris-Orly (Paris coordinates are 49°, 2.3°) terminal area. We can use the KML output in Figure 6 to better visualise the centroids for the flows departing from and approaching Paris.

**FIGURE 5 — ED-based clustering results**
The number of outliers (32%) is high, but it depends on the size of the cluster (in our case 50) as shown in Figure 7. Also, it can be explained by the fact that the chosen day was the one with the largest volume of traffic for 2015 in Reims. This particularly high traffic density may have induced a higher complexity and required exceptional measures to resolve conflicts or hotspots.

B. **SSPD-based clustering results**

In this case, we use the SSPD-based clustering method to analyse the flows in the LFEEXR sector, with the parameters in the second column of Table I.

After filtering out the flights not entering this sector, a total number of 349 trajectories remain. The RDP algorithm, with a threshold of 200 meters, is next applied to remove redundant trajectory points. Then the SSPD distance is calculated, taking 44 seconds. Adding the 98 seconds of interpolation time for the initial set of 9442 trajectories, the total computation time is approx. 2 minutes and 30 seconds, which is relatively high compared with the 3 minutes for the over 9000 trajectories of the first experiment.

Five clusters are identified with 25% of outliers. We can see the five centroids and the number of flights per cluster in Figure 8. It has to be noted that with SSPD a cluster can have trajectories with opposite directions because of the already-mentioned limitation of this distance and the fact that the clustering is performed in 2D. Therefore, the direction displayed for the cluster centroids may be the opposite direction of some of the trajectories in the cluster.

In Figure 9, we superimpose the identified clusters centroids in red over the air routes in green to make sure that both are consistent in terms at least of geographical location. For instance, cluster number 2 is a northwards flow located in the east of Paris matching perfectly with air route UM733.
C. ED versus SSPD comparison

The last experiment consists of comparing both clustering methods by applying them to the same use case. The goal is to identify the major flows in 2D (minimum of 50 flights per day) in Reims ACC between FL300 and FL500. Clustering is performed in 2D because it is more convenient for visual verification, as controllers are used to a 2D view of the traffic. Parameters are shown in the third column of Table I and the results in Table II.

<table>
<thead>
<tr>
<th></th>
<th>ED</th>
<th>SSPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of trajectories after filtering</td>
<td>6133</td>
<td>6133</td>
</tr>
<tr>
<td>RDP threshold</td>
<td>None</td>
<td>300</td>
</tr>
<tr>
<td>Number of clusters</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>Noise/Outliers (%)</td>
<td>34</td>
<td>43</td>
</tr>
<tr>
<td>Execution time (sec)</td>
<td>306</td>
<td>89463</td>
</tr>
</tbody>
</table>

There is a significant difference in time performance between the two methods, with ED-based clustering being almost 300 times faster than SSPD-based clustering (about 5 minutes versus more than 24 hours). Additionally, ED-based clustering produces about 10% less outliers, which can be explained by the fact that SSPD better takes into account the differences in shape and physical distance between trajectories.

As for the quality of the clustering, ED performs worse in some cases like in cluster number 8 (see Figure 10). Even if the trajectories in this cluster are all southbound, it would be more appropriate to separate them into different clusters to better account for the diversity of headings. However, the ED distance is not sensitive enough to operate this separation and the only alternative would have been to decrease the minimum number of trajectories per cluster, which would result in an even a greater number of clusters. With SSPD, we obtain fewer and more homogeneous clusters, but the number of outliers (43 %) is considerable.
In order to further analyse the identified clusters, Table III gives the statistics for two of the flows identified by the SSPD method. Thus, 5% of the flights in cluster 7 follow the Paris Charles-de-Gaulle (LFPG) – Heathrow airport (EGLL) route, whereas 39% of the flights in cluster 8 follow the Nice (LFMN) – Paris Orly (LFPO) route.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Heading (°)</td>
<td>325</td>
<td>336</td>
</tr>
<tr>
<td>Avg. Alt (ft)</td>
<td>36199</td>
<td>32753</td>
</tr>
<tr>
<td>Avg. Length (NM)</td>
<td>255</td>
<td>75</td>
</tr>
<tr>
<td>Flights/hour</td>
<td>14.7</td>
<td>2.5</td>
</tr>
<tr>
<td>Main orig/dest</td>
<td>LFPG-EGLL</td>
<td>LFMN-LFPO</td>
</tr>
<tr>
<td></td>
<td>5% (17)</td>
<td>39% (24)</td>
</tr>
<tr>
<td>Main origin</td>
<td>LFPG</td>
<td>LFMN</td>
</tr>
<tr>
<td></td>
<td>16% [56]</td>
<td>43% [26]</td>
</tr>
<tr>
<td>Main destination</td>
<td>EGKK</td>
<td>LFPO</td>
</tr>
<tr>
<td></td>
<td>17% [59]</td>
<td>89% [54]</td>
</tr>
<tr>
<td>Main A/C type</td>
<td>A319</td>
<td>A320</td>
</tr>
<tr>
<td></td>
<td>22% [76]</td>
<td>62% [38]</td>
</tr>
</tbody>
</table>

It may not at first be evident as to why cluster 8 overlaps cluster 7, rather than being merged into one single cluster. In fact both clusters contain a majority of northbound flights from south-east France and Italy. However, most of the flights in cluster 8 are for Paris (89% LFPO), whereas in cluster 7 the main destination is London-Gatwick (17% EGKK), therefore there are two distinct routes that are represented by these two clusters.

Figure 11 displays the centroids and their intensities (number of trajectories) to compare the results of both methods. It is not difficult to match the major flows, e.g. ED flows 3 (green), 10 (pale yellow), 1 (blue) correspond to SSPD flows 4 (salmon), 7 (orange), 5 (red). However, except for cases like the anomalous cluster ED 8 and the corresponding SSPD 14, SSPD clusters have a higher number of trajectories as this distance is unable to discriminate similar trajectories with opposite directions. For instance, we can observe that clusters 17 (red), 23 (brown), 25 (blue) identified by the ED method have been replaced by the single and bigger SSPD cluster 19 (dark orange).
D. Verification with planned trajectories

All of the previous experiments are based on the executed trajectories only (M3 in DDR2). However, the difference between the planned and executed traffic may be significant for days with an important volume of traffic as in our case, so it may be interesting to check how well the clustering generated from the planned trajectories (M1 in DDR2) match the ATS Route Network (ARN). With the executed traffic, we have already seen that some of the clusters correspond to air routes in a previous experiment (see Figure 9). In the case of the planned trajectories, which are based on ARN and have not been modified by ATC in the tactical phase, the matching should be even more evident if our clusters are consistent with the operational reality.

We use the same parameters as those for ED (first column in Table I), i.e. with a minimum number of trajectories equal to 50. Even with the ED method, some of the obtained clusters are quite accurate e.g. clusters 4 and 20 in Figure 12.

**FIGURE 12 — Clusters 4 and 20 generated with ED-based clustering**

These two clusters have 74 and 84 trajectories respectively. In each cluster, all trajectories are almost completely overlapping since no deviations from the flight plans were introduced by ATC. In Figure 13 it can be observed that sections of the published air routes UM133 and UM728 over Reims ACC (in green) match perfectly with the two clusters (in black). More generally, most of the clusters correspond to sections or links between sections of air routes. It is also interesting to note that even though the two routes are geographically close, the clustering is accurate enough to properly identify two separate flows.
IV. Conclusion

In this article, we have presented two methods based on the HDBSCAN clustering algorithm to identify air traffic flows from a set of trajectories. The choice of the method to use depends on the volume of trajectories to be processed and the desired accuracy. For a precise clustering, over a limited area, such as a sector or small family of sectors, SSPD-based clustering is the best-adapted, except if we want to make sure that similar 2D flows in opposite directions are properly separated. Otherwise, ED-based clustering is probably the best option as SSPD may be prohibitively slow for some applications unless parallel computing for SSPD is implemented.

Our results have shown that we can match the obtained clusters to existing operational air routes. For the planned trajectories, clusters fit the route structure over Reims as expected, which reinforces our confidence in the framework results. On the other hand, outliers are in general quite high and further analysis is needed to characterise them. The best way to validate the framework would be to use it in a real application and have the operational experts (Flow Management Position or controllers) check that the identified flows make sense.

In particular, we are considering using the framework within the scope of the SESAR PJ08 project (Advanced Airspace Management) [21], where sector configurations could be optimised by minimizing flow cuts. In addition, we need to add further analytic capabilities in order to better understand the contribution of each flow, as well as that of the outliers, to the traffic complexity in a sector.

We would like to extend the framework by adding other clustering algorithms. A segment approach such as the one in TRACLUS seems particularly promising, where a trajectory could potentially be associated not only to a single cluster, but to a sequence of clusters, to better explain the different flight phases. Thus, in the en-route phase, the trajectories sharing the same air routes could have their en-route segments clustered, independently of whether or not they share the same departure/arrival procedures.

Acknowledgment

The authors would like to thank Sébastien Aubry, Judicaël Bedouet, Thomas Dubot, Antoine Joulia and Xavier Olive from ONERA and Pascal Lezaud from ENAC for their contributions to the present work.
REFERENCES

COMBINING VISUAL ANALYTICS AND MACHINE LEARNING FOR ROUTE CHOICE PREDICTION

Application to Pre-Tactical Traffic Forecast

Rodrigo Marcos, Oliva G. Cantú Ros, Ricardo Herranz
Nommon Solutions and Technologies Madrid, Spain
rodrigo.marcos@nommon.es

Abstract—One of the key enablers of ATM Network Management is the forecasting of the volume and complexity of traffic demand at different planning horizons. This paper proposes a visual analytics and machine learning approach for the prediction of airline route choice behavior in the pre-tactical planning phase, when few or no flight plans are available. Visual analytics is used to identify relevant variables determining airline route choices. The output of this analysis serves as a starting point to develop a multinomial logistic regression model that predicts route choices as a function of the identified relevant variables. We evaluate the predictive power of the model, showing its potential to outperform traditional forecasting methods. We conclude by discussing the limitations and room for improvement of the proposed approach, as well as the future developments required to produce reliable traffic forecasts at a higher spatial and temporal resolution.

I. Introduction

The goal of Air Traffic Flow and Capacity Management (ATFCM) is to make airport and airspace capacity meet traffic demand and, when capacity opportunities are exhausted, optimize traffic flows to meet available capacity. An essential enabler of ATFCM is the provision of accurate information about anticipated traffic demand. The available information (schedules, flight plans, etc.) and its associated level of uncertainty differ across the different ATFCM planning phases, leading to qualitative differences between the types of forecasting that are feasible at each time horizon. While abundant research has been conducted on tactical trajectory prediction (see, e.g., [1] and [2]), trajectory prediction in the pre-tactical phase, when few or no flight plans are available, has received much less attention. The tool currently used by EUROCONTROL for pre-tactical traffic forecast is the so-called PREDICT system [3], which transforms flight intentions into predicted flight plans by assigning to each flight the flight plan of a similar flight that occurred in previous weeks. The route assigned to each flight intention is based on limited similarity criteria found in historical flight plans, without consideration of other factors (such as airline characteristics, meteorology, etc.) that also play an important role in airline route choices [4]. These simplifications limit the accuracy of the forecast, which may lead to inefficient or sub-optimal ATFCM decision-making [5].

The starting point for the present work is the hypothesis that the quality of pre-tactical traffic forecasts can be enhanced by better exploiting historical data with predictive models that incorporate a finer characterization of airline route choices. Previous research has focused in the prediction in the tactical phase (short-and mid-term) to estimate arrival time at airports [1] or aircraft position to detect trajectory conflicts [2],
Flow management

by incorporating factors such as the actual trajectory and weather forecasts. The goal of this paper is to explore how the combination of visual analytics and machine learning can be applied to historical flight data to extract meaningful insights on route choice determinants and develop new approaches able to improve the accuracy and reliability of demand forecasting in the pre-tactical phase.

Visual analytics focuses on analytical reasoning facilitated by interactive visual interfaces, offering a way to discover unexpected patterns and relationships in big and heterogeneous datasets [7]. In this paper, visual analytics is used to identify potential explanatory variables of airline route choices and to get a first qualitative idea of the impact of each variable. A machine learning model is then developed that translates the insights obtained from the visual exploration of flight trajectories into a route choice predictor. The model is calibrated and validated with several months of historical data. We instantiate and evaluate these ideas through their application to a specific case study consisting in analyzing and modelling airline route choices for the flights departing from Istanbul airports and arriving in any of the Paris airports.

The rest of this paper is organized as follows: Section II describes the selected case study, the data sources used, and the approach and methodology followed for route choice analysis and modelling; Section III describes the set of route choices between Istanbul and Paris considered in the analysis; Section IV summarizes the results of the exploration of historical flight data by means of different visual analytics techniques and the main insights extracted from this analysis; Section V presents the route choice predictor and the results of model training, validation and testing, comparing the model predictions with those provided by a null model; Section VI concludes and discusses future research directions.

II. Data and methodology

A. Case Study

As an application exercise, we have selected the Origin-Destination (OD) pair Istanbul-Paris. We study the flights departing from the Atatürk (LTBA) and Sabiha Gökçen (LTFJ) airports and arriving in Charles de Gaulle (LFPG) and Orly (LFPO). The criteria used to select this OD pair were:

- to represent one of the main European air traffic flows (in this case the South-East traffic axis);
- to have a significant volume of traffic (on average, there are more than 10 flights per day from Istanbul to Paris);
- to include a sufficiently high number of alternative route options.

The period used for data exploration and for the training of the machine learning model consists of the AIRAC cycles 1601, 1602 and 1603, i.e., from the 7th of January 2016 to the 30th of March 2016. The period used for model testing comprises AIRAC cycles 1501 and 1502, i.e., from the 8th of January 2015 to the 4th of March 2015.
B. Data Sources

1. DDR

The Demand Data Repository (DDR) is a restricted-access flight database maintained by EUROCONTROL, which records data for almost all flights flying within the European airspace (ECAC area). The information stored in DDR includes:

- Trajectory description: coordinates, timing, altitude and length of the flight.
- Flight description: ID, airline, aircraft, origin, destination, date, departure time, arrival time, most penalizing regulation and ATFM delay.
- Airspace information: charging zones shape and airport coordinates.

This information is available for both the last filed flight plan and the actual flight trajectory. The 4D trajectories in the DDR are not radar tracks, but a simplification that only includes those points that deviate significantly from the Flight Plan (FP).

The current study focuses on the analysis and prediction of the routes followed by actual trajectories.

2. CRCO

The Central Route Charges Office (CRCO) is an office within EUROCONTROL that charges airspace users for air traffic services on behalf of the Member States. The CRCO calculates the route charges due to the Member States for the services provided, bills the airspace users and distributes the route charges to the States concerned [8]. The unit rates and tariffs for en-route and terminal charges are published on a monthly basis by the CRCO in the EUROCONTROL website [9].

C. Approach and Methodology

1. Route Clustering

Usually there is a vast number of route options to fly from one airport to another. The aim of this study is not to predict accurately the route followed by each aircraft, but the airspace through which the aircraft will fly. To convert this problem into a discrete-choice form, the actual trajectories of historical flights are grouped into a set of clusters represented by a mean trajectory. Density-Based Clustering (DBC) is used. In DBC, clusters are formed by a set of core samples close to each other and a set of non-core samples close to a core sample, but not considered as core samples themselves. This allows the computation of clusters with any shape, which makes it more generic than centroid-based approaches (k-means clustering). Core samples are those in areas of high density whilst non-core samples are within a maximum distance to a core sample, but without a minimum number of nearby core samples. Any sample that is not a core sample and is not within the maximum distance to a core sample is identified as noise. In our implementation, the routes assigned to a cluster with less than 5% of the total number of flights are also treated as noise. The routes identified as noise are grouped into an additional category named as "other". DBC was implemented using the function DBCScan of the Python public library scikit-learn [10].

2. Visual Exploration

The objectives of the visual exploration phase are to discover relevant explanatory variables of airline route choices. Route choice determinants are explored by means of different types of temporal and spatial representations, including heatmaps, multivariate map representations, and multivariate bar plots.
3. Route Choice Modelling

The goal of this phase is to model airline route choices as a function of the explanatory variables identified by means of the visual exploration. The modelling process comprises two steps: first, flights are segmented according to their characteristics; then, for each segment, airline choices are modelled as a function of the identified explanatory variables, using a multinomial logistic regression model [11]. The output of the model is the probability of a route option to be chosen. The model is fit to the actual observed probabilities in the training dataset, consisting of 70% of the flights during the training period. The rest of the flights in that period are reserved to validate the model by comparing predicted and actual figures. The training and validation datasets are separated randomly. Once validated, the model is applied to a different period of time (testing period) to evaluate its predictive power. The testing period may include routes and airlines not present in the training dataset. Hence, route options are re-computed with data of the first AIRAC cycle in the testing period. The rest of the testing data are used to measure the performance of the model. The results obtained with the model are compared with those of a null model that assigns a route to a flight with a probability equal to that observed for flights in his segment in the training dataset.

III. Route clustering

A. Route Clustering Results

The average trajectory of the clusters and the trajectories assigned to each cluster are shown in Figure 1. The trajectories are grouped into 8 clusters: Cluster 0 (red) enters LF through ED avoiding LR; Cluster 1 (green) enters LF through ED, LK and LZ; Cluster 2 (gray-green) avoids ED through LO; Cluster 3 (light blue) goes through LD, LI and South LS; Cluster 4 (orange) goes through LD, LI and North LS; Cluster 5 (blue) enters LF through ED and LR; Cluster 6 (dark blue) goes through LJ and North LS; Cluster 7 (purple) goes through LK, LO, LH and LR. The main characteristics of each cluster are shown in Table I.

TABLE I — Cluster statistics.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No of flights</th>
<th>Average length (NM)</th>
<th>Average charges (EUR)</th>
<th>Regulations per flight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>139</td>
<td>1277</td>
<td>1188</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>110</td>
<td>1314</td>
<td>1144</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>190</td>
<td>1273</td>
<td>1199</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>218</td>
<td>1274</td>
<td>1203</td>
<td>0.06</td>
</tr>
<tr>
<td>4</td>
<td>117</td>
<td>1256</td>
<td>1207</td>
<td>0.07</td>
</tr>
<tr>
<td>5</td>
<td>73</td>
<td>1274</td>
<td>1204</td>
<td>0.1</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
<td>1271</td>
<td>1229</td>
<td>0.03</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>1304</td>
<td>1152</td>
<td>0.04</td>
</tr>
</tbody>
</table>

FIGURE 1 — Results of route clustering: a) Average trajectories. b) Actual trajectories colored by assigned cluster. The background shading indicates the unit rate of each charging zone: red means more expensive, blue means cheaper.

a)
b) Visual exploration

A. Exploration of Flight Efficiency Metrics

First, we study the characteristics of individual flights and their relationship with the average values of the corresponding cluster. Figure 2 shows the most direct routes (in green) and also the variability inside a cluster. Horizontal length varies from 1,230 to 1,360 kilometers. Clusters 0, 2, 3, 5 and 6 have a medium length and include routes with a wider range of lengths. Cluster 4 has the shortest average length, with little dispersion among the flights that form the cluster. Clusters 1 and 7 have higher distance values, and also low dispersion. The most selected clusters (3, 2 and 0) show intermediate values of horizontal length, despite having a much lower achievable length. As an example, the lowest length flown in Cluster 3 is 1,247 kilometers, which is lower than the average value of route 4 (1,256 km), whilst the average length of Cluster 3 is 1,274 km. This suggests that, in addition to the average distance values, the achievable distance values may also have an impact on route choice. In any case, it is clear that the horizontal length is not the only variable that determines route choice.

FIGURE 2 — a) Horizontal length of individual trajectories. b) Average value per cluster. Length is expressed in Nautical Miles (NM).
B. Exploration of Route Charges

Figure 3 shows en-route charges per flight and average route charges per cluster. Charges are in general homogeneous inside a cluster. We can observe that Cluster 1, despite having the highest average length, is the fifth most flown route due to having the lowest charges. The same applies to Cluster 0, with high length but low charges, which is the third most flown route. On the other hand, the shortest route (Cluster 4) is the fourth most flown due to its high charges. Clusters 3 and 2, the most flown, offer a longer but much cheaper alternative.

FIGURE 3 — a) En-route charges of: individual trajectories. b) Average value per cluster. Charges are expressed in EUR.

C. Exploration of Flight Duration

Another variable affecting route choice is flight time. This parameter is highly correlated with horizontal length, but can be adjusted during the flight, thus resulting in a high variability inside a cluster (see Figure 4). The yellowish colors indicate that the average values per cluster are far from the extreme values achieved by some individual flights. Cluster 5 has the lowest average flight time although its average length is longer than that of other clusters and its charges are moderate. This suggests that this route could be suitable to recover delay.

FIGURE 4 — a) Flight duration of individual trajectories. b) Average value per cluster. Time is expressed in minutes.
D. Exploration of Arrival Time

The arrival time may influence route choice in several ways, e.g. flights departing earlier may be prone to fly non-congested routes in order to avoid reactionary delay. However, Figure 5 shows a high variability within clusters, and therefore the direct use of average values per cluster is meaningless. The relevance of arrival time becomes clearer when congestion is taken into account.

FIGURE 5 — Arrival time of individual trajectories. Green means early morning flights, red means late evening flights.

E. Exploration of Congestion Metrics

To explore the impact of congestion on airline route choices, two metrics are considered at cluster level: average deviation of the actual flight level (FL) flown during cruise with respect to the reference FL in the last FP (Figure 6) and average number of regulated flights (Figure 7). Regarding the average deviation of FL with respect to the FP, Clusters 2, 7 and 1 have the highest values, whilst Clusters 6, 0 and 3 have the lowest values. Regarding the number of regulations, clusters 5, 1 and 0 (i.e., the ones flying through central Europe, which is highly congested) have values above 10%. On the other hand, Clusters 6 and 7 have the lowest number of regulations. Combining both metrics, Clusters 3 and 6 seem to be less congested than the rest, whilst Clusters 0, 2 and 5 appear to be the most congested.

FIGURE 6 — Average deviation of FL: a) Individual trajectories. b) Average value per cluster. The values are given in FL.
The average deviation of FL (Figure 6) has high dispersion inside a cluster. The reason is the intra-day variability of congestion. It seems therefore interesting to study the relationship between the selection of routes and the arrival time and its corresponding level of congestion (Figure 8), as airlines may tend to avoid congested routes at traffic peak hours. Early morning flights (Figure 8a) choose in general Clusters 2, 3 and 0. Cluster 2 is the most congested, while the rest show low FL deviation, i.e., they are less congested. Flights at the morning traffic peak (Figure 8b) do not consider Cluster 3 and tend to fly more deviated routes like Cluster 5 and 7, or even Cluster 4, with low FL deviation but high charges. Cluster 2 is still used in spite of being congested. At this point it is important to note that average congestion metrics of deviated routes might appear higher than those of the direct routes, even when those deviated routes are actually less congested. This is because the average is calculated over the total number of flights taking each route, and deviated routes are selected mainly during high traffic peaks. Flights in the afternoon (Figure 8c) continue to choose deviated routes due to congestion in the more direct routes (Cluster 2). In this case the preferred route is Cluster 3, due to its low level of congestion. In the evening (Figure 8d), the tendency is the same as in the afternoon. In the early evening (Figure 8e), congestion levels are similar to those in the afternoon, resulting in similar route choices. The last flights of the day (Figure 8f) tend to choose Cluster 5 (fastest) or 3 (shortest).

**FIGURE 7 — Average number of regulations per cluster.**

**FIGURE 8 — Variations of FL of actual trajectories arriving between:**
a) 6:00 and 8:30; b) 8:30 and 12:00; c) 12:00 and 16:00; d) 16:00 and 20:00; e) 20:00 and 22:00; f) 22:00 and 00:00. The colour scale is the same as in Figure 6.
F. Exploration of Airline Behaviour

When analyzing route choices per airline (Figure 9), differences between airlines arise. Turkish Airlines (THY) flies virtually all the clusters, with preference for Clusters 1, 2 and 4. Air France (AFR) and Pegasus Airlines (PGT) also use most of the available routes. AFR has a marked preference for Cluster 0, while PGT fairly divides its flights among the Clusters 1, 2, 3 and 6. On the contrary, Onur Air (OHY) flies almost only Cluster 3 regardless of external variables. Atlasjet (KKK) and MNG Airlines (MNB) fly a narrower set of two or three clusters.

**FIGURE 9 — Number of flights of each airline per cluster.**

These results suggest that the influence of the route choice determinants identified in the previous sections depends on other, airline-specific factors (e.g., cost of delay) that may be driven by the business model of each airline, the structure of its network (point-to-point vs hub-and-spoke), etc.

G. Conclusions of Visual Exploration

The present visualization exercise allows the extraction of relevant insights regarding airline route choice criteria. The factors identified as route choice determinants are:

- Horizontal length, which is the most significant parameter to explain fuel costs.
En-route charges, which explain air navigation costs. Longer routes often avoid expensive charging zones, thus reducing the amount of charges paid.

Congestion. Some routes may provide a stable flight time, less delays or regulations, or allow airlines to fly their desired FL, thus reducing fuel consumption. Congestion is not constant and it is more relevant during traffic peaks. Thus, an accurate route choice model should be able to capture the different levels of congestion at different times of the day.

Flight time. This variable is highly correlated with the horizontal length of a flight. However, it presents high dispersion inside clusters because of its link with factors such as wind and assigned FL.

Weather, which can affect route choice in two ways: weather events as CBs may deviate a route, and tail winds may make one route choice better than other.

Airline. All the above factors may have different importance depending on the structure of costs of each airline. Point-to-point carriers tend to use routes with low air navigation charges, while hub-and-spoke airlines may prefer to choose routes that are more stable in time. It may also be the case that smaller airlines are not always able to optimize their route choices taking into account all these factors due to their more limited resources.

While some factors are intrinsic properties of the routes (e.g., average horizontal length), their influence may depend on certain characteristics of the airline (e.g., cost of delay). There are also factors that change daily (e.g., wind). Additionally, route choices might depend on other variables that have not been explored in the analysis, such as the reactionary delay due to previous flights or the availability of certain routes as a function of military activity, thus generating an additional variability that cannot be explained by the observed variables.

V. Route choice modelling

A. Explanatory Variables and Mathematical Model

The explanatory variables selected from the visual exploration can be classified into:

- flight attributes: airline and arrival time;
- route attributes: average horizontal flight efficiency [12], average air navigation charges and probability of being subject to a regulation.

Flights are segmented according to the flight attributes by means of a k-means clustering. Then, for each segment, route attributes are used as input to a multinomial logistic regression function [11] to obtain the choice probability for each option:

$$P_i = \frac{\exp \left( \sum_{k=1}^{m} \beta_k x_{ik} \right)}{1 + \sum_{j=1}^{n} \exp \left( \sum_{k=1}^{m} \beta_k x_{jk} \right)}$$

where $P_i$ is the probability of option $i$, $\beta_k$ is the model constant associated to the k route attribute, $x_{ik}$ is the route attribute k of the option i, $m$ is the number of route attributes and $n$ the number of route options.
B. Model Training

For each flight, airline route choice is assimilated to one of the 8 clusters depicted in Figure 1a, by selecting the cluster to which the actual trajectory belongs. Flights are segmented by airline (6 classes) and arrival time (4 classes), resulting in 24 segments. For each segment, the training dataset is used to calibrate the parameters of the route choice model so as to fit the observed airline choices.

The model achieved a good fitting of the training dataset, with all predicted values within ±5% of the actual values. Errors are mainly generated by clusters with very similar characteristics, such as Clusters 0 and 5, both with intermediate length and relatively low charges (see Table I): these clusters cannot be distinguished by the model and return very similar probabilities, so that flights choosing one of these clusters are incorrectly assigned to the other cluster. This suggests that there is a missing factor in the current model explaining the difference in the choice probability of these two clusters.

C. Model Validation

Figure depicts the comparison of the choices predicted by the model with the actual route choices for the validation dataset. The results show a fair approximation of route choice, with an error within ±10% of the actual values. The worst results are again obtained for Clusters 0 and 5, due to their similarity along the considered explanatory variables. This could be improved by including other route choice determinants, such as wind, airport configuration, delay at takeoff, etc., as well as by using a dynamic congestion indicator, as discussed in Section IV.E.

**FIGURE 10 — Validation results. Early flights arrive before 12:00; midday flights between 12:00 and 16:00; late flights after 16:00.**

D. Model Testing

Testing gives a final estimation of the predictive power of the model. The results of the testing are shown in Figure 11:

- In general, the clusters for which the validation results were less accurate, such as Clusters 0 and 5, are also the ones providing the worst results in the testing experiment. The case of Cluster 0 is remarkable, as the model would be expected to reduce the number of flights assigned to it due to the higher charges in 2015. Instead, the prediction is higher. The reason for this is the model training: in the training period, Cluster 3 has more flights than Cluster 0, despite having similar length and higher charges (see Table I). In order to fit this behavior, the model gives little weight to charges, assigning a similar probability to both clusters.

- The worst performance is obtained for midday flights, coinciding with the peak of congestion (see Figure 8c).
As previously discussed, these results reveal the need for additional explanatory variables able to account for the factors not captured by the current model (e.g., by using dynamic congestion metrics).

Table II shows the correlation between the routes predicted by the proposed model and the actual route choices, compared with the results obtained with the null model, which assigns routes according to the empirical probability distributions observed within each flight segment during the training period. This null model aims to emulate current PREDICT algorithm used by EUROCONTROL [3]. Despite the room for improvement, the model predictions show much better correlation with actual choices than the null model. The poor results of the null model are explained by the steep change in unit rates between 2015 and 2016, which cannot be predicted with such a simple model.

**FIGURE 11 — Comparison of actual, testing and null model results. Flights are grouped per arrival time as in Figure 10.**

**TABLE II — Comparison of testing results and null model. Early flights arrive before 12:00; midday flights between 12:00 and 16:00; late flights after 16:00.**

<table>
<thead>
<tr>
<th></th>
<th>Pearson’s correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.9588</td>
</tr>
<tr>
<td></td>
<td>Null model</td>
</tr>
<tr>
<td>Early Flights</td>
<td>0.9360</td>
</tr>
<tr>
<td></td>
<td>Null model</td>
</tr>
<tr>
<td>Midday Flights</td>
<td>0.6956</td>
</tr>
<tr>
<td></td>
<td>Null model</td>
</tr>
<tr>
<td>Late Flights</td>
<td>0.7352</td>
</tr>
<tr>
<td></td>
<td>Null model</td>
</tr>
</tbody>
</table>

**VI. Conclusions and future directions**

In this paper, we have presented a combined approach to pre-tactical route choice prediction based on the joint application of visual analytics and machine learning techniques to historical flight data. Visual analytics is used to unveil the main determinants of airline route choices, which are then included as explanatory variables in a multinomial logistic regression model. The model provides a fair prediction performance, showing the potential of the proposed approach to outperform current pre-tactical forecasting methods, which result often in over deliveries [13] after the ATFCM process. However, further improvements of the presented model are needed in order to achieve acceptable levels of predictability.
Future research directions are outlined below:

- Other machine learning techniques (e.g., decision trees, neural networks) could be tried to evaluate which technique(s) provides the best results and under which conditions.

- The explanatory variables used by the model could also be improved. In particular, the indicators used as a proxy of congestion could be enhanced by considering a dynamic variable (e.g., depending on the arrival time) able to capture the different levels of congestion along the day.

- The predictive models should incorporate other relevant route choice determinants, such as wind and availability of routes. In the current approach, the influence of wind is not taken into account; doing so would require a dynamic variable that should be computed for each flight and for each cluster, e.g. using the wind forecasts at the departing time. Additionally, in the model presented in this paper, airspace design is only taken into account implicitly, through the routes followed by historical flights. This approach is expected to provide good results when the airspace structure is stable. However, some elements of the airspace, such as military areas, vary over time. The model could therefore be improved by considering only the choice set formed by the routes available at the departure time.

- The model presented here has been trained with a dataset of historical flights corresponding to one single season. Extending the training dataset to encompass data from several seasons could help improve prediction across seasons.

- More generally, the proposed approach could be extended to develop an adaptive approach in which models are recalibrated on a continuous basis to account for the most recent changes in the network.

- Airline decisions are usually driven by a cost optimization process. An interesting line of research would be the combination of data-driven approaches such as the one presented in this paper with optimization methods for trajectory prediction, in order to estimate variables such as the distribution of the cost of delay for different airlines.

A prospective application of the proposed modelling approach is the aggregation of route predictions into traffic demand volumes in order to predict the appearance of hotspots. To do so, the current approach should be applied to all OD pairs for which one or more possible routes cross the hotspot. Then, predictions should be aggregated in a probabilistic manner to obtain the predicted traffic volume in the hotspot.

On a more strategic level, the modelling approach developed in this paper could also be used to investigate questions related to the interrelationship between ATM Key Performance Areas, e.g. the trade-offs between environment (flight efficiency), capacity (delay) and cost-efficiency.

**Acknowledgment**

The project leading to these results has received funding from the SESAR Joint Undertaking under grant agreement No 699303 under European Union’s Horizon 2020 research and innovation programme. The authors would like to thank the INTUIT project team, as well as the members of the INTUIT Advisory Board, for their valuable inputs. We would also like to thank the SESAR JU Project Officer, Ivan de Burchgraeve, for his continuous and timely support throughout the project.
REFERENCES

Human factor
ANALYSIS OF A WORKLOAD MODEL LEARNED FROM PAST SECTOR OPERATIONS

David Gianazza
ENAC, F-31055 Toulouse, France
Univ. de Toulouse, IRIT/APO, F-31400 Toulouse, France
Email: {lastname}@recherche.enac.fr

Abstract—In this paper, we assess the performance of a workload model trained on a subset of sectors, focusing on how it generalizes on fresh sectors. The model of the air traffic controller workload is learned from historical data made of workload measurements extracted from past sector operations and ATC complexity measurements computed from radar records and airspace data (sector geometry).

The workload is assumed to be low when a given sector is collapsed with other sectors into a larger sector, normal when it is operated as is, and high when it is split into smaller sectors assigned to several working positions. This learning problem is modeled as a classification problem where the target variable is a workload category (low, normal, high) and the explanatory variables are the air traffic control (ATC) complexity metrics.

In previous work, we compared several classifiers on this problem. The models were trained on one week of traffic, and their generalization performance was assessed on another week of traffic, using the same sectors in both the training and test sets.

In the current work, we examine if models learned on a specific set of sectors can be performant on any other sector, or not. We also give a closer look at how the workload varies with the ATC complexity measures in our data, using bagplots of the data points for a few sector instances. The results allow us to better understand the strengths and limits of our data-driven model.

I. Introduction

Predicting the workload of air traffic controllers is of crucial importance to the safety of the air traffic management (ATM) system at large. Overloads might lead to potentially dangerous situations where some conflicts might not be detected in time by the controllers.

Predicting the workload with good accuracy is also a question of efficiency. In day-to-day operations, the airspace is dynamically reconfigured according to the controller workload. Underloaded sectors are collapsed to form larger sectors, and overloaded sectors are split into several smaller sectors operated separately. When it is not possible to absorb the traffic simply by reconfiguring sectors, the traffic is delayed or rerouted so as to avoid the congested areas. This needs to be done well in advance, usually before the aircraft take off. Predicting with greater accuracy which ATC sectors shall be operated at what time and which of these sectors might get overloaded would improve the whole traffic regulation process. This requires a realistic and accurate workload model, which is the subject of this paper.

This problem has been addressed in several ways since the beginnings of air traffic control (ATC). Depending on the context and purpose, one might count the movements on an airport, or the number of aircraft within the boundaries of an en-route sector, or the incoming flow of traffic over a time period. Such basic metrics – and the associated
threshold values (capacities) – provide simple and straightforward answers to the question of deciding whether the controllers are experiencing a normal workload when handling given traffic, or if they are overloaded.

However, it has been acknowledged for a long time that simple metrics, such as aircraft count, do not adequately reflect the complexity of air traffic control. ATC complexity covers dynamic aspects relative to the traffic, static aspects relative to the sector geometry and route network, and aspects relative to the air traffic control procedures.

In this paper, we are interested in examining more closely the relationship between ATC complexity metrics and workload, using a gradient boosted tree model selected in our previous work [1]. This model is trained on historical data of aircraft trajectories and past sector operations. This data is made of complexity measures computed from radar tracks and sector data, and workload measures extracted from the status of the control sectors (collapsed, opened, or split into smaller sectors). We have shown in previous works that such a model provides correct predictions in more than 80% of the cases, when training the model on one week of traffic in all the French sectors, and when assessing the results on another week of traffic in approximately the same sectors.

Our aim here is to check if such a model can be trained on a subset of sectors, and still generalize well on fresh sectors. In addition, we want to verify if some intuitive notions on the relationship between complexity and workload – such as “if this complexity value increases, then the workload should be higher” – are actually true for our data. In other words, we would like to know if our model actually captures the relationship between the cognitive workload of the controller and the complexity metrics, or if our data-driven approach also captures some of the characteristics of the sectors and traffic patterns in our data.

The remainder of this paper is organized as follows: Section II gives some background on ATC complexity and air traffic controller workload, and states the objectives of the work presented in this paper. Section III is a short introduction to machine learning. Section IV describes the gradient tree boosting method used in this study. The data and experimental setup are described in section V. The results concerning the generalization performance on fresh sectors are given in section VI, and those concerning the relationship between complexity and workload are given in section VII. The paper concludes with a brief summary of our findings and the perspectives of future works, in section VIII.

II. Background and objectives

A. ATC complexity and air traffic controller workload

In this paper, we are interested in the relationship between ATC complexity and workload. Both of these concepts are loosely defined in the literature ([2]), and before building models relating one to the other, we need to quantify them. Many ATC complexity indicators have been proposed in the literature [2], [3], [4], and this paper proposes nothing new in that matter. Quantifying the controller’s workload has been done through different kinds of measures: physical activity ([5], [6]), physiological indicators ([7], [8], [9]), or subjective ratings ([10], [11]). Some of these indicators are difficult to interpret, and others are subject to biases (such as the recency effect denounced in [8], and the possibility of raters errors in the case of “over-the shoulder workload ratings” [12]). Collecting these data requires heavy experimental setups, often resulting in relatively small datasets and potential overfitting issues when trying to adjust a model on too few examples.

In order to avoid these drawbacks, we have proposed in previous work ([13], [14], [15]) to use historical records of the past sector operations to quantify the workload. These records are available in large quantity, for a large number of sectors. The information that can be extracted from the past sector operations is the following: we can assume that the workload was normal when the sector was operated, low when it was merged with other sectors to form a larger sector, and high when it was split into smaller sectors assigned to several working positions.
Several approaches have been tried to build models relating ATC complexity to workload. For example, taskload models ([16], [17]) compute the cumulative time required to execute control tasks. Linear regression models such as the popular dynamic density models ([18], [10]) approximate subjective workload ratings by a linear combination of a number of ATC complexity measures. Other works use a neural network instead of a linear model ([11]) to approximate subjective ratings.

In previous work, we also used neural networks, but our target variable was the workload measured from the past sector operations instead of subjective ratings. Considering an initial set of 27 complexity metrics found in the literature, we selected a subset of relevant metrics for the purpose of building a model that could be used to predict future airspace configurations ([13], [14], [15]). We showed that this concept was feasible and could be used to forecast airspace configurations that were much more realistic than the actual sector opening schedules made by the Flow Management Positions ([19], [20]). The concept was demonstrated on a mock-up HMI using static data ([21]). In [1], we compared the performances of several machine learning methods using the six selected ATC complexity metrics as input. The results showed that gradient boosted trees and neural networks performed better than basic classification methods such as linear discriminant analysis, quadratic discriminant analysis, or naive Bayes classifiers.

B. Objectives

In all our previous works, the ATC sectors from which were drawn our samples were approximately the same for both the training and test sets. The test data was simply taken in a different time interval than the training data, in order to assess the generalization performance of the trained models on fresh inputs (although in the same sectors).

In this paper, our first objective is to check if a model trained on a subset of sectors can perform well on another subset of sectors. Our concern is that the trained model might not perform well on elementary sectors, for which there are no occurrences of “high” workload in the training data (as well as in the test data). A second objective is to examine more closely the relationship between the input variables (ATC complexity metrics) and the workload to determine if we actually learn a model of the cognitive workload of the air traffic controller, or if we also learn some characteristics of the observed data.

III. A short introduction to machine learning

This section is a brief introduction to machine learning. The reader may refer to [22], [23], [24] for a more thorough view of this active research field.

Learning from data with a computer can be done in different ways, through supervised learning, unsupervised learning, or reinforcement learning. In reinforcement learning, a software agent takes actions in a given environment so as to maximize a cumulative reward. In supervised or unsupervised learning, given some features $x$ of an observed phenomenon, the objective is to learn a model from a set of examples $(x_1, ..., x_N)$. Unsupervised learning considers the explanatory variables $x$ either to produce clusters of data, or to estimate the probability density of $x$, using the examples $(x_1, ..., x_N)$. In supervised learning, we assume a relationship $y = f(x)$ between $x$ and a target variable $y$, and we use examples of the outputs $(y_1, ..., y_N)$ associated with the inputs $(x_1, ..., x_N)$ to learn a model $h$ approximating $f$.

In this paper, supervised learning techniques are used to predict the workload from ATC complexity indicators. The target variable $y$ is here a workload category (low, normal, or high) and the input $x$ is a vector of complexity indicators computed from the traffic or the sector geometry.
Such learning problems where the target is a categorical variable are usually referred to as classification problems, as opposed to regression problems where \( y \) is a floating-point value or a vector of floats.

Given a loss function \( l \) such that \( l(y, \hat{y}) \) is the cost of the error between the computed output \( y = h(x) \) and the observed data \( y = f(x) \), our objective is to choose \( h \) minimizing the following risk (i.e. the expected loss), where \( X \) and \( Y \) are the random variables from which are drawn \( x \) and \( y \):

\[
\mathcal{R}(h) = \mathbb{E}_{(x, y)} \left[ l(h(x), y) \right] = \int_{x, y} l(h(x), y) p(x, y) \, dx \, dy
\]  

\[ \text{(1)} \]

A. Learning from a finite dataset

In practice, the joint distribution of \( X \) and \( Y \) is not known and one can only approximate \( f \) using a set of examples \( S = \{ (x_1, y_1), \ldots, (x_N, y_N) \} \) of finite size.

The most straightforward idea is then to select the model \( h \) minimizing the following empirical risk:

\[
\mathcal{R}_{\text{emp}}(h, S) = \frac{1}{|S|} \sum_{(x_i, y_i) \in S} l(y_i, h(x_i))
\]

\[ \text{(2)} \]

Unfortunately, minimizing the empirical risk on \( S \) might not lead to the most desirable model. The selected model might fit the examples \( \{ (x_1, y_1), \ldots, (x_N, y_N) \} \) of \( S \) very well, while performing poorly on new instances of \( x \).

Statistical models can be more or less ‘flexible’ when fitting the data, depending on their analytical expression. For example, a linear model is much less likely to overfit the data than a polynomial of high degree. Selecting the best model among a collection of models of various ‘flexibilities’ requires a bias-variance tradeoff. Simple models tend to have a high bias (i.e. they are far from truth) and a low variance (i.e. the response of the model is about the same, whatever the training set used to tune it). In contrast, complex models have low bias and high variance. A complex model tuned on too few examples tends to overfit these examples and to perform poorly on new inputs.

B. Model assessment and selection

There are several ways to control overfitting and to find a suitable bias-variance tradeoff. One can use an information theory criterion, such as AIC (Akaike’s “An Information Criterion”) or BIC (Schwartz’s Bayesian Information Criterion). These asymptotic criteria add a penalty \( P \) depending on the model complexity to the empirical risk \( \mathcal{R}_{\text{emp}}(h, S) \) defined in equation (2). The model having the lowest value of \( \mathcal{R}_{\text{emp}}(h, S) + P \) is selected.

Another way to proceed is to assess empirically the generalization error. Let us denote \( \mathcal{A} \) the algorithm used to learn a model from a dataset \( S \). In holdout cross-validation, the initial dataset \( S \) is split into two sets: a training set \( S_t \) used to learn the models, and another set \( S_v \) used to assess the holdout validation error \( \text{Err}_{\text{val}} \) as defined by the equation below:

\[
\text{Err}_{\text{val}}(\mathcal{A}, S_t, S_v) = P_{\text{emp}}(\mathcal{A}[S_t], S_v).
\]

\[ \text{(3)} \]

The model having the lowest holdout validation error is selected.

\[ \text{For example, let us assume we fit a polynomial curve on 10 points. For this regression problem, a polynom of degree 9 will fit exactly the examples, but will give poor predictions at other points. For a classification problem, the same overfitting problem might occur when using a K-nearest-neighbours method with } K = 1. \]
**K-fold cross-validation** is another popular empirical method, where the dataset $S$ is partitioned into $k$ folds $\left(S_i\right)_{i=1}^k$. Let us denote $S_i = S \setminus S_i$. In this method, $k$ separate predictors $\mathcal{A}(S_i)$ are learned from the $k$ training sets $S_i$. The mean of the holdout validation errors is computed, giving us the cross-validation estimation below:

$$CV_k(\mathcal{A}, S) = \frac{1}{S} \sum_{i=1}^k \left| S_i \right| \text{Err}_{\text{val}}(\mathcal{A}, S_i, S).$$

(4)

When used for model selection, cross-validation can be performed successively on a collection of models. The model having the best cross-validation error is selected.

**C. Hyperparameter tuning**

In many methods, the bias-variance tradeoff is controlled through one or several parameters. For example, one can think of the number of hidden units in a neural network, or the weight decay hyperparameter. Hyperparameter values can be selected through cross-validation.

Let us denote $\lambda$ the vector of hyperparameters of an algorithm $\mathcal{A}$. In this paper, a 5-fold cross-validation has been used to tune hyperparameters, as described in algorithm 1.

**Algorithm 1** Hyperparameter tuning for an algorithm $\mathcal{A}$ and a set of examples $T$ (training set).

```latex
\begin{algorithm}
\textbf{function} TUNEGRID($\mathcal{A}, \text{grid}$)[\$T\$] \\
\quad $\lambda^* \leftarrow \arg\min_{\lambda \in \text{grid}} CV_{5}(\mathcal{A}, T)$ \\
\quad \textbf{return} $\mathcal{A}_{\lambda^*}[\$T\$] \\
\end{algorithm}
```

**IV. The gradient tree boosting method**

In our experiments, we used the statistical software $R$, and more specifically the $\text{Xgboost}$ library for gradient boosted trees.

The stochastic gradient boosting tree algorithm was introduced in [25], [26], [23]. It applies functional gradient descent to classification or regression trees ([27]).

The functional gradient descent is a boosting technique. The model $h$ is iteratively improved. Denoting $h_m$ the current model at iteration $m$, we consider the opposite gradient of the loss $g_i = -\frac{\partial \hat{r}(\hat{y}, y)}{\partial y}(h(x_i), y_i)$. A model $g$ is then tuned to fit this opposite gradient, using a set of examples $(x_i, g_i)_{\text{train}}$. The model $h$ is then updated as follows: $h_{m+1}(x) = h_m(x) + \rho g(x)$, where $\rho$ is a constant minimizing the empirical risk. The next iteration repeats the same procedure for $h_{m+1}$ until a maximum number of iterations is reached. In stochastic gradient boosting, the dataset is randomly resampled at each iteration.

In the Gradient Tree Boosting, the machine learning algorithm boosted by the functional gradient descent is a classification or regression tree algorithm. Before continuing our description of gradient boosted trees, let us say a few words on classification and regression trees (CART) which were introduced by Breiman in [27]. In this algorithm, a binary tree is used to represent a binary recursive partition of the input space. At each node, the input space is split in two regions according to a condition $x_j \leq s$. The $J$ leaves of this tree describe a partition $\left(R_j\right)_{j=1}^J$ of the input space. Each region $R_j$ is associated to a constant $\gamma_j$. In the case of regression, it will be a constant float value (usually the
average value of the examples in region $R_j$. In classification trees, $y_j$ will be a class (the most represented class among the examples in $R_j$). When the tree is used to make a prediction on a new input $x$, the value $y_j$ is returned when $x$ falls into $R_j$.

CARTs have some advantages. For example, they are insensitive to input monotonic transformations: Using $x_j$, $\log(x_j)$ or $\exp(x_j)$ leads to the same model. As a consequence, this algorithm is robust to outliers. It can easily handle categorical variables and missing values. However, it is known to have a poor performance in prediction. This performance is greatly improved however when applying gradient boosting to CART.

In gradient boosted trees, the equation of the model update is the following, where $\nu$ is a shrinkage parameter:

$$ h_m : x \rightarrow h_{m-1}(x) + \nu \sum_{R_j \in \text{tree}} y_j \cdot 1_{R_j}(x) $$

(5)

We can denote $\text{GBM}(m,J,n)$ the gradient boosted tree algorithm, where $m$ is the number of boosting iterations, $J$ is the number of leaves of the tree and $\nu$ is the shrinkage parameter. The final model obtained after boosting is a sum of regression or classification trees. $J$ allows us to control the interaction between variables, as we have $J - 1$ variables at most in each tree. $\nu$ is the learning rate. In [23] (chap. 10), it is recommended to take small values for the shrinkage parameter ($\nu < 0.1$) and small values for $J$ as well ($4 \leq J \leq 8$). The hyperparameter grid used for this algorithm is presented in section V-E.

V. Data and experimental setup

A. Explanatory variables

In this study, ATC complexity indicators are used as inputs to our models. In previous works ([13], [28], [15]), we selected 6 basic complexity metrics among 27 metrics found in the literature. We used a principal component analysis to reduce the dimensionality of the inputs, and then selected the most relevant metrics related to the significant components. The 6 metrics that were found to be the most relevant for the purpose of building airspace configuration prediction models are the following:

- $\text{vol}$, the airspace volume of the considered ATC sector,
- $\text{nb}$, the number of aircraft within the sector boundaries at time $t$,
- $\text{flow}_{15}$, the incoming traffic flow within the next 15 minutes,
- $\text{flow}_{60}$, the incoming traffic flow within a 1 hour time horizon,
- $\text{avg}_v$, the average absolute vertical speed of the aircraft within the sector,
- $\text{inter}_h$, the number of speed vector intersections with an angle greater than 20 degrees.

These metrics are fairly simple and can be computed from radar tracks and static sector data (geometrical boundaries).

In the current paper, we have chosen to use the same metrics. They are standardized so as to obtain explanatory variables with mean 0 and standard deviation 1. These standardized variables are used as input vector $x$ in our models.
B. Target variable

The target variable \( y \) we are trying to predict with our models is a workload category: low, normal, or high. In order to build our examples, we extracted this workload variable from historical airspace configuration data. In many cases, the workload in an ATC sector \( s \) at a past time \( t \) can be quantified by considering how it was operated at that time \( t \). We simply make the following assumptions about the relationship between sector operation and workload:

- **Low workload** when sector \( s \) is collapsed with other sectors to form a larger sector operated on a working position,
- **Normal workload** when the sector \( s \) is operated as is,
- **High workload** when \( s \) is split into several smaller sectors operated on different working positions.

The other possible states – such as when a part of \( s \) is collapsed with one sector and another part is collapsed with another sector – are useless for quantifying the workload and are not used.

C. Dataset

The datasets used in this study are built from radar tracks and recorded sector operations from two weeks in October 2016 (13th to 26th), from the five french ATC control centers (Aix, Bordeaux, Brest, Paris, and Reims). For a given ATC sector, we sample the data so as to balance the occurrences among the workload classes having non-zero occurrences in the initial data. For example, for elementary sectors (or other sectors) for which there is no occurrence in the high workload category, we sample an equal number of instances in the low and normal categories.

We then select all the sectors with non-zero occurrences of the “normal” workload category (i.e., sectors that were opened at one moment or another). As a result, we obtain 50389 samples in the low workload category, 57539 in the normal workload category, and 21372 samples in the high workload category. This dataset is then completed by drawing samples from the sectors having no occurrences of normal workload (i.e., sectors that were never opened). The resulting dataset comprises 57539, 57539, and 32800 samples in the low, normal, and high workload categories, respectively.

This procedure is different from the one adopted in our previous work [1], where our dataset was built so as to obtain an overall balance among the three classes, without considering the balance in each subset corresponding to each sector. The resulting dataset contains 147878 complexity and workload samples concerning 163 elementary sectors and 369 ATC sectors made of several airspace sectors.

Note that the chosen data sampling procedure does not provide an exactly equal representation of the three classes in the resulting subset. One reason is that there exists no instance of the “high” workload class in the initial data for elementary sectors. Even for ATC sectors made of several airspace sectors, the sample might be imbalanced, for example when the sector is opened for very large periods throughout the day (i.e. there might not be enough data of the low or high class from which to draw samples).

---

2 By definition, elementary sectors cannot be split, so there is no occurrence of “high” workload according to our definition of workload based on the sector status (merged, collapsed, or split).

3 Note that the dataset description in section III-C of [1] is partly incorrect: although only sectors that were actually opened were initially selected, the dataset was completed using sectors that were not opened during the considered time period, in order to balance the number of instances among the three workload categories.

4 By definition, elementary sectors cannot be split into several smaller sector, so we cannot measure high workloads for such sectors with the chosen target variable (the sector status).
D. Performance evaluation and model selection

Our aim is to assess the performance of a model trained on a dataset extracted from past ATC sector operations, checking if this model generalizes well on new sectors that were not used to train the model. Consequently, our dataset must be split so that any ATC sector present in the subset used to train a model is not represented in the subset used to assess the model.

In addition, we would like examine how the trained model performs on ATC sectors made of only one elementary airspace sector. Considering these objectives, we use a nested cross-validation procedure, stratified so that the proportion of elementary sectors and collapsed sectors is about the same in all subsets.

The nested cross-validation consists in an outer 7-fold cross-validation for model performance assessment, embedding an inner 5-fold cross-validation for hyperparameter selection. The dataset $S$ is split in 7 subsets $S_i$, $1 \leq i \leq 7$. For each iteration $i$, a model is trained on $S_i = S \backslash S_i$ and its performance is assessed on $S_i$. The training on $S_i$ follows a 5-fold cross-validation procedure (the inner cross-validation), in order to select the best hyperparameter values for the chosen machine learning method (model selection). This is done by splitting $S_i$ into 5 subsets $S_{ij}$, $1 \leq i \leq 5$. A grid of hyperparameter values is used to tune several models on $S_{ij} = S_i \backslash S_{ij}$, and the performance of these models are evaluated on $S_{ij}$. The hyperparameter $\lambda^*$ providing the best performance on the subsets $S_{ij}$ is then selected, and a final model with hyperparameter $\lambda^*$ is trained on $S_i$. This is the model evaluated on $S_i$ in the outer cross-validation.

This procedure is more computationally intensive than the one chosen in our previous work [1], where the outer crossvalidation was a simple holdout validation, where the dataset was split into a training set and a test set only, instead of a 7-fold cross-validation. The new procedure has the advantage to provide some evaluation of the distribution of the performance results (considering the subsets $S_i$).

Also, in our previous paper, there were a large number of sectors overlapping in both the training and test sets. We were interested in evaluating how a model trained on one week of data would generalize on another week of data (the test set).

In the current paper, a different question is being adressed, motivating the change of procedure. We want to check if a model trained on a subset of sectors can generalize well on another subset of sectors, and if our model overfits the data for elementary sectors.

Knowing that there is no data with high workload available for the elementary sectors in our training sets, we might expect the model to overfit the training data and to generalize poorly on fresh examples, for elementary sectors. However, it should generalize correctly on the other sectors.

E. Hyperparameter grids

The hyperparameter selection of the inner cross-validation is performed on the training set, using function $\text{TuneGrid}$ of algorithm 1 and the grids described in table I.

**TABLE I — Grid of hyperparameters used in our experiments.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Hyperparameter grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = [5000, 6000, 7000]$</td>
<td></td>
</tr>
<tr>
<td>$J = [2, 3, 4]$</td>
<td></td>
</tr>
<tr>
<td>$v = [1e-4, 5e-4, 1e-3, 1e-2, 1e-1]$</td>
<td></td>
</tr>
</tbody>
</table>
F. Classification performance metrics

In this paper, the performance of a classifier is assessed through its accuracy, recall and precision. Accuracy is the total number of correct predictions made divided by the total number of predictions. Recall is the number of correct predictions made for one class, divided by the actual number of occurrences in the considered class. Precision is the number of correct predictions made for one class divided by the total number of instances predicted to be in that class.

**FIGURE 1 — Illustration of a confusion matrix.**

In other words, considering the example of a confusion matrix on Figure 1, accuracy, precision and recall are computed as follows:

\[
\text{Accuracy} = \frac{a + e + i}{a + b + c + d + e + f + g + h + i}
\]

\[
\text{Recall}(C1) = \frac{a}{a + d + g}
\]

\[
\text{Precision}(C1) = \frac{a}{a + b + c}
\]

VI. Results when generalizing on fresh sectors

In this section, we examine the performance of the GBM model when generalizing on fresh data taken from sectors that were not used when tuning the model. The model performance is assessed on several sub-populations of ATC sectors:

- All ATC sectors (elementary, or not),
- Elementary sectors, *i.e.* sectors that cannot be split into smaller sectors, and for which there is no occurrence of the "high workload" class\(^5\),
- Non-elementary sectors, *i.e.* ATC sectors made of several elementary airspace sectors, for which we detail the following results:
  - Overall performance on the non-elementary sectors
  - Model performance on non-elementary sectors for which there is no instance of the "high workload" class in the data.

A. Overall performance, all sectors included

Table II shows the performance of the GBM model, when including all sectors in the performance assessment. The first line shows the mean rates of correct classifications, and the standard deviations (within brackets). The overall rate of correct classification

\(^5\) With the chosen target variable, high workload can be observed only when the sector is split into several smaller sectors.
[2nd line, 2nd column] is the accuracy, and the class-specific rates [2nd line, columns 3, 4, and 5] are the recall. The precision of the model is given on the last line, where the overall precision is the average over the three classes.

**TABLE II — Model performance averaged over 7 folds, for all ATC sectors, using the GBM method.**

<table>
<thead>
<tr>
<th>Overall</th>
<th>Low</th>
<th>Normal</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct classif.</td>
<td>0.759</td>
<td>0.757</td>
<td>0.736</td>
</tr>
<tr>
<td>Precision</td>
<td>0.771</td>
<td>0.817</td>
<td>0.680</td>
</tr>
<tr>
<td>[0.03]</td>
<td>[0.046]</td>
<td>[0.074]</td>
<td>[0.100]</td>
</tr>
<tr>
<td>[0.034]</td>
<td>[0.065]</td>
<td>[0.042]</td>
<td>[0.112]</td>
</tr>
</tbody>
</table>

The mean values and standard deviations are computed from the 7 folds of the cross-validation, considering only the validation subsets Si (not used to train the model).

**B. Results for elementary sectors**

Table III shows the correct classification rates (overall accuracy, and class-specific recall), as well as the precision of the GBM model, for the control sectors made of only one elementary airspace sector. Such sectors cannot be split so as to alleviate the workload. As a consequence, there is no data concerning the “high” workload for these sectors.

For these results, the models trained in each fold of the cross-validation are exactly the same as in the previous subsection. They are trained on the same training subsets as before. However, the model performance is here evaluated considering only the elementary sectors in the validation subsets.

**TABLE III — Model performance averaged over 7 folds, for elementary airspace sectors only, using the GBM method.**

<table>
<thead>
<tr>
<th>Overall</th>
<th>Low</th>
<th>Normal</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct classif.</td>
<td>0.82</td>
<td>0.904</td>
<td>0.507</td>
</tr>
<tr>
<td>Precision</td>
<td>0.519</td>
<td>0.856</td>
<td>0.539</td>
</tr>
<tr>
<td>[0.092]</td>
<td>[0.062]</td>
<td>[0.197]</td>
<td>[NA]</td>
</tr>
<tr>
<td>[NA]</td>
<td>[0.112]</td>
<td>[0.271]</td>
<td>[NA]</td>
</tr>
</tbody>
</table>

**C. Results for ATC sectors made of several airspace sectors**

Table IV shows the results (correct classification rates and precision) obtained for the control sectors made of several elementary airspace sectors.

**TABLE IV — Model performance averaged over 7 folds, for ATC sectors made of several elementary airspace sectors, using the GBM method.**

<table>
<thead>
<tr>
<th>Overall</th>
<th>Low</th>
<th>Normal</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct classif.</td>
<td>0.757</td>
<td>0.739</td>
<td>0.748</td>
</tr>
<tr>
<td>Precision</td>
<td>0.759</td>
<td>0.814</td>
<td>0.679</td>
</tr>
<tr>
<td>[0.029]</td>
<td>[0.041]</td>
<td>[0.060]</td>
<td>[0.100]</td>
</tr>
<tr>
<td>[0.031]</td>
<td>[0.068]</td>
<td>[0.047]</td>
<td>[0.112]</td>
</tr>
</tbody>
</table>

Table V details the results for the non-elementary sectors for which there is no occurrence of the “high workload” class in the data, similar in that respect to the elementary sectors.

**TABLE V — Model performance averaged over 7 folds, for non-elementary sectors having no data instances in the “high workload” category.**

<table>
<thead>
<tr>
<th>Overall</th>
<th>Low</th>
<th>Normal</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct classif.</td>
<td>0.738</td>
<td>0.713</td>
<td>0.760</td>
</tr>
<tr>
<td>Precision</td>
<td>0.504</td>
<td>0.784</td>
<td>0.729</td>
</tr>
<tr>
<td>[0.033]</td>
<td>[0.058]</td>
<td>[0.089]</td>
<td>[NA]</td>
</tr>
<tr>
<td>[0.020]</td>
<td>[0.087]</td>
<td>[0.056]</td>
<td>[0.000]</td>
</tr>
</tbody>
</table>
When comparing the results in tables IV, V and III, we see that the model performance drops for the elementary sectors, with a recall barely above 50% for the “normal workload” class, but that it remains good for non-elementary sectors in general, and even those with no occurrence of high workload. This tends to show that our model overfits the data taken from elementary sectors specifically, and generalizes poorly on these sectors but not on the others. One reason might be that there are no other sectors in our data that would be similar to the elementary sectors in size and characteristic, and that would be well-balanced among the three classes. By contrast, we can find a lot of data samples from well-balanced (non-elementary) sectors having sizes and characteristics similar to the non-elementary sectors having no high workload occurrence. This might explain why our model still generalizes correctly on these sectors.

This is a good result, as it means we can still use our workload model to find optimal combinations of sectors, although we should replace or amend our model when assessing the workload in elementary sectors.

VII. Relationship between the input metrics and the workload classes

Figure 2 shows bagplots for sector N, for all pairs of explanatory variables and for the three classes (low, normal, or high workload). Bagplots are bivariate generalizations of the well-known unidimensional boxplots. The centroid corresponds to the median of the data. The inner bag is a convex hull containing 50% of the data points. The outer bag (called the loop) is obtained by expanding the inner bag by a factor 3. All points outside the loop are flagged as outliers.

![Bagplots for the sector N (Brest ATCC).](image)

Let us consider the graph on the second line and first column of Figure 2, representing the bagplots for the $\text{nb, flow15}$ pair of explanatory variables. We see that both $\text{nb}$ the number of aircraft (on the x-axis) and $\text{flow15}$ the flow of incoming traffic within the next 15 minutes (y-axis) are globally higher for data points in the “high” workload class than in the “normal” workload class. Similarly, the values of these explanatory values are higher in the “normal” class than in the “low” class. The centroids of the three workload
classes are distributed along the diagonal axis of the graph, with the lowest workload closest to zero. The same observations can be made for all pairs of explanatory variables, looking at the other graphs of the same figure, except those involving avg_vs, the average absolute vertical speed of all aircraft within the sector.

Considering the graphs on the 4th line or the 4th column, we see that for avg_vs the centroids of each workload class are more or less aligned. They are also fairly close to the x-axis, indicating that this en-route sector usually handles leveled traffic. Moreover, the dispersion of the vertical speeds is higher for low workload classes than for higher workload classes. Although this remains to be confirmed, this is probably because the biggest flows, inducing the highest workload in this sector, are the transatlantic flights and the northbound or southbound flights between the UK and southern Europe, which contain mostly leveled flights.

Now, considering Figure representing bagplots for another sector AENB in the Paris area, we see on the graphs of the 4th line that the median value for avg_vs is around 1200 fpm, whatever the workload class. This is probably because sector AENB is a pre-approach sector for Paris airports, in which nearly all the traffic is either climbing or descending. We see also that the three bagplots representing the low, normal, or high workload classes are more or less aligned horizontally, as for the sector N. The dispersion of the data is also higher in low workload classes than in higher workload classes.

The fact that the median value of avg_vs on Figures 2 and 3 remains more or less on a flat line whatever the workload class seems to contradict our intuitive notion that the cognitive workload of the controller should increase when more aircraft are climbing or descending in the sector, just as for the other explanatory variables.

**FIGURE 3 — Bagplots for the sector AENB (Paris ATCC).**

However, looking more closely at the outer bag of the “normal workload” class, on Figure 2, 4th line, 1st column, we see that its shape is more or less triangular, with the pointy end towards high values of nb, the number of aircraft. The obliqueness of the upper boundary of this triangle tends to show that high values of avg_vs are more acceptable to the controllers for small values of nb than for high values, at least for this sector. A similar shape can be observed on many other sectors. So, within the normal workload category, our intuition that the cognitive workload increases with the number of climbing/descending aircraft might actually be true.
This relationship is simply difficult to observe at the macroscopic level: the median values of \( \text{avg}_v \) are about the same whatever the workload category, which simply reflects the fact that there are no more climbing/descending aircraft, in our data, when the sector is split than when it is normally operated or collapsed.

In addition, the influence of \( \text{avg}_v \) must be considered across all the sectors. When doing so, and when combining it with other variables such as the sector volume, the average vertical speed \( \text{avg}_v \) does influence the workload categorization, as was shown in previous work on the selection of relevant explanatory variables ([13], [28], [15]). It remains to be seen if, for our purpose, it could be replaced with a categorical variable characterizing the sector (\textit{en-route}, or \textit{pre-approach}, for instance).

VIII. Conclusion

Let us now conclude this paper by summarizing our approach and our findings. We have looked into the performances of a workload model learned from historical data, using gradient boosted trees. The examples used to learn the model were made of ATC complexity measurements computed from radar records and sector data, and workload measurements extracted from past ATC sector operations. The three levels (low, normal, high) only give a rough indication of the workload. However, this workload measurement has the advantage of being easily available, in large quantities and for a great number of ATC sectors, because it can be directly extracted from historical records of past sector operations.

In previous works, this model showed an 82% rate of correct classifications, when training the model on one week of traffic, and assessing it on another week, considering approximately the same set of sectors in both the training and the test set. In the current work, our first objective was to look into the model’s performances when the model is trained on a subset of sectors and assessed on a different subset. Our second objective was to examine more closely the relation between the input ATC complexity variables and the output (i.e. the workload class).

The results show that the overall performance of the model is slightly degraded, with a rate of correct predictions around 76%, when the training and test sets are geographically segregated (different sectors) instead of being temporally segregated like in our previous approach. The detailed results show that our model probably overfits the training data for elementary sectors, leading to poor generalization performance for these sectors. However, the model remains performant on all non-elementary sectors, even those having no occurrence of the high workload class in the data.

Concerning the relations between the input variables and the workload output, the bagplots of section VII confirm some natural intuitions on the sense of variations of these quantities for most variables, except maybe for one, the average absolute vertical speed of all aircraft in the sector. For all variables except this one, we observe on two instances of sectors (one en-route and one pre-approach sector) that when the input ATC complexity metric increases, we are more likely to be in a higher workload category.

A natural intuition is that the cognitive workload of the controller should also increase when there are more climbing and descending aircraft in the sector, just as for the other variables. This relationship is most probably true, but it is difficult to observe in our macroscopic workload categorization, except maybe by looking closely at the dispersion of the data in the normal workload category.

To conclude, the model obtained with the GBM machine learning method cannot be interpreted only as a model of the cognitive workload of the controller. The model expresses relations among variables emerging from the data, and it can only be as good as the data that was used to train it. It cannot be transposed to any context without precautions. The fact that our workload model remains performant for all non-elementary
sectors confirms that it can actually be used to predict optimal configurations of ATC sectors (sector opening schemes), where we only search to split or merge sectors optimally. This model should be completed or replaced by a more simple model when evaluating the workload in elementary sectors.

In future works, we might try to produce some artificial data samples of the “high” workload class for elementary sectors. This would force our model to correctly assess the boundary between normal and high workload for these specific sectors. Another approach that we could try is to apply one-class classification methods on the “normal” workload class, to detect when non-normal instances (underloads, or overloads) occur in elementary sectors. Other work might consider the seasonal variability in our data. It would be interesting to compare the performances of a same model tuned several times on data samples of different months.

REFERENCES


DEVELOPMENT OF NEUROMETRICS FOR SELECTIVE ATTENTION EVALUATION IN ATM

Gianluca Borghini¹, Martina Ragosta², Pietro Aricò¹, Stefano Bonelli², Gianluca Di Flumeri¹, Nicolina Sciaraffa¹, Paola Tomasello², David Mancini¹, Alfredo Colosimo¹, Fabio Babiloni¹
¹Sapienza University of Rome, Rome, Italy
²Deep Blue Research and Consulting, Rome, Italy
[gianluca.borghini; pietro.arico; gianluca.diflumeri; nicolina.sciaraffa; david.mancini; alfredo.colosimo; fabio.babiloni]@uniroma1.it
[martina.ragosta; stefano.bonelli; paola.tomasello]@dblue.it

Abstract—The European Air Traffic Management (ATM) system is expected to face challenging situations, with the growth of air traffic, the increase of its complexity, the introduction of innovative concepts and increased automation. The roles and tasks of Air Traffic Controllers (ATCOs) will change in the future and it is vital to enhance the comprehension of human responses to their role changing, that is, from active control to monitoring of complex situations and managing unexpected system disruptions. ATCOs performance is recognised to be impacted by several aspects such as stress, emotions, available attentional resources, attention focus and so on. In the recent years, the concept of Human Performance Envelope (HPE) has been introduced as new paradigm in Human Factors. Rather than focusing on one individual factor (e.g. fatigue, situation awareness, etc.), the HPE considers their full range, mapping how they work alone or in combination leading to a decreased performance that could affect safety. At the EU level, the STRESS project is currently addressing the research goal of monitoring the operator’s performance by including all the available behavioral and neurophysiological data in order to characterize the Human Factors involved when dealing with the considered task. In line with this, the proposed study will show the results coming from the experiments performed in controlled environment for the evaluation of the user’s attentional level, and the definition of the neurophysiological indexes for the characterization and assessment of the selective attention along the execution of ATM activity.

I. Attention in air traffic management

The latest SESAR Global demonstrations, such as the one regarding the System-Wide Information Management (SWIM), have shown that the development and evolution of the “next” generation of innovative and unconventional ideas, concepts and technologies that define the performance of the future European Air Traffic Management (ATM) system, and contribute to its successful evolution, are already in place [1]. This new era not only concerns the introduction of advanced technologies, but it will deal with a revised founding principles and building blocks of information sharing, service orientation, federation, open standards, and information and service lifecycle management. In compliance with the International Civil Aviation Organization (ICAO) and European Aviation Safety Agency (EASA) regulations and directives [2], [3], SESAR is delivering the performance necessary to meet the growing demand for air transport from a worldwide perspective in order to achieve performance ambition levels for 2035 through significant operational changes.
These consist of improvements to technical systems, procedures, human factors and institutional changes supported by standardisation and regulation [4].

In the near future, we may see the introduction of a new generation of highly automated supporting technologies that are able to autonomously (or partially autonomously) manage tasks that are currently carried out by human operators and/or to provide inputs to human decisions that the operators will hardly be in a position to question. The introduction of higher levels of automation will bring about a new task allocation between the human and the machine [5]. Tasks previously carried out by the human, for example the provision of separation, are supposed to be partially delegated to the system. Highly automated systems are expected to take over operators’ repetitive tasks, while human role is expected to be focused on strategic planning, intervening on exceptions and monitoring the system’s behavior [6][7]. In general, rather than governing directly flight operations, pilots and Air Traffic Controllers (ATCOs) will likely supervise the automated systems. So, the ATCOs role would be shifted from active to passive (e.g. system monitoring) [8].

This implies the need for a radical revision of the competences required to perform the tasks [9] as well as a refinement of the Human Performance Envelope (HPE) and all its aspects such as stress, emotions, available attentional resources, attention focus and so on. Indeed, assuming that the equipment and the pilots all perform correctly, the controller’s workload would be expected to decrease in nominal situations, since they are interacting with fewer aircraft and supported by improved automation systems. It also seems plausible that the controller’s situational awareness of the airspace would decrease as well, since they would not be focusing as much attention as they previously did on many of the aircraft [10]. However, for an unknown period of time, certain aircraft in the system would be properly equipped to participate according to the new rules, allowing more autonomy, but others would not be, and the controller would be required to manage their flight path. Consequently, ATCOs have to process selectively the vast amount of information whom they have to face, prioritizing some aspects of information while ignoring others by focusing on a certain location or aspect of the visual scene [11].

Attention can be classified into three main components [12]:

1. Selective attention: The ability to process or focus on one message in the presence of distracting information.

2. Divided attention: The ability to process more than one information at the same time.

3. Visual attention: The mechanism determining what information is or is not extracted from our visual field.

One important concern for a notification system in ATM field is that seemingly prominent objects in the visual field can sometimes elude attention despite their relevance and importance to the primary task [13]. This lack of attention or distraction usually affects Human Performance (HP) by causing the omission of procedural steps, forgetfulness to complete tasks, and taking shortcuts that may not be for the better. A performance decrement can be noticed when attention, workload and task difficulty increase; the reaction time and number of errors increase as well, while accuracy and number of completed tasks decrease [14][15]. Reduction of the performance in monitoring, tracking, auditory discrimination, and reduction of visual field can be observed too[16].

One of the objectives of the STRESS project (human performance neurometricS Toolbox foR highly automatEd SystemS design) is to support the aforementioned transition to higher automation levels, by addressing, analysing and mitigating its impact on the HP [17]. In fact, the roles and tasks of ATCOs will change in the future and it is vital to enhance the comprehension of human response to changes in role, monitoring of complex situations, unexpected disruptions. It is also vital to develop tools to investigate such aspects and to monitor in real time controllers’ fitness to the task, anticipating risks and problems.
II. Neurometrics for Selective Attention

Attention is the ability to select only the relevant part of an information ignoring the useless ones\cite{18}. Attention is adopted for a wide range of everyday activities, like driving a car, watching a movie, or talking with friends, and it becomes very important and even necessary in most of the workplaces like a hospital, or an airport or an air traffic control room. Because of the significant role played by the attention in so many different fields, researchers from multiple disciplines like neuroscience, psychology, and ergonomics focused their work on the study of the attentive processes. Lacks of attention during safety risk activities may lead to catastrophic outcomes. Attention in operational environment might be studied observing and analysing people while performing their job allowing to understand what are the aspects of the workplaces, i.e. interactions with colleagues or instrumentation, or work demands, i.e. time pressure or co-working, that may interfere with attention, and operate on the environment and technologies according with the findings. To get more replicable and general information about attention, the most appropriate method is the laboratory setting, by adopting tasks and protocols created ad-hoc for a proper evaluation of the attention with full control on the participants, experimental conditions, and surrounding environment.

According to the most solid theories, attention is a multifaceted concept. Firstly, it is defined as a set of cognitive processes that lead to discriminate useful information in a framework of distractors \cite{18}, but this is only one of the characterizing aspects. In line with \cite{19}, \cite{20}, attention is divided in two main domains: intensity and selective aspects. The Intensity aspect of attention embraces alertness and sustained attention (or vigilance \cite{21}): task execution with an optimal level of performance is possible because for the entire duration of the task there is an appropriate level of arousal managing resources involved in orienting and selecting \cite{22}. This capacity of controlling the focus represents the second main aspect of attention that involves the Selective and the divided attention. Each of these attentive components is elicited by opportune tasks and shows neurophysiological correlates. The role of the Autonomic Nervous System during attentional processes has been less investigated, with respect to the Central Nervous System (i.e. brain activity) without taking into account pure clinical research. In general, the widely demonstrated concept is that intensive cognitive processes, such as mental effort and attention, are accompanied by increased autonomic arousal \cite{23}. In this regard, the objective of the proposed work was to assess the level of user’s selective attention through a multimodal approach, by considering several behavioural and neurophysiological signals, such as the Electroencephalogram (EEG), Electrocardiogram (ECG), Electrooculogram (EOG), Galvanic Skin Response (GSR), Reaction Time (RT), and number of correct responses.

For the selective aspect of attention, scientific literature reports significant theta and beta EEG activity increasing, and alpha EEG decreasing in correspondence of a target revelation\cite{24},\cite{25},\cite{26}. Furthermore, higher desynchronization (i.e. decrement) in the alpha EEG band over the left brain hemisphere than in the right one, and on the anterior brain cortex than on the posterior brain areas were found\cite{27}. Evidences on the sustained aspect of attention showed that an attentional increase was reflected by an Heart Rate (HR) decreasing, because of the assumption that during HR deceleration the brain cortex would be activated, which in turn would facilitate processing of external stimuli \cite{28}. Also, studies in selective attention found a decrement of low frequencies in heart rate variability (HRV) \cite{24}. Skin Conductance Level (SCL) and Response (SCR) are considered the gold standard measures of the attention \cite{29}: higher the user’s attention level is, higher SCL and SCR are. Also, the magnitude of the SCR, i.e. its peaks amplitude, is affected by short term changes in the general level of attention of the participant and also by the amount of attention being directed to a particular stimulus\cite{30}. These evidences are summarized in the following table.
TABLE I — Literature evidences

<table>
<thead>
<tr>
<th>Features</th>
<th>Modulation for higher attention level</th>
<th>Ref.</th>
</tr>
</thead>
</table>
| EEG      | Increment of theta and beta band activity.  
Decrement of alpha activity more in left and frontal brain areas than in right and posterior brain | 24÷27 |
| HR       | Decrement                             | 28    |
| HRV      | Decrement of low frequencies          | 24    |
| SCL      | Increment                             | 29    |
| SCR      | Increment in peaks amplitude          | 29-30 |

III. Material and methods

A. Participants

The experiments were conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000. It received the favorable opinion from the Ethical Committee of the Sapienza University of Rome. The study involved only healthy students from the Sapienza University of Rome recruited on a voluntary basis. Informed consent was obtained from each participant on paper, after the explanation of the study. Thirteen healthy volunteers (27±3 years old) their informed consent for taking part at the experiment. In particular, 7 males and 6 females took part at the experiments. All the participants took part to a practice session before starting with the experiment in order to avoid results due to learning processes.

B. Conjunction Visual Search Task (CNJ)

The Conjunction Visual Search Task (CNJ) consists in presenting visual stimuli on a screen and finding out the target among distractors, and reacting as fast as possible by pressing the space bar on the keyboard [31]. Both target and distractors were rectangular bars (size: 0.5*1.6 Visual Angle - VA): in particular the target was a red vertical bar, while the distractors could be green or red and vertical or oblique bars depending on the conditions (see below for details). No action was required when the target was not presented. All the stimuli were presented on a black background on a 25 position’s matrix filled with 8 elements: 7 distractors and 1 target (target events), or 8 distractors (no-target events). The matrix was presented at the participants for 2 seconds and between two trials a fixation cross was presented at the centre of the screen for a random interval between 0.25 ÷ 1 second.

C. Experimental Protocol

The experimental protocol consisted in accomplishing the Conjunction Visual Search Task (CNJ) under three conditions requiring different levels of attention (Figure 1). In particular, the Easy (E) condition was a pre-attentive level (low attention) based on one feature (colour); the Hard (H) condition was based on an orientation feature (medium attention), while the Conjunction (C) condition was based on two features (colour and orientation - high attention). In particular, in the E condition, the distractors were green vertical rectangular bars. In the H condition, the distractors were red rectangular bars rotated by 45°.
FIGURE 1 — The CNJ task has been proposed under three different conditions. The Easy condition was a pre-attentive level (low attention) based on one feature (colour); the Hard (H) condition was based on an orientation feature (medium attention), while the Conjunction (C) condition was based on two features (colour and orientation - high attention).

Finally, in the C condition the target was defined by two different features, colour and orientation, while the distractors were vertical and 45° rotated green and red rectangular bars. The conditions within the CNJ task (Easy, Hard and Conjunction) were randomized in order to avoid any habituation and expectation effects. The task was divided into two blocks comprehending 180 trials each. In each block participants performed 60 trials of three different conditions which required different level of attention. The participants performed 10 practice trials per condition before starting with the experiments. Each experimental trial included 30 target events and 30 no-target events. The experimental protocol lasted about 45 minutes in total.

During the entire experiment, the EEG, ECG, and EOG (used to remove eye-related artefacts from the EEG signal) signal were recorded using a high-resolution 64-channel system, while the Galvanic Skin Response (GSR) was recorded using the Shimmer GSR+ devices. Participants seated at a distance of 60 cm from the monitor (Figure 2). This preparation was followed by the baseline period of data collection for all neurophysiological variables. During the baseline period, the participants were asked to sit calmly and rest for a minute with their eyes open, and then a minute with the eyes closed. Right after the baseline, the participants filled the Visual Analogue Scale (VAS) in order to collect their subjective baseline, and during the experiment, at the end of each condition (E, H and C), participants were asked to rate their perceived attentional level using the VAS scale. The VAS comprised four items: Alertness, Attentiveness, Interest and Motivation. Each of the items consisted in a 100 points scale between opposite states (i.e. Attentive/Dreamy, Interested/Bored). This modified version of the scale was taken from [32].
FIGURE 2 — Experimental setup. The participant seated comfortably in front of the screen where the stimuli were presented on. During the entire experiment, the EEG, ECG, EOG signal were recorded using a high-resolution 64-channel system, while the Galvanic Skin Response (GSR) was recorded using the Shimmer GSR+ devices.

To assess both the accuracy and the speed of the user within one synthetic index, the combination between the Reaction Time (RT) of the corrected responses divided by the percentage of correct responses has been used (Equation 1).

$$Inverse\ Efficiency\ Score\ \ (IES) = \frac{RT_{TRUE}}{\%TRUE_{resp}}$$

In cognitive research, such an index is called the Inverse Efficiency Score (IES) [33], and it has been used to compare the performance across the three different levels of attention required during the CNJ task (Easy, Hard and Conjunction).

D. Statistical Analysis

Subjective Data - By the VAS questionnaire, the participants rated the perception of the attention demand across the different experimental phases (E, H, and C). Repeated measure ANOVA has been performed on such scores by considering as within factor the attention conditions (3 levels: E, H and C).

EEG Data - The Power Spectral Density (PSD) of the different EEG rhythms have been normalized with respect to the corresponding References conditions by means of two-tail paired t-tests. Then, repeated measures ANOVAs have been performed on the normalized PSDs values of each EEG channel and band by considering as within factor the different experimental conditions. When the results of the ANOVA was significant (p<0.05), the post-hoc analysis was performed as well. Then, the differences between couple of conditions (e.g. E vs H, E vs C, and H vs C) in correspondence of significant results (p<0.05) were saved, while no significant comparisons were set equal to zero. At the end, for each EEG channel, we obtained a structure were only the statistically significant comparisons, EEG channels and bands were kept. Finally, on such EEG channels, repeated measure ANOVAs have been performed with the aim to assess how the considered EEG rhythm changes across the different experimental conditions (within factor: experimental conditions) and to define the neurometrics for the evaluation of the investigated cognitive phenomena (e.g. selective attention).
**ECG Data** - The Heart Rate (HR) and LF/HF ratio of the Heart Rate Variability (HRV) [34] across the experimental conditions of the CNJ have been normalized by computing the z-score for each participant [35]. Then, repeated measures ANOVAs and Duncan’s post-hoc tests have then been performed on the two ECG-based indicators with the aim to find out eventual differences across the experimental conditions.

**GSR Data** - The mean value of the SCL and the mean amplitude of the SCR peaks [30] during the experimental conditions of the Conjunction Visual Search (CNJ) have been normalized, by computing the z-score, for each participant in order to perform group analysis. In particular, the two GSR-based indicators have been compared by means of repeated measures ANOVA. Duncan’s post-hoc test has been then performed to assess possible differences between the attention conditions.

**IV. Results**

**A. Subjective Results**

The result of the repeated measures ANOVAs on the VAS scores is reported in Figure 3. The perception of the attention across the experimental conditions did not report any statistically significant changes \((F(2, 22) = 0.36; p = 0.7)\). However, the trend of the graph reveals that the Conjunction condition was perceived more attentional demanding than the previous ones [Easy and Hard].

**B. Behavioural Results**

Figure 4 reports the variation of IES index across the Easy, Hard, and Conjunction conditions. The ANOVA showed that the participants reacted significantly faster in the Easy condition than in the Hard and Conjunction ones \((p < 10^{-4})\).
FIGURE 4 — IES values throughout the Easy, Hard and Conjunction conditions. The ANOVA showed that the participants reacted faster in the Easy condition than in the Hard and Conjunction ones (all p<10^{-4}).

C. Neurophysiological results

EEG Data – In the following figures (Figure 5÷10), the results of the repeated measures ANOVAs have been reported for each EEG rhythms over the brain areas in which significant changes were found. In particular, the spectral maps report only significant changes (PSDs increment and reduction were plotted in red and blue shades, respectively; on the contrary, no significant differences were coloured in grey), and above them the statistical trends over the brain areas (frontal, central, parietal, and occipital). The asterisks highlight the experimental conditions resulted statistically different (p<0.05) from the others. When no asterisks are reported, the considered conditions were not statistical different. When the attention demand increased from Easy to Hard, and finally to the Conjunction condition, significant increments of the theta, beta and gamma EEG bands have been found over the whole brain.

In particular, theta (F(2, 22) = 13; p < 10^{-3}), and beta (F(2, 22) = 4; p < 0.03) reported the highest increments over the parietal and occipital brain areas (Figure 5÷8).

FIGURE 5 — ANOVA results on the EEG theta activity performed over the different brain areas. The asterisks highlight the experimental conditions resulted statistically different (p<0.05) from the others.
FIGURE 6 — Cortical maps of the EEG theta activity over the different brain areas. PSDs increment and reduction were plotted in red and blue shades, respectively; on the contrary, no significant differences were coloured in grey.

FIGURE 7 — ANOVA results on the EEG beta activity performed over the different brain areas. The asterisks highlight the experimental conditions resulted statistically different (p<0.05) from the others.

FIGURE 8 — Cortical maps of the EEG beta activity over the different brain areas. PSDs increment and reduction were plotted in red and blue shades, respectively; on the contrary, no significant differences were coloured in grey.

On the contrary, the gamma band exhibited significant increment over the central (F(2, 22) = 8.15; p < 0.01) and parietal (F(2, 22) = 6.11; p < 0.007) brain areas (Figure 9 and 10).

Finally, between the medium and high attentional conditions, significant reduction of the frontal left gamma band has been reported (F(2, 22) = 6.11; p < 0.007).
FIGURE 9 — ANOVA results on the EEG gamma activity performed over the different brain areas. The asterisks highlight the experimental conditions resulted statistically different (p<0.05) from the others.

FIGURE 10 — Cortical maps of the EEG gamma activity over the different brain areas. PSDs increment and reduction were plotted in red and blue shades, respectively; on the contrary, no significant differences were coloured in grey.

In addition, significant asymmetry within the beta and gamma bands over the posterior lobes was found, where the right hemisphere was more involved than the left one. Finally, between the medium (H) and high attentional (C) conditions, significant reduction of the frontal left gamma band has been reported (p<0.007).

No significant differences have been found within the alpha EEG band over the considered brain areas.

ECG Data - Neither the Heart Rate (HR) nor the Heart Rate Variability (HRV) showed significant differences among the different attention demand conditions. In other words, they did not change in response to variations of selective attention.

GSR Data- By one hand the SCL did not show significant variations among the different conditions of the Conjunction Task. On the contrary, by the other hand the peaks amplitude of the SCR increased significantly (p = 0.035) from the Easy to the Conjunction condition (Figure 11). In particular, the Duncan’s post-hoc analysis highlighted that both the Hard and the Conjunction conditions induced SCR peaks significantly higher (respectively p = 0.03 and p = 0.02) than those ones induced by the Easy condition, whilst no significant difference was found between them (i.e. Hard and Conjunction).
V. Discussions

The analysis of behavioural data (IES index) showed significant slower reaction time and higher percentage of correct responses in the Easy condition with respect to the others two. This can be explained by the pre-attentive nature of this condition. Instead, in the Hard and Conjunction conditions the IES increased significantly due to higher attentive resources required by the subjects to accomplish the task and discriminate two features in order to identify the target.

The self-reported measures during the CONJ task did not provide any significant differences in terms of attention perception.

When the attentional demand became high (conjunction condition) the theta band increased significantly over the posterior areas, while the beta and gamma bands increased significantly along the midline throughout the whole brain areas. Also, the beta band kept the asymmetry on the right hemisphere by showing an enhanced activity.

Concerning the activation of the Autonomic Nervous System in response to the different types of attention, the cardiac parameters, HR and HRV appeared to be insensitive to the variations of selective attention.

On the contrary, the SCR peaks amplitude seems to be very sensitive to the selective sphere of attention. Although the Hard and Conjunction conditions appeared no significantly discriminable, this effect probably depends on an intrinsic poor discriminability of the two conditions, since also neither the subjective (i.e. VAS questionnaire) nor the behavioural measures (i.e. IES index) were able to discriminate them.

Despite the small experimental sample, since the results were derived from controlled settings, they provided robust evidences for the assessment of the attention level while dealing with realistic tasks. In fact, the evidences suggest to select the theta, beta and gamma EEG bands and the SCR component of the GSR in order to define an index able to track the user’s attention level.
VI. Conclusions

The results showed the possibility to assess different levels of the user’s attention. In particular, they highlighted the current limitation in using single neurophysiological signal rather than a combination of them. In fact, by considering only the behavioural or GSR data, it was possible to discriminate only two levels of attention [Figure 4 and 11]. On the contrary, if the EEG was integrated with them, the resolution of the neurometric would allow to measure three levels of attention (Figure 5÷10). In addition, the results suggest to define such a neurometric as a combination of the EEG estimated along the midline of the brain, and the SCR component of the GSR.

These evidences will be used for the next phase of the STRESS project: the First Validation. In the first validation, experiments will be performed at the University of Anadolu (Eskişehir, Turkey) by recruiting professional ATCOs and asking them to manage a realistic ATM scenario in which specific events will be designed with the aim to elicit different level of attention, stress and workload. In particular, such events will be designed with the support of Subject of Matter Expert (SME) and HF Experts, and by considering the experimental tasks used in the laboratory experiments. In fact, the CNJ task was selected by taking into account both the scientific validity and reliability of the task itself, and the similarity with the ATC activities, such as identify a specific aircraft among all the others, time and stressing pressure. The evidences will be then used in the first validation with the aims of 1) testing the proposed indexes for attention, stress, and workload evaluation in ecological settings, and 2) eventually combining them by considering the new evidences and conditions coming from realistic activities (e.g. more talking and movements).

Acknowledgment

This paper describes some activities performed in the context of the STRESS project which has received funding from the SESAR Joint Undertaking under grant agreement No 699381 under European Union’s Horizon 2020 research and innovation.

We thank our project partners Jean-Paul Imbert, Géraud Granger and Railane Benhacene from École Nationale de l’Aviation Civile (Toulouse, France) and Uğur Turhan, Birsen Açıkel, Ali Ozan Canarslanlar and Serkan Bayar from University of Anadolu (Eskişehir, Turkey) who provided insight and expertise that greatly assisted the research.

REFERENCES

[33] R. Bruyer and M. Brysbaert, "Combining speed and accuracy in cognitive psychology: is the Inverse Efficiency Score (IES) a better dependent variable than the mean Reaction Time (RT) and the Percentage of Errors (PE)?," Psychol. Belg., vol. 51, no. 1, pp. 5–13, 2011.
MACHINE LEARNING OF CONTROLLER COMMAND PREDICTION MODELS FROM RECORDED RADAR DATA AND CONTROLLER SPEECH UTTERANCES

Matthias Kleinert¹, Hartmut Helmke¹, Gerald Siol¹, Heiko Ehr¹, Michael Finke¹, Youssef Oualil², Ajay Srinivasamurthy³
¹ Institute of Flight Guidance, German Aerospace Center (DLR), Braunschweig, Germany,
² Spoken Language Systems Group (LSV), Saarland University (UdS), Saarbrücken, Germany,
³ Idiap Research Institute, Martigny, Switzerland
firstname.lastname@{dlr.de, lsv.uni-saarland.de, idiap.ch}

Abstract—Recently, the project AcListant® related to automatic speech recognition has achieved command recognition error rates below 1.7% based on Assistant Based Speech Recognition (ABSR). One main issue to transfer ABSR from the laboratory to the ops-rooms is its costs of deployment. Currently each ABSR model must manually be adapted to the local environment due to e.g. different accents and models to predict possible controller commands. The Horizon 2020 funded project MALORCA (Machine Learning of Speech Recognition Models for Controller Assistance) proposes a general, cheap and effective solution to automate this re-learning, adaptation and customization process to new environments, by taking advantage of the large amount of speech data available in the ATM world. This paper presents an algorithm which automatically learns a model to predict controller commands from recorded untranscribed controller utterances and the corresponding radar data. The trained model is validated against transcribed controller commands for Vienna and Prague approach. Command error rates are reduced from 4.1% to 0.9% for Prague approach and from 10.9% to 2.0% for Vienna.

I. Introduction

A. Problem

One of the main causes hampering the introduction of higher levels of automation in the Air Traffic Management (ATM) world is the cost factor. ATM system suppliers try to reduce costs by developing generic systems, e.g. one basic Arrival Manager like MAESTRO [1] which fits for many airports. Therefore, the deployment of decision and negotiation support tools in current ATM business still requires a strong and manual adaptation to the local environment to avoid low end-user (controller) acceptance. Every single process of adaptation yields a significant cost increase for a core ATM system so that total system costs easily exceed the threshold of one million Euros.

The work is partly funded by SESAR Joint Undertaking (Grant Number 698824).
To ensure the acceptance of any new feature developed by any ATM project, it is imperative that its benefits are clearly recognizable for the end-user at the very beginning when they are confronted with new tools. If the outcomes of the development are not of certain quality, new problems or additional workload are introduced to the end-user and they will likely refuse the acceptance and cooperation, i.e. when a new tool is demonstrated to end-users it should already be tailored to the local environment. On the other hand if we provide end-users with innovative and helpful tools with a promising perspective that further improvements can be achieved, strong support from the end-users can be expected for the further development.

In aviation, Automatic Speech Recognition (ASR) is a known technology used with considerable success in ATC training simulators. Recently, the venture capital funded project AcListant® [2] has achieved command error rates below 2% in operational environments and fuel reduction of 60 liters per aircraft (based on an A320) with a specific speech recognition technology called Assistant Based Speech Recognition (ABSR), developed by Saarland University (USAAR) and DLR [3]-[6]. ABSR combines ASR with an assistant system, which generates context information to reduce the search space of the speech recognizer.

To manually adapt the ABSR models to different local environments, large amounts of speech data need to be collected and manually transcribed, especially, if the mismatch between the original models and the new environment is large. It requires significant human effort and the process of re-training requires vast expertise. The new data is generally used to adapt the acoustic model, the language model, and the rules to generate the context for the local environment. Moreover, partial adaptation may be required also during the life time of the system given that involved controllers, waypoints, procedures or used phraseology, etc. may change, resulting in new costs. Today the adaptation process requires a notable amount of resources and hence causes considerable costs. The estimate is that these adaptations costs represent at least 10% of the total costs [7]. Both end-user partners, i.e. Austro Control and ANS CR, are the first test sites who will explore proposed solutions. To conclude: Less expensive adaptation to the local needs without compromises with respect to end-user acceptance is needed.

B. Solution

The Horizon 2020 SESAR project MALORCA (Machine Learning of Speech Recognition Models for Controller Assistance) proposes a general, cheap and effective solution to automate this adaptation and customisation process. Adaptation of speech recognition models were selected as a first show-case of MALORCA [8]. The MALORCA consortium consists of two members from academia, Saarland University (Germany) and Idiap Research Institute (Switzerland), Air Navigation Service Providers from Czech Republic (ANS CR) and Austria (Austro Control) representing the user needs, and the German Aerospace Center (DLR) as the connecting element between fundamental research and business needs. The proposed solution builds on the huge amount of target data recorded every day in the operation rooms.

As shown in Figure 1, each Air Navigation Service Provider (ANSP) generates Mega Bytes or even Giga Bytes of radar data and voice recordings on a daily basis. These recordings can be the input for machine learning algorithms. The outputs would be improved speech recognition models which are adapted to the local needs. These initial models can be improved day by day. If a new waypoint is added it would be learned, if a waypoint is removed it would be “unlearned” etc.
C. Paper Structure

In the next chapter we present related work with respect to machine learning and speech recognition applications in ATM. In chapter III the Command Prediction Model which is the focus of this paper is shown in detail. The performed proof-of-concept experiments are described in chapter IV. The results can be found in chapter V. Afterwards results and next steps are discussed in chapter VI before we draw our conclusions in chapter VII. First results on learning the acoustic and language models can be found in [9] and [10].

II. Related work

A. Speech Recognition Applications in ATM

Speech Recognition applications have dramatically improved during the last decade (e.g. Siri®, Alexa, Google Assistant). The integration of ASR in ATC training started already in the late 80s [11]. Today ASR applications go beyond simulation and training. ASR is e.g. used to get more objective feedback of controllers’ workload [12]. Chen and Kopald used speech recognition to build a safety net for airport surface traffic to avoid aircraft using a closed runway [13]. Most recently they presented an approach to detect pilot read back errors [14].

Although the vocabulary in controller pilot communication is quite limited and phraseology is restricted, recognition rates are still far from being perfect. One promising approach to improve ASR performance is using context knowledge regarding expected utterances. These attempts go back to the 80s [15], [16]. This information may heavily reduce the search space and lead to fewer misrecognitions [17]. The approach developed by DLR and Saarland University uses the output of an Arrival Manager (AMAN) as context information [3]. The AMAN (4D-CARMA) in Figure 2 analyzes the current situation of the airspace and predicts possible future states used by the “Hypotheses Generator” to predict a set of possible commands. This dramatically reduces the search of the “Lattice Generator” [18], [19]. The search lattice is dynamically regenerated and contains a search tree for all possible phoneme sequences determined by the “Hypotheses Generator”. The “Speech Recognizer” finds the most probable path in the search tree. The output of the “Command Extractor” is checked again by the “Plausibility Checker”, determining whether the recognized commands are reasonable in the current situation.
In [5] command recognition rates (RecR) for ABSR of 95.2% and command recognition error rates (ErrR) below 2% are reported for Dusseldorf Approach Area.

B. Supervised and unsupervised learning

Recent advances in machine learning have significantly improved human-machine interaction systems by understanding the context of interaction and adapting to it. Given features describing data and possibly output labels, machine learning aims to model the rules that can map the input features to output labels. Output labels can be categorical, in which the task is called classification, or could be continuous valued in a regression task.

Machine learning methods can be supervised, unsupervised or semi-supervised [20], [21]. In supervised learning, we require data samples and corresponding output labels, and several different algorithms can be used to learn the input-output relationship. However, in unsupervised learning, the output labels are missing and the machine learning algorithm just uses the data examples to learn both output labels and the rules to model data. Typical unsupervised learning approaches include data clustering to partition data according an optimization criterion. In semi-supervised learning, partially labeled set of examples are used to build a machine learning model.

Supervised learning approaches utilize different methods to model the input data to output labels. Decision trees [22], [23] use a series of nested rules to compare input data to arrive at the output label. The rules of the tree are pre-determined from expert knowledge or learned from input examples to obtain an optimal partitioning of data using those rules. Decision trees can effectively learn complex input-output relationships with limited data and computational resources. Recently, neural network based models have been shown to accurately learn arbitrary input output relationships. Neural network models require extensive computational resources and are mainly effective when large amount of examples are available, e.g. to build acoustic models for the ABSR system [24].

III. Machine Learning of Command Prediction Models

Assistant Based Speech Recognition (ABSR) normally uses three main models, which need to be trained/adapted for each ATC environment [approach area] separately:

1. Acoustic Model,
2. Language Model (e.g. grammar) and
3. the Command Prediction Model (CPM).
A. Model Interaction within ABSR

Figure 3 shows in the upper part how those three models are used within ABSR. The dark blue ellipses represent the models; rectangles describe tasks and ellipses with a lighter blue show additional inputs and outputs.

At first the CPM is used by the Hypotheses Generator in Figure 3 to derive a set of commands (Command Hypotheses), which are possible in the current situation. These commands are used as input for ASR to reduce the search space size and to guide the search process of the speech recognition system.

FIGURE 3 — Interaction of Acoustic, Language and Command Prediction Model

The other two models (acoustic and language) are directly used by ASR. For a controller utterance given as audio signal, the acoustic model and the language model are used to extract the sequence of spoken words, e.g. “austrian two zero one descend five thousand feet qnh one zero two two”. A sequence labeling approach is additionally needed to extract the relevant concepts from the recognized word sequence. In this case, one concept is the callsign “AUA201”. The next is the “DESCEND” command with the value “5'000 feet” and we have the concept “QNH” with the value “1022”. These concepts are combined to two recognized commands, here “AUA201 DESCEND 5000 ft” and “AUA201 QNH 1022”. A common ontology for command transcription is being developed by SESAR 2020 exercise PJ 16-04 [25].

From the three models described above, the focus of this paper is on the CPM and how it influences the results of the ABSR.

B. Command Prediction Model as a Decision Tree

Figure 4 shows the structure of the CPM, which is modelled by a decision tree.
For each command type (e.g. DESCEND, REDUCE, INCREASE, CLEARED_ILS) and flight type (e.g. Arrival, Departure, Overflight) a prediction area is needed as shown in Figure 5. If the “Hypotheses Generator” detects that a lat/long position of an aircraft is inside an area of a specific command type, the command values related to that flight and command type are predicted for that aircraft.

Each symbol in the prediction area (see Figure 5) represents a square of 1 nm by 1 nm. These areas can be created manually [26] or learned automatically from transcribed controller utterances and corresponding recorded radar data. This, however, requires either expert knowledge for manual creation and/or expensive manual transcription work of recorded utterances. In order to remove the need of manual work, our approach tries to learn these areas from automatic transcriptions (task “Area Learning” in Figure 3). For each controller utterance the corresponding lat/long positions are known from the recorded radar data, but the (correct) controller commands, however, are unknown. The only things we know are the recognized commands from the Automatic Speech Recognition in Figure 3.

C. Filtering of Recognized Commands by Checker

If we have a controller utterance like, “sky_travel two five zero nine descend to flight level nine zero”, ABSR should normally recognize the expected command “TVS2509 DESCEND 90 FL”. Afterwards this command could be used, together with the corresponding radar data (which amongst others includes flight plan information) for automatic learning of the command prediction model, but the automatic speech recognition could be wrong. That means that the same controller utterance could result in other commands than the expected one, e.g. in “TVS2509 REDUCE 190” or in “DLH109 DESCEND 90 FL”. Without further filtering this would either result in an entry of 190 in the area of the REDUCE command or in an entry in the DESCEND area for the correct flight level value, but with the wrong lat/long position, given that DLH109 also has radar data at the same time.
To prevent these cases the “Checker” in Figure 3 tries to filter out false recognitions. The challenge for this task is to filter out false recognitions on the one hand, but not to exclude unexpected, but correct transcriptions, on the other hand. To filter out false recognitions, the “Checker” applies the following rules for rejection:

- Commands with unlikely values (e.g. runway that is not available at the airport, values for reduce or descend command to low/high etc.)
- Commands in one transmission that are contrary and usually do not appear together (e.g. turn left and turn right for one aircraft)
- If one of the recognized commands in a transmission is wrong (as described by the rules above) the other commands could be wrong too and will be rejected as well.

D. Command Prediction Model Learning

All command recognitions that are not filtered by the “Checker” are included in the Area Learning task in Figure 3. This task does not only mark the areas in which a command type is given, but it also stores the values that occurred for a command type and counts how often a specific command was given in the 1nm by 1nm areas of the CPM (see Figure 5).

If we take a closer look at Figure 5 we can easily see that Cleared ILS commands only occur inside a small area. Two problems become obvious. The model consists only of a limited amount of training data which also contains false recognitions. On the one hand this results in outliers which are probably the result of false recognitions that the “Checker” did not catch. On the other hand there are small gaps between the learned areas where no Cleared ILS command occurred in the training data, but is very likely in reality. To close the gaps and also expand the border of the learned areas we assumed that a valid command that appears at a certain position in the training data is not only valid for this position but also for the surrounding positions. We do not only mark the respective 1*1 nm area in which a command occurs, but also the surrounding areas. An expansion window size of 13 means that we also mark the 168 neighbors [13*13-1] of a lat/long position. In order to reduce the influence of outliers while enlarging the CPM through expansion windows, we added an additional filter starting at a window size of 13x13 nm. The filter removes every 1nm by 1nm area from the CPM that is even after expansion by the window only marked once. The result of this approach for the Cleared ILS area of Figure 5 is shown in Figure 6.

FIGURE 6 — Prediction area of CPM for Cleared ILS-Command for Arrivals (expansion window 13x13 nm)
E. Iterative Improvement of the Recognition Models

As shown in the bottom yellow shaded part of Figure 3 automatic learning of the predictions areas will result in an im-preved “Command Prediction Model”, which we expect will improve the “Command Hypotheses” iteratively resulting in better “Recognized commands”. The aim of the MALORCA project, however, is to learn/improve also the other ABSR models. The “Checker” in Figure 3 helps also to improve “Acoustic Model” and “Language Model”, because the learning algorithms for acoustic and language model use the feedback from the additional sensor “Radar data” to decide whether an automatic transcription is good or improvable. In this paper we concentrate on CPM improvement without using inputs from an Arrival Manager as described in Figure 2.

IV. Experimental Set-Up for Proof-of-Concept

The set-up to demonstrate that automatic learning of the CPM is possible and how CPM quality improves with the amount of provided learning data is described now. Radar data for Vienna approach was recorded from July to September 2016 for runway configuration 34 for inbounds and runway configuration 29 and 34 for outbounds. Prague approach data was recorded from August to November 2016 for runway configuration 24 and 30 for inbounds and runway configuration 24 for outbounds. Recordings consist of controller communication to pilots and the corresponding radar data and flight plan information.

We considered four controller positions and learned CPMs for each of them:

1. Prague Arrival Executive Controller (AEC),
2. Prague Director Executive Controller (Feeder, PEC),
3. Vienna sector BALAD executive controller (BALAD),
4. Vienna feeder executive controller (Feeder).

The data was split into two parts. The first part was manually transcribed and used for testing. The majority was automatically transcribed based on the speech recognizer software being developed until May 2017 in the MALORCA project. Currently our training data set for Vienna consists of 18.7 hours of clean speech after removing the silence. For Prague approach 18.1 hours are available.

Table 1 shows the total number of commands resulting from automatic transcription by the acoustic and language model from May 2017 resulting in RecR of 89.0% (Prague) and 60.7% (Vienna) without error filtering by the “Checker” in Figure 2. The ErrR is 4.1% (Prague) resp. 10.9% (Vienna). RecR and ErrR will be iteratively improved also by CPM improvement within following MALORCA work packages. Transcribed commands for which automatic transcriptions fails to recognize callsign (output NO_CALLSIGN) or command type (output NO_CONCEPT) are already excluded from Table 1. Correctly transcribed data sets were used to generate the four different command predictions models. We excluded from area learning all suspicious recognitions as described in section III.C.

| TABLE I — Command Number Resulting from Automatic Transcription |
|-----------------|-----------------|-----------------|-----------------|
| Configuration   | # Total Cmds    | # Descend cmds  | # ILS clearances |
| AEC             | 11'103          | 2'184           | 351             |
| PEC             | 5'365           | 920             | 458             |
| BALAD           | 5'929           | 1'062           | 13              |
| Feeder          | 6'959           | 1'100           | 245             |
TABLE II — Size of Test Data Set

<table>
<thead>
<tr>
<th>Approach</th>
<th>Area # Utterances</th>
<th># given</th>
<th># given</th>
<th># sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prague</td>
<td>2'582</td>
<td>4'563</td>
<td>4'580</td>
<td>27</td>
</tr>
<tr>
<td>Vienna</td>
<td>2'427</td>
<td>3'556</td>
<td>3'556</td>
<td>19</td>
</tr>
</tbody>
</table>

For evaluation we used different test utterances which were manually transcribed (see Table 2), i.e. for these utterances the correct transcription (so called gold commands) were known. For each of the 27 resp. 19 controller sessions we calculated different metrics:

- Total number of given commands (#TgC),
- Command recognition rate (RecR): number of correctly recognized commands divided by #TgC,
- Command rejection rate (RejR): number of rejected recognized commands divided by #TgC,
- Command recognition error rate (ErrR): number of recognized commands which were not spoken and not rejected, divided by #TgC,
- Beta: number of rejected recognized commands which were included in gold commands, i.e. they were wrongly rejected, divided by number of total rejections,
- Command prediction error rate (CpER): number of commands included in gold commands, which were not predicted, divided by #TgC.
- Average number of predicted commands (#NPC).

We reject a recognized command if it is not predicted by the learned command prediction model. The sum of the values RecR, RejR and ErrR could be greater than 100% [see 4] for detailed definition of the rates], because sometimes more commands are recognized than given. This at least results in an increase of RejE or ErrR. If more commands are given than recognized, this is always counted as a contribution to RejR.

V. Results

In this chapter we present the results of our experiments and compare them to the baseline. As baseline we choose the set of predicted commands when all inputs from automatic data transcription are ignored, i.e. the set of predicted commands includes all possible commands.

Table 3 shows the baseline results. It might be confusing that command prediction error rate (CpER) is not 0%, if all possible commands are predicted. We, however, use some basic heuristics already in this case to reduce number of pre-dicted commands. The same heuristics are applied when the area based ”Command Prediction Model” is used (Figure 5 and Figure 6):

- Altitude commands are only generated in steps of 1000 respectively 10 for flight levels.
- Inbounds get no climb and increase commands, outbounds no descend and reduce commands. Sometimes provided flight plan information is wrong so that this heuristic fails.
- Aircraft do not get commands after a (recognized) handover command.
- Predicted QNH and ATIS INFORMATION command values (e.g. 1013 or charly) depend on values used in previous commands.
Only IFR flights which are in a defined polygon around the airport are considered (Vienna test radar data on average contains 70 aircraft whereas on average only for 29 aircraft commands are predicted.)

| TABLE III — Metrics if All Commands are Predicted (Baseline) |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                 | #Tg C           | Rec R [%]       | Err R [%]       | Rej R [%]       | Beta [%]        | CpE R [%]       | #NPC            |
| Vienna                          | 3556            | 60.4            | 3.7             | 36.1            | 0.8            | 3.6            | 11505           |
| Prague                          | 4566            | 88.1            | 1.1             | 11.0            | 10.0           | 1.0            | 2054            |

If these hypotheses are not valid a command prediction error is observed. These errors and others are observed as well when area based CPM is used (Figure 5). The only exception is that altitude values which are no multiple of 1000 (e.g. 2700 or 3400 feet) are learned from automatic transcription. To conclude: Area based command prediction will at least include 3.7% (Vienna) resp. 1.1% (Prague) CpE.

In a first approach for automatic learning of the CPM we just used the automatically recognized commands (filtered commands are excluded) and stored the areas in which those commands occurred, according to the aircraft radar data (see chapter III). Table 4 shows the evaluation results of this approach.

| TABLE IV — Metrics For Simple Command Prediction Model |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                 | #Tg C           | Rec R [%]       | Err R [%]       | Rej R [%]       | Beta [%]        | CpE R [%]       | #NPC            |
| Vienna                          | 3556            | 39.4            | 1.4             | 59.4            | 34.5           | 35.2           | 924             |
| Prague                          | 4566            | 70.3            | 0.4             | 29.6            | 63.3           | 21.5           | 350             |

Compared to our baseline the average number of predicted commands is reduced by a factor of 12 resp. 6, but this reduction comes with a price. The reduced set of predicted commands drops the RecR for the Vienna test data to 39.4% (baseline 60.4%) resp. 70.3% (baseline 88.1%) for Prague.

The poor quality of the CPM and the resulting large loss in RecR comes from the limited amount of training data. To improve the quality of the CPM we used the window based approach (see section III.D). We experimented with different expansion windows from 3x3 nm to 29x29 nm. The results of this approach are shown in Table 5 (Vienna) and Table 6 (Prague).

| TABLE V — Vienna - Metrics For window Based CPM |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                 | #TgC            | Rec R [%]       | Err R [%]       | Rej R [%]       | Beta [%]        | CpE R [%]       | #NPC            |
| 3x3                             | 3556            | 51.9            | 1.68            | 46.6            | 17.6           | 16.60           | 1026            |
| 5x5                             | 3556            | 54.8            | 1.75            | 43.6            | 12.4           | 12.30           | 1104            |
| 7x7                             | 3556            | 56.8            | 1.90            | 41.5            | 9.0            | 9.61            | 1170            |
| 9x9                             | 3556            | 57.9            | 1.93            | 40.4            | 6.7            | 8.22            | 1229            |
| 11x11                           | 3556            | 58.4            | 1.96            | 39.8            | 5.6            | 7.40            | 1274            |
| 13x13                           | 3556            | 58.8            | 1.93            | 40.1            | 6.1            | 8.14            | 1215            |
| 15x15                           | 3556            | 58.5            | 1.93            | 39.8            | 5.4            | 7.59            | 1264            |
| 17x17                           | 3556            | 58.6            | 1.96            | 39.6            | 5.0            | 7.24            | 1277            |
| 19x19                           | 3556            | 58.9            | 1.98            | 39.3            | 4.4            | 6.89            | 1304            |
| 21x21                           | 3556            | 59.0            | 1.98            | 39.1            | 4.1            | 6.74            | 1328            |
| 25x25                           | 3556            | 59.1            | 2.01            | 39.1            | 4.0            | 6.44            | 1368            |
| 29x29                           | 3556            | 59.2            | 2.22            | 38.7            | 3.8            | 6.21            | 1400            |

By enlarging the CPM through the window based approach we observe an increase of RecR, but also an increase of ErrR and average number of predicted commands (#NPC). If we put the gain in RecR against the loss in ErrR we can see a benefit up to a window size of 25x25 for Vienna and Prague. At this window size we still have a gain in RecR of 0.1% for Vienna and Prague, but only a loss in ErrR of 0.03% for Vienna resp. 0.00% for Prague.
If we compare the results of the 25x25 window to our baseline the loss in \( \text{RecR} \) is 1.3% (Vienna) resp. 0.4% (Prague) and the improvement in \( \text{ErrR} \) is 1.69% (Vienna) resp. 0.25% (Prague). The difference between baseline and the window based approach is relatively small. The benefit of the window based approach becomes visible, when we take the predicted command set size into account. Compared to the baseline the window based approach delivers fewer predicted commands. For Vienna and Prague we get an average \( \#\text{NPC} \) of 1368 resp. 604. That means a reduction of the predicted command set size by a factor of 8 resp. 3 compared to baseline with all possible commands predicted. The smaller the predicted command set size is, the better the output of the speech recognizer and the better the filtering of the "Checker" will be. The search lattice size (see explanation of Figure 2) exponentially depends on predicted command set size.

The acceptable size of the predicted command set that is applicable depends on the application and how fast it has to update. For an application in which the predicted command set only significantly changes every 60 seconds, a bigger predicted command set is not an issue.

To determine the relevance of training data for a CPM we trained the model with different amounts of data from 10% to 100% of the available training data. We executed those evaluations for all window sizes for Vienna and Prague. The results for \( \text{RecR} \) can be seen in Figure 7 and Figure 8. If we look at the results with only 10%, 20% and 30% of the training data, there is a relatively big increase in recognition rate especially with small window sizes. This difference decreases with larger window sizes, because the larger windows compensate some of the missing data. With more training data the increase in \( \text{RecR} \) gets visibly smaller, but if we look at the difference between 90% and 100% of the data, we still get an increase in \( \text{RecR} \) of 0.14% (Vienna) resp. 0.02% (Prague) with a window size of 29x29. That does not seem like a large improvement, but we have to take into account that the 29x29 window already compensates for a huge amount of not available training data. To conclude: With more training data a small improve in \( \text{RecR} \) is still possible. Also the window size for the CPM could be reduced with more training data, since the need to compensate missing training data would be smaller.

**FIGURE 7 — Dependency of ReCR from training data size (Vienna)**
VI. Discussion of Results and Next Steps

The output of the current version of the acoustic and language model are quite noisy with respect to command recognition (RecR) and command error rate (ErrR). Although we have only RecR of 60.7% resp. 89.0% in the learning data, the learned models for the checker could reduce ErrR by a factor of 5 for both Vienna and Prague approach (from 10.9% to 2.0% for Vienna resp. 4.1% to 0.8% for Prague). Obviously the size of training data also increases the RecR, which is of course limited by the recognition rate without using the "Checker".

Figure 7 and Figure 8 shows a logarithmic dependency of recognition rate from data size (ds). If we assume the relation

$$ds = m \ln(RecR) + b$$

we could expect an increase of RecR for Vienna from 58.6% to 59.4% by increasing recorded speech data size by a factor of 2 (i.e. from 18.7 to 35.4 hours) and of 61.6% if we increase by a factor of 10 (window size 17x17 nm). For Prague we could expect an increase from 87.1% to 88.3% (two times more data) resp. 90.9% (10 times more data). These numbers should just show that more data can improve recognition quality, but the effects will be even smaller, because RecR after rejection could not be better than the RecR without rejection.

If data size is limited (which is always the case) we can also improve recognition rate by increasing the window size which emulates the availability of more data. Increasing window size also increases the error rate, because the areas of outliers and false rejection are enlarged (see Figure 6). Currently we are improving our window model by [1] introducing a dynamic filter for outliers (not always just remove fields which are marked once), [2] varying the effect the expansion window has on surrounding areas (not increase all areas in the window by 1.0) and [3] using expansion windows which are command type dependent. The area for an ILS clearance is much smaller than for a descend clearance. Therefore, smaller windows are more suitable for the ILS clearance. On the other hand the area for a reduce command is comparable to the area size of a descend clearance. We, however, only have 450 commands for learning the reduction area, but 2'150 command for learning the descend area. Therefore, window size should also depend on number of available training data for a command type.

Recognition rates of 60% [Vienna] resp. 88% [Prague] seem to be low compared to 95% reported by the AcListant® project [5], [6]. We should, however, keep in mind, that we only concentrated on the "Checker" component. An improved CPM also improves the ABSR output itself (yellow part in Figure 2) and it will help to improve acoustic and language model improvement which will result in better inputs for CPM learning.
VII. Conclusions

We presented an algorithm which automatically learns a model to predict radar approach controller commands using only radar data, flight plan information and recorded untranscribed controller utterances. Compared to a neural network based approach resulting in a black box model the presented model is based on a decision tree. The command prediction model (CPM) was validated against transcribed controller commands for Vienna and Prague approach for the feeder and sector position, i.e. four command prediction models were learned. In this study the CPM was used for filtering the output of an automatic speech recognizer with low performance (89.0% for Prague and 60.7% for Vienna), i.e. for rejecting wrongly recognized commands. The presented machine learning based algorithm for controller command prediction was successfully validated: Command error rates could be reduced from 4.1% to 0.9% for Prague approach and from 10.9% to 2.0% for Vienna approach.

Compared to AcListant® project with recognition rates of 95% the presented recognition rates seem quite low. This study, however, only uses pure radar and flight plan data and does not use the outputs of an arrival manager, which also contributed to the high recognition rate in the AcListant® project. Furthermore the improved CPM will improve machine learning of acoustic and languages models again resulting in an improved CPM.

Overall, the impact of the solutions of the MALORCA project when integrated into the current ATM procedures is expected to be high, especially due to minimizing the total costs related to the implementation of decision and negotiation support systems and related to the maintenance and system changes towards new ATM procedures.

Acknowledgment

We would like to thank all the controllers who anonymously provided us with real world command examples and also our MALORCA partners from Austro Control and from Air Navigation Service Provider of Czech Republic.

REFERENCES


Meteo and the environment
FUEL SAVING BY GRADUAL CLIMB PROCEDURE

Ryota Mori
Air Traffic Management Department
ENRI
Tokyo, Japan
r-mori@mpat.go.jp

Abstract—This paper proposes a new climb procedure called a gradual climb procedure. When aircraft climb to the cruise altitude or change the flight level, maximum climb thrust is usually applied. However, this maximum cruise thrust leads to a higher fuel consumption compared to the cruise thrust, which results in non fuel optimal higher climb rate. An optimal rate of climb which minimizes the fuel consumption exists, but aircraft usually opt for a faster rate of climb. This paper focuses on the part of the climb phase just prior to the cruise altitude, and clarifies the relationship between fuel consumption and the rate of climb via numerical simulations. The author also proposes a practical fuel-saving climb procedure considering actual air traffic control constraints and pilot operation. The expected fuel saving is of the order of 50 lb, which corresponds to 0.1 % of total cruise fuel consumption. However, the proposed procedure will be applicable to almost all aircraft and flights worldwide, so the cumulative effect will be significant. In addition, the negative effects to air traffic control and a pilot are minor, so the proposed gradual climb procedure can be applied in the near future operation.

Nomenclatures

\( x \): longitudinal distance [m]
\( z \): altitude [m]
\( \nu \): true air speed [m/s]
\( \gamma \): path angle [rad]
\( m \): weight [kg]
\( T \): thrust [N]
\( M \): Mach number
\( L \): lift [N]
\( D \): drag [N]

This product or document has been created by or contains parts which have been created or made available by the European Organization for the Safety of Air Navigation (EUROCONTROL). EUROCONTROL ©2013. All rights reserved. EUROCONTROL shall not be liable for any direct, indirect incidental or consequential damages arising out of or in connection with this product or document, including with respect to the use of BADA4.

\( g \): acceleration of the gravity [m/s\(^2\)]
\( \nu_z \): target rate of climb (vertical speed) [m/s]
I. Introduction

Fuel saving is a keyword in air traffic management, and there are many researches which consider the issue from various aspects. In the ATM field, most researches relate to multiple aircraft to minimize fuel consumption, because aircraft can rarely follow their optimal trajectory due to the aircraft interference of others. For example, optimal operation of multiple aircraft are considered in arrival manager (AMAN) [1][2] and airport surface management[3][4].

On the other hand, CCO (Continuous Climb Operation) and CDO (Continuous Descent Operation) have been proposed for a single flight operation[5][6]. According to these concepts, an aircraft should fly near its optimal profile and stay high as long as possible, minimizing level segment during descent or ascent. However, these operations are also affected by airspace congestion, and are therefore often applied in non-congested times only. The potential benefit of CDO and CCO is significant [7][8][9], so various researches have been conducted on the operation of CDO/CCO in congested airspace [10][11].

The problem of single aircraft trajectory optimization is an old one mainly studied in 1970-80s. Its objective function is set to minimize fuel consumption or flight time, or a mixture of those. The trajectory optimization was described as an optimal control problem, and was solved by maximum principle theoretically [12][13][14]. The computational burden is quite little and easy to implement onboard, so recent aircraft calculate their climb/descent profile using this theory. There are also some recent extended studies, one of which proposes an improvement in calculation with limited information[15].

As for a single trajectory optimization of previous studies, the whole optimal trajectory is categorized into three phases; climb phase, cruise phase, and descent phase. However, the transition between two phases cannot be easily optimized theoretically by the maximum principle. Since the impact of the trajectory optimization is not so big in the transition phase, this transition phase has been overlooked by other researchers and practitioners. However, if the fuel saving is possible and not negligible, it is worth optimizing the climb profile during this transition phase. There are some recent studies calculating the optimal trajectory and obtain the possible fuel saving[16]. However, such a pure “optimal” trajectory often cannot be flown by current aircraft and is therefore difficult to implement the near future.

This paper proposes a new practical climb procedure to reduce fuel burn by changing the climb profile. In the current climb operations, maximum climb thrust (MCT) is usually used (de-rated thrust is also used actually, which will be explained later). The past works[12][13][14][15] also assume the MCT during climb phase. The MCT climb is simple and currently widely in use, but MCT climb is not optimal in terms of fuel consumption. Therefore, this paper clarifies the following points: 1) possible fuel saving by changing the climb profile, 2) proposition of an optimized climb profile for practical use.

This paper is organized as follows: Sec II provides an overview of a typical operation of two climb procedures (climb operation to top of climb (TOC) and step-up climb operation). Sec III explains the simulation model and the optimization method applied in this research. Sec IV shows simulation results for both step-up climb operation and climb to TOC operation under the current operation and the proposed operation in various conditions. Sec V summarizes this paper.

II. Research scope and CLIMB operations

A. Research scope

First of all, the scope of this paper is clarified. As described in the Introduction, this paper proposes a new procedure for the transition phase between climb and cruise, i.e.
climb phase near TOC. On the other hand, the “optimal” trajectory cannot be flown by the current aircraft due to ATC constraints and aircraft capability. This paper also considers these aspects, and proposes a practical way to fly on a sub-optimal trajectory providing a better fuel burn performance than the current one. Regarding the optimization, due to difficulties associated with accurate theoretical description of the transition state, here a numerical analysis is applied.

During the climb, MCT (the maximum thrust) is widely used in the current operation. Therefore, the baseline scenario is assumed that the aircraft climbs to a cruise altitude with MCT. During the cruise, it is well-known that the optimal cruise altitude gradually increases as the aircraft weight decreases[13]. However, the rate of climb (ROC) becomes very small to track the optimal trajectory, so the aircraft cannot fly on the optimal trajectory due to ATC requirements. Instead, the aircraft often applies a step climb. Although the optimal altitude changes gradually, ATC usually assigns altitude to each aircraft every 1000 ft, so the step climb allows the aircraft to fly on sub-optimal altitude by changing the altitude by 1000 ft. The optimal timing of the step climb exists, and some researchers consider the optimal step climb points including wind effects [17][18], which are also a well-discussed problem. However, ROC during step climb has not been discussed by other researchers. This step climb can also be handled as the transition from climb to cruise, and there is room for improvement by changing the climb profile during step climb. Therefore, this paper considers two cases: climb profile near TOC and climb profile during step climb. The flight trajectories at low altitude climb and the cruise phase flight are assumed to be the same as the current operation, and the proposed climb profile does not affect other phases.

B. Current aircraft capability and possible sub-optimal climb profile

This subsection considers the aircraft operational aspects. The current aircraft may not be able to follow the optimal flight profile, because the aircraft should select a certain FMC (Flight Management Computer) mode. FMC can provide the target path and speed profile by considering various aircraft information such as route structure and aircraft weight. The target path and speed profile is calculated by the maximum principle explained in the previous section. Since the lateral route is usually decided by a flight chart, longitudinal and vertical motion are optimized by FMC. Here, speed and ROC are controlled by two control devices (pitch angle and engine thrust). Each control variable (speed and ROC) can be controlled by either control device, so there are two options. The first option is to keep the target speed by controlling the pitch angle and setting the thrust constant (usually MCT during climb). ROC is automatically determined by the thrust setting. The second option is to keep the target ROC by controlling the pitch angle and using the engine thrust to control speed. The first option is usually used during climb operation as VNAV SPD mode (Boeing) or flight level change (FLCH/LVL CHG) mode. During the climb, the thrust is set to MCT, so only the target speed profile is required. The second option can be flown using VNAV PATH mode (Boeing) or vertical speed (V/S) mode. Since both speed and ROC are controlled, both the target speed profile and the target ROC profile are required. Both options are summarized in Table 1.

<table>
<thead>
<tr>
<th>TABLE 1 — Two FMC mode options during climb.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Option 1</td>
</tr>
<tr>
<td>Pitch angle</td>
</tr>
<tr>
<td>Thrust</td>
</tr>
<tr>
<td>Target speed profile</td>
</tr>
<tr>
<td>Target ROC profile</td>
</tr>
<tr>
<td>Option 2</td>
</tr>
<tr>
<td>Pitch angle</td>
</tr>
<tr>
<td>Thrust</td>
</tr>
<tr>
<td>Target speed profile</td>
</tr>
<tr>
<td>Target ROC profile</td>
</tr>
</tbody>
</table>

In reality, Option 2 is rarely used during the climb, so only the speed profile is given during this stage. In order to change the climb profile, it is recommended to use Option 2 but the target ROC should be given. If the pilot wants to use Option 2 during the climb, it can only be done with a manual target ROC setting.
C. Current climb operation

This research focuses on two different climb operations: the first operation is the climb operation to TOC, and the second operation is a step-up climb operation.

First, the climb operation to TOC is considered. When an aircraft climbs to TOC, it usually uses VNAV SPD mode (Boeing), i.e. Option 1. FMC calculates the optimal speed during climb. During the climb, the thrust is set to MCT, so the aircraft climbs at the maximum possible ROC while maintaining its target speed.

To extend the engine cycle, the derated thrust setting is often used. The derated thrust sets the climb thrust smaller than the MCT by up to 30%, which depends on the aircraft type. During the ascent, if the ROC is set smaller than the optimal one, the aircraft consumes additional fuel. However, the engine life span can be extended, and such a benefit is usually larger than the additional fuel consumption. However, the rate of derated thrust becomes smaller with higher altitude, and most aircraft do not activate the derated thrust above 30000 ft altitude, which means that MCT is applied when climbing above 30000 ft. This research considers the flight above 30000 ft only, so the derated thrust is not considered.

As describe above, the aircraft uses the MCT (or derated thrust) to reach TOC, but the engine is designed to be most efficient at a certain cruising thrust, not at MCT. When an aircraft climbs, thrust larger than cruising thrust is used, but the fuel consumption can be theoretically reduced by setting the thrust near the cruising thrust during the climb, so ROC decreases. On the other hand, there is an optimal cruising altitude once the aircraft type and weight are determined. Too small a ROC means that the aircraft has to fly on the non-optimal altitude, which consumes additional fuel. Therefore, in regard of these two factors, it is safe to say that there exists a fuel-optimal ROC for each altitude.

Second, the step-up climb is considered. The optimal flight altitude depends on the aircraft weight. With time the aircraft becomes lighter due to the burnt fuel and so the optimal flight altitude becomes higher as the cruise proceeds. At a certain timing, the pilot can place a request to ATC to fly on a higher altitude. If clearance is obtained, the pilot usually sets the designated altitude in MCP and pushes the altitude nob, then the aircraft starts climbing with MCT. For reasons analogous to the ones stated earlier in the climb to TOC explanations, there should be an optimal ROC.

D. Possible sub-optimal climb profile

To account for the implementation of the proposed climb procedure in the real world, the operational aspects should be considered. From ATC perspective, the aircraft should fly on a certain cruise altitude and should not apply too low ROC. From aircraft control perspective, the target ROC should be set, i.e. the climb by Option 2.

First, the comparison of the various climb profiles is shown in Fig. 1 under a certain environment (a certain aircraft type and a certain weight). This is just an example, and the calculation method will be shown in the next section. If MCT is applied, the aircraft can climb with the highest possible ROC. On the other hand, the optimal trajectory shows a fast climb at the beginning, but as the ROC gradually reduces, a very low ROC (optimal cruise ROC) is observed. However, as mentioned before, such a ROC is not acceptable from ATC perspective. The proposed climb profile is within the allowable window defined by these two profiles. First the aircraft climbs with MCT, but at a certain point (33000 ft in this example) small ROC (1000 ft/min in this example) is set. After reaching the cruise altitude, the aircraft maintains this altitude. By setting a small ROC at 33000 ft, TOC is shifted by about 10 NM compared to MCT climb. The proposed profile is possible for the current aircraft, and will be acceptable for ATC. (The details of the possible negative impacts will be discussed in Sec. IV (D).) The questions are how much fuel can be saved by the proposed procedure, and what ROC should be set during the climb. These questions are investigated in the following sections.
III. Simulation and optimization

A. Simulation model

To investigate the effect of ROC on fuel consumption, numerical simulations are conducted. Here, the details of the simulation model are described. It is assumed that an aircraft flies on a straight track, and only the longitudinal and vertical dynamics are considered. The aircraft dynamics are given by the following point mass model with five state variables \( \mathbf{x} = [x, z, \nu, \gamma, m]^T \) and two control variables \( \mathbf{u} = [\dot{\gamma}, T_{\text{ratio}}]^T \). Here, null wind is assumed.

\[
\begin{align*}
\dot{x} &= \nu \cos \gamma \\
\dot{z} &= \nu \sin \gamma \\
\dot{\nu} &= \frac{T - D}{m} - g \sin \gamma \\
\dot{\gamma} &= \frac{L}{m \nu} - \frac{g \cos \gamma}{\nu} \\
\dot{m} &= f(M, T)
\end{align*}
\]

where \( f(M, T) \) is the fuel consumption. Since \( \dot{\gamma} \) is a control parameter, \( L \) is automatically determined. \( D, T_{\text{max}}, T_{\text{min}}, \) and \( f(M, T) \) are calculated by the Base of Aircraft Data (BADA) model\[19\]. Here, two aircraft types (B777-300 and A330-300) are used in the simulation. The operational constraints are also set based on BADA model.

B. Optimization method

To determine the optimal flight path, a trajectory optimization problem is formulated. Here, two operations are assumed: 1) climb with MCT, 2) climb at constant ROC. To account for both operations, a three-stage optimization problem is formulated. For the step-up climb operation, the first stage is for a level flight before the step-up climb, and the second stage is for a climbing operation, and third stage is for a level flight after the step-up climb. The following initial and terminal conditions and constraints are set in each stage.

\[
T = T_{\text{ratio}} \left( T_{\text{max}} - T_{\text{min}} \right) + T_{\text{min}} T_{\text{ratio}} = [0 \ 1]
\]
1st stage constraints

\[ x_0^i = 0 \]
\[ z_0^i = \text{initial altitude} \ (z_0) \]
\[ \nu_0^i = \text{initial TAS calculated from initial Mach} \]
\[ \gamma_0^i = 0 \]
\[ m_0^i = \text{initial weight} \]
\[ z_1^i = z_0 \]
\[ \dot{\gamma}_1^i = 0 \]

2nd stage constraints

\[ z_0^2 = z_0 \]
\[ z_2^2 = \text{final altitude} \ (z_f) \]
\[ t_2^f - t_0^f = \frac{z_f - z_0}{v_z} + \alpha \] \ ... [a]
\[ t_2^f - t_0^f \leq \frac{z_f - z_0}{v_z} \]
\[ t_{ratio}^2 = 1 \] \ ... (b)

Mach number is between cruise Mach number \(-0.005\) and cruise Mach number \(+0.005\).

3rd stage constraints

\[ z_0^3 = z_f \]
\[ \dot{\gamma}_3^3 = 0 \]
\[ x_3^f = \text{final distance} \]
\[ z_0^3 = z_f \]
\[ \nu_0^f = \text{final TAS calculated from final Mach} \]

where the superscript indicates the stage number, the subscript 0 indicates the initial condition in each stage, and the subscript \( f \) indicates the final condition in each stage. Either constraint [a] or [b] in 2nd stage are used depending on the problem, i.e. climb with MCT or constant ROC. The constraints [a] are used for 2) climb with constant ROC, and the constraint [b] is used for 1) climb with MCT. \( \alpha \) is a parameter to account for the aircraft movement. If \( \alpha \) is set to 0, the aircraft has to climb at the maximum ROC \((\nu_z)\) during the 2nd stage. However, at the beginning and the end of the stage, the ROC should be 0, so the solution becomes infeasible. Therefore, \( \alpha \) should be as low as possible if the solution is feasible, and here it is set by trial and error.

The objective function is usually set to include both the fuel consumption and flight time, and its weight is given as a cost index \([CI]\). Since \( CI \) is given in the unit of 100 lb/hour [Boeing], the following objective function to be minimized is set.

\[
J = \frac{100}{3600} C I n^3 \int_0^f -\dot{r} dt + 0.453592 \int_0^f \nu z dt
\]

The unit of \( J \) is lb. 1 s flight time corresponds to 0.0278 CI lb. Therefore, if \( CI \) is set to 100, 1 s flight corresponds to 2.78 lb fuel consumption. To solve the optimization problem, the pseudospectral discretization method is applied. Using this method, the continuous trajectory optimization problem can be formulated as a nonlinear programming (NLP) problem. In this study, IPOPT is used as a NLP solver, and PSOpt (optimal control solver software) is used for software implementation [20]. The nodes in each stage are also set by trial and error, with 10-30 nodes being set in each stage.
IV. Simulation Results

A. Optimal flight profile of the step-up climb procedure

To account for a step-up climb procedure, three initial and terminal conditions are assumed as shown in Table 2. In all scenarios, the aircraft changes altitude by 2000 ft. Once the initial weight is determined, the optimal vertical flight profile is determined, and the appropriate initial and terminal altitudes are set in each scenario.

Figs. 2 and 3 show the optimal flight profile in scenario 1 and 3. (Neither constraints (a) nor (b) are applied.) Both figures show a similar trend; the mach number is almost constant during the simulation period, and the aircraft gradually climbs from the initial altitude to the terminal altitude. The same trend is observed in scenario 2 as well. Since the optimal altitude gradually changes theoretically, these results show that the calculation is feasible. In all cases, the ROC during the climb is about 10 ft/min. This ROC is too low and this operation is never performed in practice. Instead, step climb is often used. During a step climb, MCT is usually applied. Therefore, the optimal step climb profile is obtained via optimization calculation. In the step climb calculation, the constraint (b) is applied. Fig. 4 shows the optimal step climb flight profile with MCT in scenario 1. The altitude change happens around 1000 NM point, which agree with the optimal flight profile shown in Fig. 2, because the aircraft passes around 1000 NM point at 37000 ft. Since MCT is applied, about 1500 ft/min ROC is observed. The interesting point is the difference of the objective function. The difference of the objective function between the optimal profile (Fig. 2) and the optimal step climb profile with MCT (Fig. 4) is 89.3 lb. The objective function consists of fuel consumption and flight time, but the difference of flight time is less than 1 s and therefore negligible. The difference of the objective function is almost equal to the difference of the fuel consumption. The same trend is observed in all scenarios (1-3).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial weight [lb]</td>
<td>540,000</td>
<td>610,000</td>
<td>440,000</td>
</tr>
<tr>
<td>Initial altitude [ft]</td>
<td>36,000</td>
<td>33,000</td>
<td>37,000</td>
</tr>
<tr>
<td>Terminal altitude [ft]</td>
<td>38,000</td>
<td>35,000</td>
<td>39,000</td>
</tr>
<tr>
<td>Initial/Terminal Mach</td>
<td>0.83</td>
<td>0.83</td>
<td>0.80</td>
</tr>
<tr>
<td>Flight distance [NM]</td>
<td>2,000</td>
<td>2,000</td>
<td>2,000</td>
</tr>
<tr>
<td>Aircraft type</td>
<td>B773</td>
<td>B773</td>
<td>A333</td>
</tr>
<tr>
<td>Cl</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

FIGURE 2 — Optimal flight profile in scenario 1. (J = 103,530.5lb, fuel used = 61,687.2lb, flight time = 15,063.6s)
FIGURE 3 — Optimal flight profile in scenario 3 (J = 94,523.3lb, fuel used = 51,334.6lb, flight time = 15,547.9s)

FIGURE 4 — Flight profile in scenario 1 with the maximum climb thrust. (J = 103,618.0lb, fuel used = 61778.4lb, flight time = 15,062.2s)

However, as mentioned before, climb at 10 ft/min is not realistic, so higher ROC is required in practice. Even basic FMCs provide a climb function at a fixed ROC (V/S mode) with a minimum value of 50 ft/min, and it can be set every 50 ft/min up to 1000 ft/min. This time, 50 ft/min, 500 ft/min, 1000 ft/min are chosen for the calculation. The values of objective function are summarized in each scenario and ROC, and shown in Table 3. The climb profile in each ROC in scenario 1 is shown in Fig. 5.

TABLE III — Fuel consumption in each ROC and each scenario.

<table>
<thead>
<tr>
<th>ROC [ft/min]</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>103530.5</td>
<td>111172.3</td>
<td>94523.3</td>
</tr>
<tr>
<td>50</td>
<td>103563.1</td>
<td>111208.4</td>
<td>94553.2</td>
</tr>
<tr>
<td>500</td>
<td>103590.5</td>
<td>111235.9</td>
<td>94580.3</td>
</tr>
<tr>
<td>1000</td>
<td>103604.5</td>
<td>111248.4</td>
<td>94593.0</td>
</tr>
<tr>
<td>MCT</td>
<td>103618.0</td>
<td>111265.0</td>
<td>94594.0</td>
</tr>
</tbody>
</table>
As for the climb profile as shown in Fig. 5, MCT climb achieves the highest ROC while the optimal climb profile shows the smallest ROC. All climb profiles cross at a specific point, which indicates the feasibility of the calculation. Note that the climb profiles between 500 ft/min climb and MCT climb do not differ greatly. MCT climb completes the climb for about 10 NM and 80 s, while 500 ft/min climb completes the climb for 30 NM and 240 s.

As seen in Table 3, less fuel is used with slower ROC, and the overall trend is similar for all scenarios. By using 50 ft/min climb, 40-60 lb fuel can be saved compared to MCT step climb. However, if 50 ft/min climb is applied, it takes 40 minutes to complete 2000 ft altitude change, and all altitude ranges (in this case three flight levels) have to be blocked. Such multiple altitude blocking is possible in the current ATC operation, but of course it reduces the capacity of airspace. However, it might be worth trying 50 ft/min climb if the airspace is not crowded. On the other hand, 13-30 lb fuel can be saved with 500 ft/min step climb compared to MCT step climb. According to ATC controllers, 500 ft/min climb is not slow, because the climb performance of some aircraft is less than 1000 ft/min even with MCT. 500 ft/min climb might be more realistic for implementation.

The possible fuel saving can differ depending on the flight conditions, such as cruise mach number, aircraft type, wind condition, weight, Cl, and so on. On the other hand, the possible fuel saving is 15 – 50 lb which corresponds to about 0.1 % of total fuel consumption for 2000 NM flight. Although it might seem quite little and even negligible, the proposed gradual climb procedure is applicable to almost all aircraft flying worldwide. Its cumulative effect will be significant. Under specific conditions, gradual climb might save no fuel at all, so the calculation in various conditions will be a subject of future work.

**B. Optimal flight profile for climb to TOC operation**

Next, the climb to TOC operation is considered. Here, two initial conditions (Scenario 4 and 5) are considered based on scenario 1 as shown in Table 4. Since it is already known that the optimal altitude under the initial weight in scenario 1 is a bit less than 36000 ft, the optimal profile is calculated up to 36000 ft under the same initial weight. For the comparison purpose, in scenario 5, it is assumed that the aircraft is not cleared to fly at optimal altitude (36000 ft) due to airspace congestion and is allowed to fly at 34000 ft.

**TABLE IV — Initial and terminal conditions for climb to TOC procedure.**

<table>
<thead>
<tr>
<th>Scenario 4</th>
<th>Scenario 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial weight [lb]</td>
<td>540,000</td>
</tr>
<tr>
<td>Initial altitude [ft]</td>
<td>30,000</td>
</tr>
<tr>
<td>Terminal altitude [ft]</td>
<td>36,000</td>
</tr>
<tr>
<td>Initial climb angle [deg]</td>
<td>1.0</td>
</tr>
<tr>
<td>Initial/Terminal Mach</td>
<td>0.83</td>
</tr>
<tr>
<td>Flight distance [NM]</td>
<td>500</td>
</tr>
<tr>
<td>Aircraft Type</td>
<td>B773</td>
</tr>
<tr>
<td>Cl</td>
<td>100</td>
</tr>
</tbody>
</table>
Fig. 6 shows the optimal profile and the profile with MCT in scenario 4. When MCT is used, the aircraft climbs to 36000 ft with the highest possible ROC and flies level. On the other hand, in the optimal profile, the aircraft climbs fast at the beginning, but the ROC gets lower with climb. After passing 35000 ft, the ROC becomes small. Fig. 7 shows the relationship between altitude and the ROC under MCT and optimal profile. The initial climb angle of 1.0 deg corresponds to about 800 ft/min ROC. With MCT, about 2000 ft/min ROC is achieved during the entire altitude range. However, in the optimal profile, the optimal ROC decreases with altitude almost linearly. If a small ROC is applied near TOC, the fuel burn can be reduced.

**FIGURE 6 — Optimal flight profile and the flight profile with MCT in scenario 4.**

![Optimal flight profile and the flight profile with MCT in scenario 4.](image)

**FIGURE 7 — Relationship between ROC and altitude in optimal flight profile in scenario 4.**

![Relationship between ROC and altitude in optimal flight profile in scenario 4.](image)

In order to save fuel to reach TOC in real world implementation, the following is a possible operation: climb to a certain altitude (defined as transfer altitude) with MCT and maintain a certain ROC between the transfer altitude and the cruise altitude. Fig. 1 in Sec II shows an example of flight trajectory when transfer altitude is 33000 ft and ROC is 1000 ft/min as well as the MCT climb profile and optimal climb profile. As shown in the figure, the aircraft starts climbing at 30000 ft and climbs to 33000 ft with MCT (about 2000 ft/min). After passing 33000 ft, the ROC is changed to 1000 ft/min and climbs to 36000 ft (TOC).

Since the proposed climb procedure requires two parameters (transfer altitude and ROC), Fig. 8 shows the calculated fuel saving by the proposed method compared to MCT climb in Scenario 4. First, the difference of fuel consumption between MCT and optimal climb is 64 lb, which is the maximum possible fuel saving. If 500 ft/min ROC is applied, the best transfer altitude is 32000 ft, and 42 lb fuel is saved compared to MCT climb. As for 1000 ft/min ROC case, 41 lb fuel saving is achieved when the transfer altitude is set to 30,000 ft. If an appropriate transfer altitude is chosen, the selection of ROC does not cause a big difference in this case.
Next, the same calculation is conducted for Scenario 5. The optimal profile is calculated and shown in Fig. 9. At the beginning, the optimal profile between Scenario 4 and 5 are the same, but the optimal trajectory of Scenario 5 smoothly goes below the optimal trajectory of Scenario 4. In Scenario 5, the aircraft has to fly below the optimal cruise altitude (around 35500 ft), so the optimal trajectory is to cruise the highest cleared altitude. In scenario 5, the fuel saving of the proposed procedure is calculated with various parameters as shown in Fig. 10. Compared to Fig. 8 (Scenario 4), the maximum altitude is constrained to 34000 ft, so the possible fuel saving by the optimal climb profile is reduced to 34 lb. However, using the proposed climb procedure, about 28 lb fuel saving is possible by choosing appropriate ROC and the transfer altitude. Note that 1000 ft/min ROC is overall better than 500 ft/min ROC in Scenario 5, while it is opposite in Scenario 4. This is due to the difference of the optimal profile between Scenario 4 and 5, but we have to find the best transfer altitude and ROC easily for the real implementation. These will be functions of the various factors such as aircraft weight, cleared altitude, wind and temperature. Further details will be examined and analyzed in a future work.

**FIGURE 8 — Objective function reduction by the proposed climb with various parameters in Scenario 4.**

**FIGURE 9 — Optimal trajectory in Scenario 4 and 5.**

**FIGURE 10 — Objective function reduction by the proposed climb with various parameters in Scenario 5.**
C. Possible negative effects and feedbacks from pilots and air traffic controllers

The introduction of the proposed gradual climb procedure might cause negative impacts to the aircraft operation. Therefore, the author discussed the proposed procedure with several pilots and air traffic controllers and obtained their valuable feedback.

According to the pilots, gradual climb operation should save fuel to some extent. The optimal ROC varies with the conditions, but it is preferable to have a simple rule, such as 500 ft/min for step-up climb, or 1000 ft/min for the last 2000 ft prior to TOC. In addition, TCAS monitors climb or descent when the climb/descent rate is 500 ft/min or larger, so if the ROC is too small, other aircraft might think that the aircraft is not climbing/descending. Therefore, the ROC for 500 ft/min or greater is recommended for situational awareness. To conduct the proposed gradual climb procedure, the pilot should select V/S mode by pushing the V/S button and set an appropriate ROC. After reaching the cruise altitude, the aircraft automatically starts cruise flight and V/S mode is automatically changed. Therefore, the impact of the gradual climb procedure to the pilot workload will be limited.

As for ATC perspective, it takes longer time to reach the cruise altitude by using a gradual climb procedure. However, during the normal climb, ATC does not assign ROC of the aircraft and does not know how long it will take to reach the cruise altitude. Therefore, the aircraft are sufficiently separated horizontally from each other during climb, so no safety issue will be occurred by a gradual climb procedure. As for the ATC efficiency, ATC does not feel that 500 ft/min climb rate is slow. Since sufficient horizontal separation is set, no impact will be given to another aircraft. Also, even if the pilot applies gradual climb procedure, no report to ATC is required. However, if the aircraft conducts a 50 ft/min climb, multiple flight levels should be blocked, which might affect other flights in the vicinity.

According to these comments, the negative impacts will be almost negligible by operating gradual climb procedure. Even if the fuel saving per flight by gradual climb is not big, little negative effect is expected, so it is worth performing the gradual climb procedure.

V. Conclusions and Future Works

This research considered a practical way to implement gradual climb which has theoretically been known to save fuel. Potential fuel saving was calculated considering the current ATC and pilot operation. It would be impossible to fly on the “optimal profile”, but this research showed that a sub-optimal profile such as a fixed ROC could achieve fuel saving. The possible fuel saving per flight is not significant, the order of 10-100 lb. However, the proposed gradual climb procedure is applicable for all commercial aircraft flying worldwide, and so the cumulative effect will be significant. Clarification of the conditions under which the proposed operation can be applied and implemented in practice will be a subject to future work.

Acknowledgment

I would like to thank Mr. Tsuneharu Uemura for his valuable comments from the pilot perspective.
REFERENCES


GROUND-BASED WIND FIELD CONSTRUCTION FROM MODE-S AND ADS-B DATA WITH A NOVEL GAS PARTICLE MODEL

Junzi Sun, Huy Vû, Joost Ellerbroek, Jacco Hoekstra
Control and Simulation, Faculty of Aerospace Engineering Delft University of Technology, the Netherlands

Abstract—Wind is an important parameter in many air traffic management researches, as it often introduces significant uncertainties in aircraft performance studies and trajectory predictions. Obtaining accurate wind field information has always been a challenge due to the availability of weather sensors. Traditionally, there is no direct method to measure wind data at different altitudes with the exception of weather balloon systems that cannot be easily scaled. On the other hand, aircraft, which rely heavily on atmospheric data, can be part of atmospheric model itself. Aircraft can provide wind and temperature measurements to ground observers. In this paper, aircraft are considered as a moving sensor network established to re-construct the wind field on a larger scale. Based on the powerful open-source tool pyModeS, aircraft ground velocity and airspeed are decoded from ADS-B and Mode-S data respectively. Wind observations are then derived based on the difference of these two vectors. An innovative gas particle model is also developed so that the complete wind field can be constructed continuously based on these observations. The model can generate wind field in real-time and at all flight levels. Furthermore, the confidence of wind at any 4D position can be computed according to the proposed model method. Multiple self-and cross-validations are conducted to ensure the correctness and stability of the model, as well as the resulting wind field. This paper provides a series of novel methods, as well as open-source tools, that enable the research community using simple ADS-B/Mode-S receivers to construct accurate wind fields.

I. Introduction

Using airplanes as weather sensors is a relatively new field in ATM and meteorological research. Traditionally, aircraft obtain weather updates from air traffic services to optimize their trajectory and speed, to best adapt to wind conditions, and to avoid areas of drastic weather conditions. These meteorological updates are mostly coming from ground-based operations, for example, radar surveillance, observation stations, or forecast systems [1], [2]. On the other hand, while an aircraft flies through an airspace, its local meteorological conditions can also be computed. Existing systems such as Aircraft Meteorological Data Relay (AMDAR) [3] and Meteorological Routine Air Report (MRAR) allow aircraft to down-link these meteorological data either through ACARS and Mode-S respectively. Air traffic controllers can combine these data with external sources to make better predictions of weather, which can then be updated and relayed to other aircraft.

Both AMDAR are MRAR are unencrypted broadcast data, which means that anyone can set up receivers to intercept these data. However, as part of ACARS, the legality of
intercepting AMDAR is questionable in certain countries. As for MRAR, the amount of aircraft that broadcast this information is extremely limited. Most aircraft transponders choose not to enable this capacity and is also not interrogated by many ATC. Also, like other Mode-S data, the decoding for MRAR can be troublesome for the research community due to the closed design of downlink data structure.

Several previous studies have been proposed to use flight data from various sources to estimate wind conditions at the location of aircraft. They can be generally divided in three stages or categories:

1. **Estimation of wind from ground based trajectory observation**

   This concept assumes a quasi-constant wind velocity and aircraft airspeed during a turning maneuver. The wind velocity vector can then be estimated dynamically using multiple continuous observations of aircraft ground speed in combination with Bayesian filtering. Early in 1989, Hollister et al. [4] first proposed this method. Later on, Delahaye and Puechmorel, applied variations and extensions of such methodology [5], [6], [7]. Now that ADS-B transponders have been widely deployed, monitoring aircraft states through ADSB has become a possiblity. De Leege et al. and De Jong et al. were the first to introduced the use of ADS-B data to solve this problem [8], [9].

2. **Estimation of aircraft local wind from Mode-S data**

   Mode-S provides many additional aircraft states to ground control, complimenting radar or ADS-B data. A series of studies conducted by the Dutch Meteorological Institute presented by De Haan et al. constructed wind from Mode-S and MRAR data [10], [11]. Hrastovec and Solina also implemented a similar experimental method to achieve the same goal [12]. In addition to the previously mentioned direct wind information in MRAR, airspeed of aircraft are down-linked. The wind can be computed as the difference between aircraft airspeed and ground speed. De Haan also used such a method to calculate local wind data of aircraft [13].

3. **Wind field estimation based on multiple wind measurements**

   While most of the above researches focused on deriving the local meteorological conditions of an aircraft, several of these studies also extended their methods to larger wind field or multiple aircraft scenario. For example, Hollister proposed the Hidden Markov model to update a wind grid based on measurement from multiple aircraft. Delahaye and De Leeg used non-linear Kalman filters on either radar or ADS-B data to obtain wind field. In additional to previously mention literatures, other methods also exist. For example, Hurter et al. used the least-squared method to construct wind field from multiple aircraft measurements [14], while Kapoor et al. implemented machine learning based on the Gaussian Process to predict and extrapolate wind field [15].

Based on the existing literature, researchers are also able to estimate wind vector using aircraft data. However, there are still areas missing in terms of constructing wind field based on measurements from aircraft. Unlike other types of direct sensor networks, using aircraft themselves as wind sensors has several disadvantages:

- Airplanes are moving objects. Therefore, the measurements derived from air traffic data have both temporal and spatial continuity and variance.
- As aircraft tend to fly along a predefined path, most measurements are concentrated along these flight paths. Except climbing or descending, aircraft also tend to fly at fixed cruise levels. This creates a high concentration of measurements along flight routes, alongside rare or no measurement in other spaces.
- The chaotic and temporal nature of wind makes the model highly non-persistent.
Decoding airspeed from EHS entails a certain level of ambiguity and, thus, leads to errors in individual wind calculation.

The focus of this paper is to investigate a novel and relatively fast gas particle model that estimates real-time wind field from observations of aircraft ground speed and airspeed gathered by an ADS-B/Mode-S receiver. The gas particle model can be used to estimate states of wind field and address the challenges caused by the dynamic and not evenly distributed observations. Tunable model parameters can be used to produce confidence levels of wind field and to configure model persistence.

The crucial part of decoding ADS-B and Mode-S data are open-sourced by the authors of this paper. And the process of generating wind observations based on these data will be addressed. Within the framework of this paper, wind observations are first computed using the difference between ground speed (from ADS-B messages) and airspeed (derived from Mode-S messages) using open-source software pyModeS [16]. Then the observations are used by the particle model to construct the wind field and compute wind vector at any positions with confidence indicators. The results are first validated with external wind data source (i.e.: Global Forecast System datasets) for an extended period of time. Finally, the model is self-validated to examine variances and stabilities.

The remainder of the paper is structured as follows. Section two describes the process of obtaining wind observations. Section three presents the essentials of the gas particle model with simple examples. Section four details the large number of the experiments and validations conducted based on the model. Finally, the discussion and conclusions are presented in sections five and six.

II. The wind observations

A simple ADS-B/Mode-S receiver is installed at the faculty of Aerospace Engineering at the Delft University of Technology. This device provides a constant stream of signals obtained from aircraft. Using open-source decoding library pyModeS, the ADS-B and Mode-S data that are collected can be used to derive wind observations for the particle model.

A. Processing of Mode-S data

Through Mode-S, different aircraft state information is downlinked to ground receivers. This information contains aircraft positions, velocities, operational parameters, and meteorological data, etc. The Mode-S transponder can maintain 256 different 56-bit wide Binary Data Store registers (BDS) that can be interrogated by ATC. These registers are indicated by two-digit hexadecimal numbers that can be requested via 25 different downlink formats (DF). Information in these registers are updated with a minimum interval specified by ICAO.

Among all these downlink formats, ADS-B is transmitted via DF17 (or DF18), while Mode-S EHS/MRAR is transmitted via DF20 and DF21. Decoding of ADS-B messages is well documented. However, the decoding of Mode-S data is much more challenging. The challenges include determining the source of aircraft (ICAO address), the content of message (from BDS code), and the quality of the content (certainty of the values).

Aircraft ICAO addresses can be determined by the reverse parity check of the Mode-S message (DF04, DF05, DF20, and DF21). Correct ICAO addresses can only be obtained when a signal is not corrupt. If a message is corrupt (e.g. one or more bits are flipped), it will result in an incorrect ICAO address. However, by cross-referencing resulting ICAO addresses with ADS-B streaming, error messages can be discovered.
The second challenge is that, unlike ADS-B, Mode-S messages do not contain any information on their message types (i.e. BDS codes). This is because only the interrogating ATC knows the target aircraft and what to expected in the downlink message. Such a lack of transparency in Mode-S design poses the biggest challenge in making use of these open data. With the latest version of pyModeS, much of this data is finally unveiled.

Briefly, in pyModeS, the BDS code is determined by checking several signification status bits and evaluating possible values contained in the messages. A status bit indicates whether its related register field (aircraft parameter) is available in the message. This is implemented as follows: When a status bit is set to zero, all related content bits should be zero as well. Messages with different BDS codes usually have different signification status bits. Thus, multiple checks assuming different message types need to be performed to evaluate all possible types or a combination of types. It may occur that a message matches multiple BDS codes. In this paper, only uniquely identified messages are kept and used for the propose particle model.

The last challenge is the quality of the content. Values decoded from corrected messages may be incorrect due to aircraft measurements or transmission errors. For now, no additional filtering is applied in order to provide the direct computation of wind observations. A good design of wind field model needs to cope with this uncertainty, in addition to the errors from incorrect BDS identifications.

B. Compute wind vectors

Figure 1 shows the relationship between true airspeed, ground speed, and wind. The ground speed vector $V_g$ given by Equation 1 is the sum of the true airspeed vector $V_a$ and wind vector $W$. $\chi_g$, $\chi_a$, and $\chi_w$ are the ground speed vector angle (track angle), airspeed vector angle (heading) and wind vector angle with respect to the true north respectively.

$$ V_g = V_a + W \quad (1) $$

To simplify calculations in equation 1, the vectors $V_x$, $V_y$, and $W$ are decomposed into a west-east component $V_x$ and a south-north component $V_y$. The decompositions are calculated using Equation 2.

$$ V_x = V \sin(\chi) \quad (2) $$

$$ V_y = V \cos(\chi) $$

Wind components can then be calculated with Equation 3.

$$ W_x = V_g - V_a $$

$$ W_y = V_g - V_a $$
To calculate the wind, the airspeed and ground speed should first be determined. This can be done using ADS-B and Mode-S data. In Table I, the relevant variables from ADS-B and Mode-S are given.

<table>
<thead>
<tr>
<th>ADS-B</th>
<th>BDS 5.0</th>
<th>BDS 6.0</th>
<th>MRAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track angle</td>
<td>Track angle</td>
<td>Magnetic heading</td>
<td>Wind speed</td>
</tr>
<tr>
<td>Ground speed</td>
<td>Ground speed</td>
<td>Indicated airspeed</td>
<td>Wind direction</td>
</tr>
<tr>
<td>Position</td>
<td>True airspeed</td>
<td>Mach number</td>
<td>Air temperature</td>
</tr>
<tr>
<td>Pressure alt.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geometric alt.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The wind vector can then be determined as follows:

- Other than the ground speed vectors from ADS-B directly, the ground speed, true airspeed, and track angle are also available in BDS 5.0 messages.

- In BDS 6.0, the indicated airspeed is available. This can, in turn, be converted into true airspeed, assuming that indicated airspeed is equal to the calibrated airspeed. Under the ISA condition, the true airspeed can be calculated using Equation 4.

\[
V_{\text{TAS}} = \left\{ \frac{2\kappa}{\kappa-1} \rho \left[ 1 + \frac{\rho_0}{\rho} \left( \frac{\kappa - 1}{2\kappa} \rho \frac{V_{\text{CAS}}}{V_{\text{TAS}}} \right)^{\frac{\kappa-1}{\kappa}} - 1 \right] \right\}^{\frac{1}{2}}
\]

where \( \rho \) and \( \rho_0 \) are pressure and air density respectively. \( \kappa \) is the specific heat ratio of 1.4. Parameters with subscript 0 represent their values at sea level. Furthermore, since BDS 6.0 also contains the Mach number, the true airspeed can also be derived more accurately with Equation 5.

\[
V_{\text{TAS}} = \frac{V_{\text{CAS}}}{f(\rho, M)} \sqrt{\frac{\rho_0}{\rho}}
\]

\[
f(\rho, M) = 1 + \left( \frac{\rho}{\rho_0} \right)^2 M^2 + \frac{3}{640} \left( 1 - \frac{10}{\rho_0} \frac{\rho}{\rho_0} + \frac{9 \rho^2}{\rho_0^2} \right) M^4
\]

- In order to determine the airspeed vector, heading is required. Aircraft magnetic heading can be obtained via BDS 6.0. However, the heading information refers to the magnetic heading of aircraft, which should be converted into true north depending on the location of the aircraft.

- When BDS 4.4 (MRAR) messages are detected, direct meteorological information on wind can be decoded.

### III. The particle model

The idea behind the proposed particle model is to mimic gas particles in nature. The particles are modeled to carry the states of a wind measurement. Particles are first

---

1 The magnetic declination in the Netherlands is only around 1 degree. Thus, for simplification purpose, in later datasets, the heading is assumed to be equal to the magnetic heading.
generated when a new measurement of wind is obtained and decay over time according to a certain parameterized kernel function. Using a stochastic process model, these particle propagates within the airspace. Wind fields are constructed by combining the weighted states of all neighboring particles. The propagation of particles allows for wind at areas of low measurement density to be computed. The following section will be dedicated to a more detailed explanation of the model, methods, and kernel functions used to compute wind field and confidences.

A. Measurement array

A single wind measurement is a 2D vector represented by a west-east \( w_x \) and a south-north \( w_y \) component, in 3D space \( \{x, y, z\} \). The measurement array consists of all wind measurements from different aircraft at a given time, defined as \( \{X, Y, Z, W_x, W_y\} \).

B. Particles

A particle is defined as a point object that carries the states of wind. Particle states consist of position \( \{p_x, p_y, p_z\} \), origin \( \{p_{x0}, p_{y0}, p_{z0}\} \), carried wind states \( \{s_{w_x}, s_{w_y}\} \), and age \( \tau \).

Particles are generated when new wind measurements are observed (computed). For any new measurement vector \( \{X, Y, Z, W_x, W_y\} \) with \( M \) measurements, \( N \) number of particles are created for each measurement. The total number of \( M \times N \) particles is generated from a multivariate normal distribution, using the aircraft position as the mean value:

\[
\begin{bmatrix}
    p_{x,mn} \\
p_{y,mn} \\
p_{z,mn}
\end{bmatrix} \sim N\left( \begin{bmatrix} x_m \\ y_m \\ z_m \end{bmatrix}, \begin{bmatrix} \sigma_{x}^2 & 0 & 0 \\
0 & \sigma_{y}^2 & 0 \\
0 & 0 & \sigma_{z}^2 \end{bmatrix} \right)
\]

The carried states of particles are also assigned a small variance that represents the uncertainty of the wind measurement:

\[
\begin{bmatrix}
s_{w_x,mn} \\
s_{w_y,mn}
\end{bmatrix} \sim N\left( \begin{bmatrix} w_{x,m} \\ w_{y,m} \end{bmatrix}, \begin{bmatrix} \sigma_{w_x}^2 & 0 \\
0 & \sigma_{w_y}^2 \end{bmatrix} \right)
\]

As an example, Figure 2 displays the measurement vectors in solid arrows and generated particles in tiny vectors. (Note that only 10% of the particle samples are shown for a more clear illustration.) The plot shows the 2D projection of the X-Y plane. The dashed circles indicate the variance of particle positions in relation to the measurement location.

**FIGURE 2 — Wind measurements and corresponding particles**

![Wind measurements and corresponding particles](image)
C. Particle motion model

Particle motion follows a Gaussian random walk model that takes into consideration of the actual wind vector \((\mathbf{s}_w, \mathbf{s}_v)\). At each step update, the particle age \((\tau)\) increases. The following equation describes the motion model of a particle.

\[
\begin{align*}
\mathbf{p}_{(x,y,z,t+1)} &= \mathbf{p}_{(x,y,z,t)} + \mathbf{e}_{(x,y,z)} \\
\mathbf{e}_{(x,y,z)} &\sim \mathcal{N}\left(\begin{bmatrix} k^*s_w \mathbf{s}_w \\
\mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma^2_{px} & 0 \\
0 & \sigma^2_{py} \end{bmatrix}\right) \\
\mathbf{e}_z &\sim \mathcal{N}(0, \sigma^2_z)
\end{align*}
\]

The step factor \(\varepsilon\) is different in the horizontal and vertical direction. Horizontally, the term \(k \cdot \mathbf{s}_w\) allows the random walk executed with a small biased \(\cdot\) along the direction of wind, with a scaling factor \(k\). Choosing a larger \(k\) allows the propagation becomes more biased toward the wind direction. Vertically, the propagation follows a zero mean Gaussian walk. The particle motion model is illustrated in Figure 3, where two projections \((X-Y\) and \(X-Z\)) of a possible particle update are shown. The red dot represents the position \(\mathbf{p}_{(x,y,z,t)}\) while the probability of the next position \(\mathbf{p}_{(x,y,z,t+1)}\) is shown by the contour plot. The vector equals to \(\mathbb{E}[\mathbf{e}_{(x,y,z)}]\). Also note that the length of vectors and variances are not to their real scale. In reality, \(k \cdot \mathbf{s}_w\) is much smaller than variances.

**FIGURE 3 — Possible random update of a particle position**

The updates of particles follow the Gaussian random walk as shown in Figure 4, where several possible 100-step random walks of a particle \((\text{with origin } [0, 0, 0])\) are illustrated. Different trajectories are distinguished by different colors.

**FIGURE 4 — Examples of particle random walks in 3D**

\[k: 0.05, s_w: 5, s_v: 5, i:100\]
D. Wind field construction

The wind field is represented by a grid of equally spaced coordinates, which has the size of \( I \times J \times K \). Numbers \( I, J, \) and \( K \) represent the number of grid points at each axis. From each grid point \((x_i, y_j, z_k)\), the wind is constructed using the weighted wind state values from its neighboring \( P \) number of particles:

\[
\begin{bmatrix}
W_{x(i,j,k)} \\
W_{y(i,j,k)}
\end{bmatrix} = \frac{1}{\sum_{p=1}^{P} \omega_p} \sum_{p=1}^{P} \omega_p \cdot \begin{bmatrix}
S_{wx,p} \\
S_{wy,p}
\end{bmatrix}
\]

The \( \omega_p \) is the weight of each particle that is computed based on the product of two kernel functions. Function \( f_d(\cdot) \) draws an exponential relationship between weight and distance between the particle and the coordinate. Function \( f_0(\cdot) \) defines the weight of the particles and depends on distance to their origins. Function \( f_\tau(\cdot) \) expresses the similar relationship between weight and particle-age.

\[
\omega_p = f_d[d] \times f_0[d_0] \times f_\tau[\tau]
\]

Here, \( d \) represents the spatial distance between particle and grid point. \( C_d, C_0, \) and \( C_\tau \) are control parameters for the kernel functions \( f_d(\cdot), f_0(\cdot), \) and \( f_\tau(\cdot) \).

Figure 5 displays the constructed wind field from previously generated particles, at time-step zero. At each grid point, the wind vector is shown in solid arrows. Grid points with no information yet are presented in scattered circles.

FIGURE 5 — Wind field constructed from particles (10% particle samples illustrated)
E. Wind field confidence model

The confidence level of each grid point in the wind field is computed as the combination of confidence functions that are based on several independent factors. These factors are:

1. the number of particles in the vicinity of the grid point \( N \)
2. the mean distances between particles and the grid point \( D \)
3. the homogeneity of wind states carried by particles \( H \)
4. the strength of particles due to decaying function \( S \)

1. Particle numbers and distances

The idea behind these two confidence parameters is to give the wind field a higher confidence value where more and closer measurements are observed. On the contrary, areas that are far from flight trajectories tend to have less reachable particles and should yield a lower confidence value.

2. Homogeneity of carried states

The level of homogeneity refers to the similarity of particle states. It essentially indicates whether different measurements propagated from a nearby area indicate similar evidence of wind vectors. It is computed as the norm (spectral norm) of the covariance matrix of two wind states of all particles:

\[
H = \sum \| \lambda \| = \sqrt{\lambda_{\text{max}} \left( \sum \sum \right)}
\]

where the \( \lambda_{\text{max}} \) represents the largest eigenvalue of a matrix.

3. Particle strength

Since a particle’s creation, its age parameter \( \tau \) increases at each step of propagation. The decaying strength obtained by Equation 13 regularizes not only the weights of particles in wind calculation, but also the confidence. Mean strength of all neighborhood particles are calculated as follows:

\[
S = \frac{1}{N} \sum_{\rho=1}^{N} \left( S_{wp} \right)
\]

4. Normalized and combined confidence

Values from all four confidence factors all have a distinct range. It is important to normalize these factors into the same range. A linear scaler is used to covert all values of each factor into \([0, 1]\) range.

\[
s(x) = \frac{x - \min(x)}{\max(x) - \min(x)}
\]

At any given time, the confidence vectors that represent all wind grid points are: \( N, D, H, \) and \( S \). Then, the combined confidence is expressed as:

\[
C = \text{mean} \{ s(N), s(D), s(H), s(S) \}
\]
F. Compute wind at any points

It is worth pointing out that the particle model runs independently of a pre-defined grid. When looking at the model as a dynamic object, particles are generated as measurements are observed. They are propagated independently from then on. This property allows us not to store the measurement and still be able to compute wind at any given time and space.

Hence, wind states from this particle model are not limited to any symmetric grid points. Values can be computed at any point or any set of points. Equation 9 can be used at all locations to produce accurate wind state information, as long as a sufficient number of particles exist in the neighborhood of these locations. The confidence levels can also be calculated in the same fashion.

IV. Experiments and results

Firstly, a small data set from ADS-B and Mode-S are combined and used as wind observations as a means to generate the test wind field. Results are illustrated as a sampled wind grid. Later on, different validation criteria are proposed, and comparison experiments are conducted to examine the model and related results.

From ADS-B and Mode-S data, wind observations are calculated for the area that is covered by our ADS-B/Mode-S antenna. The area is about 600 kilometers in diameter and located around Delft, the Netherlands, as shown in Figure 7. Based on a one-hour continuous streaming of measurement data, wind vectors are computed on a 3D grid consisting of both horizontal and vertical points.

A. Constructing the wind field

The dataset consists of one hour of wind data obtained, from 11:30 to 12:30 hours on July 27, 2017. In total, around 87,600 wind measurements were generated during this one-hour period. In Figure 7, the distributions of wind observations are displayed both horizontally and vertically.
In this figure, the plot on the left hand side illustrates the ground projection of all observations. On the right hand side, the plot shows a histogram with the number of observations per 2,000 feet altitude. It is apparent that horizontally, the measurements are highly concentrated along flight routes. Vertically, the majority of the observations are at cruise altitudes and lower approaching altitudes.

Despite the horizontal location of the observations, the statistic of wind at different altitudes can be computed, as shown in Figure 8.

During this hour, it can furthermore be observed that wind generally comes from a west or south-west direction, but with different levels of velocities at different altitude levels. The time-spatial variate wind state also leads to variability in both wind velocity and direction at each altitude. Both Figure 7 and 8 illustrate the challenges of using aircraft as sensors to model atmospheric conditions, there being: 1) high non-uniform distribution, and 2) large variation in the time-spatial domain.

To simulate a real-time run of the model, these recorded wind data are streamed to the particle model using the original sequence based on the data time-stamp. Each second, there are around 11 wind observations on average computed by the receiver.

The entire area is converted with Cartesian coordinates centered at the location of receiver (latitude: 51.99°N, longitude: 4.37°E). For illustrative purposes, the horizontal visualization of wind grid size is set to 10 x 10, where each set of adjacent points are 60 km apart. Vertically, 12 equally separated flight levels are chosen for visualization. A snapshot of the wind grid at 12:00 hours is shown in Figure 9.
Visually, it can be ascertained that wind speed increases with increasing altitude. At lower altitudes, the wind generally comes from a south-west direction, while, at higher altitudes, wind generally comes from a westerly direction. Both speed and direction trends correspond with the observations from Figure 8. To further validate the accuracy of the results, several additional methods are addressed below.

**B. Validation of particle model**

The validation of the particle model is focused on two indicators: correctness and variability. The level of correctness can be examined against data from existing meteorological models. Global Forecast System (GFS) Analysis data are used for this purpose [2]. Variability can be caused by both the uncertainty in the model itself and the quality of data. Firstly, all particle generations and propagations follow a stochastic process. At each run and each step, the states of each individual particle cannot be predicted. Secondly, wind measurements from aircraft also carry their own uncertainties.

To validate the model variation, multiple runs are performed for the same data. Differences in the resulting wind grid are compared. To validate the data variation, a complete dataset is sampled into different sizes of test datasets. Then, the results are compared with those from the complete dataset.

1. **Correctness**

To improve the quality in correctness validation, a much larger number of wind data samples across an entire week are used. GFS Analysis data provide global atmospheric re-analysis of all vertical levels at the highest resolution of 0.25 degrees latitudinal and longitudinal, at 00:00, 06:00, 12:00, and 18:00 hour each day. A wind observation dataset
is aggregated that contains seven days of one-hour data computed around these hours from GFS, lasting from the 24th to 30th of July, 2017.

At a GFS hour (00:00, 06:00, 12:00, or 18:00), spot values and average values are computed. The spot value is the wind grid computed from the particle model at the exact GFS hour. The average values are computed as the mean of the hour around GFS hour (per minute, +/- 30 minute of wind grids).

To compare the difference in wind vectors from GFS and the particle model at the same position and time, two distance matrices - angular difference and magnitude difference - are calculated at each GFS hour.

The angular difference is computed as

\[ \Delta \theta = \arccos \left( \frac{V_{pm} \cdot V_{gfs}}{|V_{pm}| \cdot |V_{gfs}|} \right) \]

where \( V_{pm} \) and \( V_{gfs} \) are the two wind vectors computed by the particle model and extracted from GFS respectively. \( \Delta \theta \) is the angle in degrees between two wind vectors with a range of \([0, 180]\). The smaller the \( \Delta \theta \), the smaller the angular difference between the two wind vectors.

The magnitude difference is computed as the absolute difference of wind vectors:

\[ \Delta V = |V_{pm} - V_{gfs}| \]

where the smaller the value, the smaller the difference between the two results.

Table II summarizes the results of the 7-day mean angular and magnitude differences between the wind field generated from the particle model and GFS data. The analysis is divided in two parts – part one, with up to gentle breeze wind (less than 10 m/s) and, part two, with higher speed wind (greater than 10 m/s). This distinction is needed because that small variability can cause large relative differences in low wind conditions. If analyze without such consideration, the low wind difference could be simply considered as outliers.

Both the spot value and average value are generated from the particle mode to compare with the GFS. They are also illustrated in box-plots in Figure 10. Within this 7-day data sample, the mean angular difference of the two wind fields is around 20 degrees for low speed wind and around 10 degrees for higher wind speeds. The magnitude difference is around 4-5 m/s as compared to around 20 m/s average wind speed.

When using a one hour (60 sample) average, the differences become small, but not significantly different. This is due to the fact that the particle model is already considering historical wind measurement as a result of the decaying factor. For example, in this experimental setting, historical observations of up to one minute still persist in the particle model.

It is apparent that for low wind speeds, the results are less aligned with the GFS data. However, this does not mean that the results are less accurate. Rather, the wind information generated by the GFS model is smoothed and interpolated over much larger periods of time and areas.
TABLE II — Mean grid angular and magnitude distance

<table>
<thead>
<tr>
<th>Date time</th>
<th>$\Delta \theta$</th>
<th>$\Delta V$</th>
<th>$\Delta \theta$</th>
<th>$\Delta V$</th>
<th>$\Delta \theta$</th>
<th>$\Delta V$</th>
<th>$\Delta \theta$</th>
<th>$\Delta V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-07-24 00H</td>
<td>23.22</td>
<td>4.09</td>
<td>26.42</td>
<td>3.18</td>
<td>18.96</td>
<td>7.64</td>
<td>10.75</td>
<td>5.26</td>
</tr>
<tr>
<td>2017-07-24 06H</td>
<td>56.31</td>
<td>2.97</td>
<td>43.79</td>
<td>2.03</td>
<td>21.03</td>
<td>6.38</td>
<td>14.45</td>
<td>6.11</td>
</tr>
<tr>
<td>2017-07-24 12H</td>
<td>36.46</td>
<td>2.66</td>
<td>30.14</td>
<td>2.06</td>
<td>13.80</td>
<td>5.80</td>
<td>8.30</td>
<td>4.97</td>
</tr>
<tr>
<td>2017-07-24 18H</td>
<td>22.02</td>
<td>2.52</td>
<td>20.13</td>
<td>1.92</td>
<td>9.71</td>
<td>5.02</td>
<td>8.28</td>
<td>3.49</td>
</tr>
<tr>
<td>2017-07-25 00H</td>
<td>27.85</td>
<td>7.04</td>
<td>19.54</td>
<td>2.20</td>
<td>14.81</td>
<td>4.57</td>
<td>9.50</td>
<td>3.71</td>
</tr>
<tr>
<td>2017-07-25 06H</td>
<td>14.84</td>
<td>2.11</td>
<td>13.50</td>
<td>2.60</td>
<td>12.03</td>
<td>3.56</td>
<td>9.03</td>
<td>2.47</td>
</tr>
<tr>
<td>2017-07-25 12H</td>
<td>19.52</td>
<td>1.97</td>
<td>14.29</td>
<td>2.43</td>
<td>12.69</td>
<td>2.12</td>
<td>9.30</td>
<td>2.06</td>
</tr>
<tr>
<td>2017-07-25 18H</td>
<td>29.71</td>
<td>2.82</td>
<td>20.70</td>
<td>2.24</td>
<td>17.02</td>
<td>2.71</td>
<td>11.46</td>
<td>2.27</td>
</tr>
<tr>
<td>2017-07-26 00H</td>
<td>18.39</td>
<td>3.53</td>
<td>27.14</td>
<td>3.74</td>
<td>11.82</td>
<td>7.89</td>
<td>8.87</td>
<td>5.76</td>
</tr>
<tr>
<td>2017-07-26 12H</td>
<td>23.71</td>
<td>1.98</td>
<td>18.50</td>
<td>1.88</td>
<td>16.56</td>
<td>3.57</td>
<td>14.53</td>
<td>2.51</td>
</tr>
<tr>
<td>2017-07-27 00H</td>
<td>16.87</td>
<td>3.79</td>
<td>15.71</td>
<td>6.98</td>
<td>10.95</td>
<td>4.84</td>
<td>8.05</td>
<td>3.77</td>
</tr>
<tr>
<td>2017-07-27 06H</td>
<td>20.73</td>
<td>5.43</td>
<td>14.84</td>
<td>6.06</td>
<td>10.75</td>
<td>5.58</td>
<td>7.70</td>
<td>4.12</td>
</tr>
<tr>
<td>2017-07-27 18H</td>
<td>11.19</td>
<td>3.28</td>
<td>12.97</td>
<td>4.23</td>
<td>7.40</td>
<td>6.68</td>
<td>8.15</td>
<td>4.46</td>
</tr>
<tr>
<td>2017-07-28 00H</td>
<td>17.11</td>
<td>3.21</td>
<td>17.05</td>
<td>6.76</td>
<td>8.29</td>
<td>8.85</td>
<td>8.21</td>
<td>6.11</td>
</tr>
<tr>
<td>2017-07-28 06H</td>
<td>13.70</td>
<td>5.15</td>
<td>30.79</td>
<td>6.15</td>
<td>6.00</td>
<td>6.11</td>
<td>5.62</td>
<td>5.19</td>
</tr>
<tr>
<td>2017-07-28 12H</td>
<td>13.58</td>
<td>5.38</td>
<td>30.10</td>
<td>8.48</td>
<td>6.04</td>
<td>5.78</td>
<td>7.82</td>
<td>4.86</td>
</tr>
<tr>
<td>2017-07-29 00H</td>
<td>45.10</td>
<td>13.04</td>
<td>48.62</td>
<td>10.00</td>
<td>10.58</td>
<td>5.10</td>
<td>8.35</td>
<td>3.81</td>
</tr>
<tr>
<td>2017-07-29 06H</td>
<td>25.36</td>
<td>9.34</td>
<td>28.88</td>
<td>11.60</td>
<td>5.35</td>
<td>4.57</td>
<td>4.91</td>
<td>4.25</td>
</tr>
<tr>
<td>2017-07-29 12H</td>
<td>12.45</td>
<td>6.00</td>
<td>14.61</td>
<td>9.55</td>
<td>5.30</td>
<td>3.83</td>
<td>3.16</td>
<td>3.26</td>
</tr>
<tr>
<td>2017-07-29 18H</td>
<td>19.82</td>
<td>9.84</td>
<td>15.93</td>
<td>10.84</td>
<td>8.50</td>
<td>3.75</td>
<td>5.05</td>
<td>3.27</td>
</tr>
<tr>
<td>2017-07-30 00H</td>
<td>37.08</td>
<td>9.44</td>
<td>45.83</td>
<td>10.13</td>
<td>6.66</td>
<td>5.11</td>
<td>5.34</td>
<td>3.69</td>
</tr>
<tr>
<td>2017-07-30 06H</td>
<td>18.68</td>
<td>11.31</td>
<td>24.54</td>
<td>10.66</td>
<td>6.21</td>
<td>5.34</td>
<td>4.67</td>
<td>4.26</td>
</tr>
<tr>
<td>2017-07-30 12H</td>
<td>17.89</td>
<td>5.44</td>
<td>11.02</td>
<td>6.96</td>
<td>8.01</td>
<td>4.59</td>
<td>3.75</td>
<td>3.54</td>
</tr>
<tr>
<td>2017-07-30 18H</td>
<td>33.50</td>
<td>5.15</td>
<td>25.26</td>
<td>7.91</td>
<td>5.74</td>
<td>4.63</td>
<td>3.73</td>
<td>3.96</td>
</tr>
</tbody>
</table>

FIGURE 10 — Mean wind grid angular and magnitude difference

2. Model variation

As previously stated, there is a certain level of the randomness in the particle model. The advantage of such randomness is that the model mimics and copes with the uncertainty of wind. With a large amount of particles, the general trend of wind is (hopefully) stable. To study whether the randomness of particles effects the wind field, as well as the level of the influence, the same example in Section IV-A is used with multiple runs of the particle model. The wind field at 12:00 hour (as shown in Figure 9) is computed at the end of each run.

In Figure 11, the distribution of standard deviations of wind grid speed and heading among 100 runs is displayed. Among these runs, the difference is almost negligible,
namely less than one degree for heading and 1 m/s for magnitude. Using box-plots, the baseline variance is also illustrated for later comparisons.

**FIGURE 11 — Standard deviation of wind speeds and headings of 100 runs**

3. **Data variation**

Another important validation is to determine how the quality of observation data effects the wind field estimation. More precisely, it is necessary to ascertain whether the wind grid would be different if some percentage of the observed data are not available. To study this effect, the previous dataset is randomly sampled into several new datasets that contain 90%, 70%, 50%, 30%, and 10% of the total wind observations. Then, the same processes are run to create five different wind fields at 12:00 hour.

Figure 12 illustrates the wind grid estimated at the altitude level of 12 km when different percentages of sampled data are used. From the first plot to the last, it is obvious that with increasing observation data samples, the size of estimated wind field is increased, together with an increased level of confidence. Visually, it is already possible to observe that the magnitude and headings of wind field are quite similar.

**FIGURE 12 — Wind field at a 12 km altitude from different samples**

In order to quantify the differences, mean heading and magnitude differences from the entire grid (including all altitude levels) of all wind vectors are compared with the results from the complete data, being shown in Figure 13.

It is apparent that the comparison with box-plots confirms the previous observation. Compared to the baseline variance of the model as shown in Figure 11, we can infer that with a loss of up to 50% of the total data, the differences are still within the acceptable range. This test indicates that, within a reasonable percentage of data uncertainty, the particle model can always obtain relatively stable wind field results.

---

2 Grid points with higher variances are usually along the boundary regions with fewer particles.
4. Error resistance

The factor that effects the stability and correctness most is the fundamental measurement error in raw data. It is apparent that with a better accuracy in wind observations, the wind field can be better re-constructed. In Section II, from Mode-S and ADS-B data, different inference methods were implemented that trying to produce a higher level of wind accuracy. However, there is no reference data to check the correctness of the computed wind vectors themselves at this stage.

To study how errors in data would affect the wind field, a percentage of the dataset is replaced with random wind vectors that are uniformly distributed between the minimum and maximum wind speeds with headings between 0 and 360 degrees. In Figure 14, wind grid differences between no assumed error and data error rates of 2%, 4%, 6%, 8%, 10%, and 15% are shown.

With such an aggressive error model, the particle model can maintain a reasonably correct wind field with up to an error rate of approximately 5%. One can further infer that if the magnitude and heading errors are small (in another words, wind observations distributed close to their true values), the particle model would be able to handle an even larger percentage of data error.
V. Discussion

The introduction to this paper indicates four challenges posed by using aircraft as a sensor network to construct wind fields. Throughout the paper, methods and models are proposed to address those challenges.

The source - wind observations - remains the most important factor that influences the results. Without going into much detail, the computation of aircraft true airspeed in pyModeS is a complicated task. The challenge is not only the decoding of BDS 5,0 and 6,0 messages, but rather a complete identification process of the entire Mode-S family of messages. As a third-party observer without the knowledge of Mode-S interrogations, the decoding is extremely complex. Developed by the authors of this paper and supported by open-source community, pyModeS is an effective tool to solve this problem. Sophisticated identification processes can be found in the source code of the software as referenced.

Remaining challenges include constructing a model that is able to cope with the chaotic nature of wind, moving sensors (aircraft), and extreme non-uniformly distributed observations. The particle model proposed in this paper addresses the stochastic characteristic of wind through particles, while maintaining the stability through the use of relatively large particle numbers. One must not confuse the model with Particle Filtering. The particle model mimics the gas particles' stochastic motions to propagate wind information (not actual wind) to their surrounding areas. Using the propagated information, wind filed in areas with less (or no) measurements can be estimated. Parameters on particle propagation and decaying can be tuned in order to control performance. These parameters are set empirically in this paper. However, for future work, an automatic parameter tuning method shall also be constructed.

As a novel approach, there are still a few remaining future developments. For example, together with wind, air temperature field can also be computed using Mode-S data. Since this paper is focused on the concept of the particle model, temperature has not been included in the scope of this paper. Air temperatures are generally more stable and evenly distributed spatially. Hence, using the same particle model with addition temperature state, a similar field can be generated in parallel.

VI. Conclusions

With the increasing accessibility of open ATM data from ADS-B and Mode-S, as well as related open-source decoding library development, exciting new possibilities for research are made available to the open research community. This paper proposes an open framework to construct accurate real-time wind field using aircraft as sensors.

At first, using our pyModeS library, raw wind vectors are computed from ADS-B and Mode-S down-link data. Then, a novel and fast particle model constructs wind field on a large scale. This model is self-evaluated in order to understand its variability and resistance to errors. The accuracy of calculated wind fields are also validated against GFS data, using data from 28 sets over a long period (one week).

As the result, the combination of accurate wind data from pyModeS and the fast fault resistance particle model is convincing evidence of the utility of open source solutions in ATM research. Our model clearly shows the possibility of using aircraft as large sensor networks to construct a global scale real-time meteorological measuring system under the open-source domain. In stark contrast to the current propitiatory AMDAR system, this model and the results proposed in this paper are fully open to the ATM and the wider scientific community. Without the need for any new equipment or communication protocols, the implementation of such a system is completely based on existing technology and data sources.
REFERENCES


Improving snow nowcasts for airports

Seppo Pulkkinen, Annakaisa von Lerber, Elena Saltikoff
Research department Finnish Meteorological Institute Helsinki, Finland
First.lastname@fmi.fi

Martin Hagen
Institute of Atmospheric Physics German Aerospace Center (DLR) Oberpfaffenhofen, Germany
First.lastname@dlr.de

Abstract—PNOWWA [Probabilistic Nowcasting of Winter Weather for Airports] project has studied methods to forecast snowfall for next few hours by extrapolating movement of radar echoes. Three different methods to create motion vectors [a simple method, a method used operationally and a new method] as well as three methods to produce probability forecasts with help of a motion vector field have been studied. The verification results of four case studies show a large dependence of the weather regime: widespread frontal precipitation is easier to forecast than isolated snow showers. The effect of orography can be split to quantitative enhancement of snowfall due to lower hills and mountains, and dynamic effect of the Alps which are effecting the movement of the entire weather system. Here the forecasts using motion extrapolation will often fail due to the complex interaction of synoptic-scale systems with orography. A further aspect in this study is the conversion of radar reflectivity (either forecasted or actual measured) to parameters which are of interest for the airport operation like visibility, de-icing index, or snow accumulation. Conversion formulas will be provided for easy use, even though there is a large uncertainty due to the wide variability of the shape and density of ice particles or snowflakes.

I. Introduction

The PNOWWA [Probabilistic Nowcasting of Winter Weather for Airports] project produces methods for the probabilistic short-term forecasting of winter weather. A survey of user needs has shown demand of short, detailed forecasts [nowcasts] of snowfall and related phenomena such as the decrease of visibility, and accumulation on runways. Empirical conversion formulas will be provided, however, due to the large variability of ice particles or snowflakes there is a large scatter in the retrieved parameters.

The approach taken in PNOWWA is based on extrapolation of the movement of snowfall areas in consecutive radar images. Extrapolative methods have their limits, but in the very short range forecasting they have the ability to express timing of short-lived phenomena, such as a 45 minute pause in snowfall. Very few other methods can do that. The presence of mountains in the vicinity of airports will considerably influence the behavior of precipitation systems and thus the predictability in short time ranges. This is studied in detail for the airports of Oslo in Norway and Rovaniemi in Northern Finland. Even more complex is the situation for the airports of Munich in Southern Germany and Salzburg in Austria where there is a strong interaction between the Alps and synoptic-scale precipitation systems.
II. Nowcasting methods

A. Motion of precipitation

In a method suggested by Andersson and Ivarsson [1] the wind at 850 hPa level is used to describe the movement. The wind is taken from HIRLAM (High Resolution Limited Area Model) numerical weather prediction model. This approach had been tested in SESAR1, so it was known to provide reasonable results.

The method operationally used and originally developed at FMI, applies modified correlation-based atmospheric motion vector (AMV) system by EUMETSAT [2]. It is described in detail in [3]. The AMV system was originally developed to extract wind data from METEOSAT imagery to be used as input for numerical weather prediction. For that purpose it provides a sophisticated automatic quality indicator (QI) of the vectors [4], which is also useful in application to the radar images.

The five latest 500 m PseudoCAPPI reflectivity fields combined from ten radars in Finland are used as the input to the AMV system. Radar echoes with reflectivity less than 0 dBZ are removed from the analysis. Data is processed at time steps of 5 minutes, in grid of 16 x 16 km grid boxes. Each grid box is compared to the neighboring grid boxes from the previous time step, and the best autocorrelation is chosen to show the area of origin of the precipitation cells in the grid box.

FIGURE 1 — Andersson and Ivarsson method. Colorful 60 degree sector illustrates the direction, from where radar echoes are moving towards the airport. Distribution of snow/dry in each segment represents probability of snowfall during one 15 minute timestep.

The quality indicators of atmospheric motion vectors consist of five separate parts: consistency is tested for direction, speed, vectors, spatial homogeneity and for first guess field. The five parameters are then combined to one normalized quality indicator QI, details of this are described in [4]. In the application for radar images, vectors with quality index QI greater than 0.7 are included. This allows scatter of 1-15 degrees in direction, and 5%-20% in speed.

After rejecting vectors with quality index smaller than 0.7, a smooth vector field is analyzed by applying a modified Barnes interpolation scheme [5], [6].

When the method was developed at FMI, tests showed that it is crucial to have radar data at time steps of 5 minutes; otherwise the rapid development of precipitation systems is too large compared to the change of radar image by movements of the existing precipitation systems. European wide OPERA data is available only at steps of 15 minutes. For the area of Finland, the motion vectors from FMI’s operational nowcasting were archived for comparison purposes in case studies.
The new motion vector analysis schema [7] is based on approach of optical flow, introduced by Proesmans et al. [8]. The novel, consistency-driven technique implemented in PNOWWA is based on the intuition that for a reliable estimate, the forward- and backward-computed motion vectors should have opposite directions. When this is not the case, there is likely to be growth or decay of precipitation or measurement artifacts. All of such phenomena make the motion estimation problem ill-defined, as the optical flow methods are based on the assumption that intensity of the tracked features is preserved.

**FIGURE 2 — RAVAKE method.** Example of using vectors backwards to determine, which pixels will arrive to the airport (star) after 3 timesteps. The ellipse indicates uncertainty of the vector field: in this case the deterministic forecast would be “dry”, but there is a small probability that the radar echoes in the upper half of the ellipse arrive at the airport at the validation moment.

![Image of RAVAKE method](image)

The new method aims at minimization of a cost function that penalizes intensity changes and motion inconsistencies. This leads to a set of coupled differential equations, for which we have implemented a numerical solver. The computations are done in multiple spatial scales in order to increase robustness to large advection velocities. An example motion field estimated with the method is shown in Figure 3.

**FIGURE 3 — Proesman method** Motion field estimated from two radar reflectivity images, snowfall case Vantaa 12th January 2016.

![Image of Proesman method](image)

A key feature of the proposed approach is that it provides confidence estimates for motion vectors based on their consistency. The proposed method was compared to four state of the art optical flow methods and it showed to be more robust and to provide the most reliable confidence estimates.

During testing for research demonstration campaign, it was noticed that the method was not yet suitable for extrapolative use without further development. Because of
residual, non-moving targets in radar composites, the method was not able to generate proper motion vectors either to these areas or the areas without precipitation. This issue was improved later on by introducing additional quality filters and thresholds to the input data.

B. Approaches of probability forecasting

The method by Andersson and Ivarsson [1] was used with 60 degree movement uncertainty sector (Figure 1). The sector in each airport is divided to sections corresponding to the movement of radar echoes during each 15 minute nowcast interval. The content of each section is analyzed to get the probability distribution of precipitation intensity.

The number of pixels in each intensity class was divided by number of pixels in the entire section (assuming that each pixel has equally large probability to arrive at the target point at the validity moment of the forecast).

In the FMI operational method the smoothed motion vector field is used “backwards” step by step: first finding out where is the pixel which should be at airport in 15 minutes, then seeing where that pixel comes from etc. Distribution of pixels in an ellipse at the source point is used to estimate the uncertainty and hence converted to probability distribution. The dimensions of the ellipse come from quality indicator of the vector field (lower quality, larger ellipse), and pixels within the ellipse get a Gaussian weighing, the pixels in the center having the largest weight.

The stochastic ensemble method can use any motion vector field as an input. It is the only method which assesses also the uncertainty due to growth and decay of the precipitation systems, not only the uncertainty in the motion field.

It is known that forecast uncertainty increases with lead time, and predictability is proportional to spatial scale (i.e. small-scale features have shorter lifetime). In the stochastic ensemble method this is modeled by autoregressive process in each spatial scale. Unexplained variance is gradually replaced with spatially correlated noise field.

Perturbations are added to the deterministic nowcast based on the motion field. 51 ensemble members are obtained by perturbing precipitation intensities and motion field. The ensemble mean represents the “most probable” precipitation intensity. The mean field becomes smoother when the forecast time increases: badly predictable scales are filtered out. The ensembles also yield probability distributions of precipitation intensities. At a given location, an empirical probability distribution for precipitation intensity can be constructed from the ensemble members.

III. Verification as case studies

These verification exercises are limited to comparing the nowcasts of radar reflectivity to observations of radar reflectivity. The research area was Southern Finland, and a period of 12 nowcasts at 5 min intervals were studied. Four cases were considered – radar images of all these are in Fig. 4:

- Case W is widespread precipitation 1 Feb 2015
- Case K was isolated snow showers 13 December 2015
- Case L was lake effect snow 3 – 9 January 2016.
- Case T was frontal precipitation 22 February 2017
The parameters selected for assessing quality of probability forecasts are Brier score, CSI and ROC. These are described in detail at website of WWRP/WGNE Joint Working Group on Forecast Verification Research or in book by Jolliffe and Stephenson [9]. Brier score answers the question: What is the magnitude of the probability forecast errors? It measures the mean squared probability error. Brier score range 0 to 1, with perfect score being 0. Brier scores for all four cases as function of forecast length are shown in Fig. 5a.

The Brier score can be decomposed into 3 additive components: Uncertainty, Reliability, and Resolution. The reliability term measures how close the forecast probabilities are to the true probabilities, given that forecast. For example, if we group all forecast instances where 80% chance of snowed was forecast, we get a perfect reliability only if it snowed 4 out of 5 times after such a forecast was issued. The resolution term measures how much the conditional probability given the different forecasts differs from the climatic average. The higher this term is the better. In the worst case, when the climatic probability is always forecast, the resolution is zero. In the best case, when the conditional probabilities are zero and one, the resolution is equal to the uncertainty.

In this short verification set, widespread precipitation case W gets almost perfect scores, because high probabilities are forecasted and snow is almost always observed. The lake effect case gets worst scores.

The Brier score includes comparison to climatology, which is not straightforward in comparing such disparate events. Our plan is to perform longer verification timeseries using Brier Skill Score, and the persistence as a reference forecast.
A more common verifications score, critical success index CSI, was calculated assuming that probability over 50% means “yes”. The CSI measures, how well did the forecast “yes” events correspond to the observed “yes” events? CSI combines probability of detection and the false alarm rate. CSI score ranges from 0 to 1, with perfect score being 1. CSIs for all four cases as function of forecast length are shown in Fig. 5b.

From the CSI scores we can see, as expected, that the predictability in case W (widespread precipitation) is high for all forecast lengths (because in whatever direction the snowfall area moves, it is still snowing everywhere), while for case K (scattered showers) the quality decreases rapidly. The frontal precipitation case T is nearly as good as W, and the lake effect case is somewhere between showers and the others.

Relative operating characteristic ROC is presented by plotting hit rate (POD) vs false alarm rate (POFD), using a set of increasing probability thresholds (0.1, 0.2, 0.3, etc.) to make the yes/no decision. The area under the ROC curve is frequently used as a score. ROC answers the question: What is the ability of the forecast to discriminate between events and non-events? A perfect curve travels from bottom left to top left of diagram, then across to top right of diagram. Diagonal line indicates no skill. ROC curves for 60 minute forecasts are shown in Fig. 6. It is easy to see, how for widespread precipitation the probability of detection stays relatively high but also the false alarm rate is relatively high, while for the showers case false alarms are more rare.

Longer verification periods and comparison of different nowcasting methods in the same situations are still to be calculated.

**FIGURE 6 — ROC curves for 60 minute forecasts for cases W (top left; 1 Feb 2015), Case K (top right; 13 December 2015), Case L (bottom left; 9 January 2016), and Case T (bottom right; 22 February 2017).**
IV. Conversion of parameters

A. Needs for conversions

All the nowcasting methods produce probability distribution of radar reflectivity in dBZ. PNOWWA Survey indicated the parameters which are most useful for different activities at the airport.

Several conversion equations are needed, using the reflectivity forecasts and a number of auxiliary data such as temperature and dewpoint, which are achieved from the METAR observations, TAF forecasts or HIRLAM numerical weather prediction model.

These conversion equations were used to express the user-defined parameters in radar reflectivity dBZ. The results, as used in the first scientific demo, are shown in tables 2 – 4.

The microphysical properties of different types of snowflakes are different. Some of the differences are related to temperature: ice needles fall in very cold temperatures, large “monster snowflakes” in near-zero temperatures, while wet snow is observed in even warmer weather.

To select the right dBZ thresholds, type of snow had to be determined. ICAO has defined the types of snow as follows [10]

- Dry snow – can be blown if loose or compacted by hand, will fall apart again upon release.
- Wet snow – can be compacted by hand and will stick together and tend to form a snowball.
- Compacted snow – can be compressed into a solid mass that resists further compression and will hold together, or break up into lumps, if picked up.

The most useful parameter to distinguish between snow and rain is the wet bulb temperature. Near 0°C of the wet-bulb temperature, both rain and snow are possible, and the probability of liquid rain increases with increasing wet bulb temperature. The image below (Fig. 7) shows the cumulative distribution of the non-zero rain-components as function of wet-bulb temperature, analyzed with a video-disdrometer Particle Imaging Package (PIP). As coarse analysis, it can be stated that until -0.45°C, snow can be treated as dry, with rain-component being less than 10% in the CDF and over 1.9°C, it can be treated as rain as 90% in CDF is considered to be composed of rain. In between wet snow values can be applied.

As wet bulb temperature is not included in standard METAR airport observations, the snow type for PNOWWA demos was determined based on temperature and dewpoint, read from the METAR.

The microphysical properties of different types of snowflakes were studied using the video-disdrometer Particle Imaging Package (PIP), OTT Pluvio2 weighing gauge and laser snow depth sensor (Jenoptik SM30).

In Fig. 8 the snow ratio for every 15 minutes is plotted as function of wet-bulb temperature. The median value is computed for every half degree bin. There are only few observations of the cold snow events, and there are not enough data points to make any conclusions. The larger values such as 40 even with temperatures close to 0°C are most likely because of the low precipitation rate, when resolution accuracy of Pluvio2 accumulation might be too coarse for computing the ratio. The mean value of snow ratio, 10.1, calculated between temperatures -4°C and -0.2°C, is selected to present the snow ratio in dry snow. The polynomial third-order fit is performed in between temperatures -0.2°C and 2°C to define the ratio in wet snow. Approximately it can be assumed to be 5.
FIGURE 7 — The 15 minute values of snow ratio for 95 snow cases, the median values calculated for every half degree bins of wet-bulb temperature between -14°C - 4°C are depicted in bracket line and solid line show the mean value of snow ratio in temperature region of -4°C – (-0.2)°C and the fit describing the change of snow ratio as function of wet-bulb temperature.

FIGURE 8 — Cumulative distribution function of rain component of the precipitation rate higher than zero as a function of the wet-bulb temperatures.

<table>
<thead>
<tr>
<th>Visibility dBZ for dry snow</th>
<th>dBZ for wet snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=600</td>
<td>&gt;29.0</td>
</tr>
<tr>
<td>600-1500</td>
<td>24.5-29.0</td>
</tr>
<tr>
<td>1500-3000</td>
<td>15.5-24.5</td>
</tr>
<tr>
<td>&gt;3000</td>
<td>&lt;15.5</td>
</tr>
</tbody>
</table>

TABLE I — The dependency between visibility and radar reflectivity

<table>
<thead>
<tr>
<th>Liquid water equivalent mm/h dBZ for dry snow</th>
<th>dBZ for wet snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;=4</td>
<td>&gt;29.0</td>
</tr>
<tr>
<td>2-4</td>
<td>24.5-29.0</td>
</tr>
<tr>
<td>0.4-2</td>
<td>15.5-24.5</td>
</tr>
<tr>
<td>&lt;0.4</td>
<td>&lt;15.5</td>
</tr>
</tbody>
</table>

TABLE II — The dependency between liquid water equivalent and radar reflectivity

<table>
<thead>
<tr>
<th>Snow accumulation mm/15 min dBZ for dry snow</th>
<th>dBZ for wet snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;10</td>
<td>&gt;29.0</td>
</tr>
<tr>
<td>5-10</td>
<td>24.5-29.0</td>
</tr>
<tr>
<td>1-5</td>
<td>15.5-24.5</td>
</tr>
<tr>
<td>&lt;1</td>
<td>&lt;15.5</td>
</tr>
</tbody>
</table>

TABLE III — The dependency between snow accumulation and radar reflectivity

<table>
<thead>
<tr>
<th>De-icing dBZ for dry snow</th>
<th>dBZ for wet snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>&gt;24.5</td>
</tr>
<tr>
<td>2</td>
<td>15.5-24.5</td>
</tr>
</tbody>
</table>
B. Visibility

Both meteorological visibility and radar reflectivity are related to scattering properties of snowflakes: their size, type and amount. Still, using radar measurements for describing visibility in snowfall is a challenging task, mainly because of the strong dependence of extinction on microphysical parameters. For a given snowfall rate, the visibility has a large range of values, changing from a factor of 3 to 10 [11]. Particle size distribution has large effect on the relation between the radar reflectivity factor and the visibility. For example if the snow particle mass in the radar measurement volume stays the same, but though aggregation process the snow particles aggregate to snowflakes, the radar reflectivity increases strongly, but the visibility increases although the scattering cross-section enlarges [12], [13]. In her BSc thesis using measurements from Finland, Kaisa Ylinen found that the visibility of the same radar reflectivity factor was less than in the published studies. It can be speculated that this is because the cases she studied were in very cold weather (-10 °C to -30°C), while many other researchers have mainly studied cases in near-zero temperatures, and the type of snowflakes is strongly connected to temperature [14].

Fig. 9 shows an example from visibility measurements at Munich airport using radar measurements with the DWD weather radar located at Isen about 30 to the South-East of the airport. It should be noted that the radar measurements are about 300 m above the airport and therefore not always represent the visibility observations at the surface. This might partly explain the large scatter, but a considerable part of the scatter is related to the large variability in size, density and shape of snowflakes. The parametrizations indicated in Fig. 9 refer to Table 3 in [14].

FIGURE 9 — Visibility vs. reflectivity for Munich airport using Isen radar. Symbols indicate METAR snowfall intensities; fitting lines indicate different empirical relations [14].

V. Effect of mountains

A. Quantitative studies

When airflow approaches or comes over mountains, snowfall is more difficult to forecast than in other situations. The predictability is worse for all studied methods: extrapolation of radar images (which is the subject of PNOWWA), but also for TAF forecasts written by human forecasters, and for numerical weather prediction models.

The quantitative effect of sea and orography was estimated using the nowcasting system developed for SESAR1, which was run on additional periods. The forecasted parameter is DIW, de-icing weather, which is an index getting values 0-3. For the comparison, DIW index is calculated in three ways:

- DIW_e - Extrapolating the movement of radar echoes using the method described by [1]
The orographic effect was studied using the SESAR1 methods at two airports: Rovaniemi EFRO and Oslo Gardemoen ENGM.

Days were counted as orographic effect days if at 850 or at 925 hPa (in the case of EFRO also 950 hPa was taken into account, as the terrain and height differences are rather low there) was from the sector (180° – 250°) in Rovaniemi and from (80° – 180°) in Oslo. In most days the direction of the flow varies with time; the flow was considered coming from the valley when it remained in the sector at least two hours.

In almost all the situations, the radar-based extrapolation method (DIWe) was slightly better than the others. Only in average of all cases for the 2 – 3 h period model forecasts outperformed the radar extrapolation. In orographic situations DIWe was best for the whole 3-hour period. Fig. 10 shows the performance of DIWe for Rovaniemi and Oslo. If the flow is affected by mountains forecast quality is less than for all cases.

FIGURE 10 — Summary showing the extrapolation performance in Oslo (red) and Rovaniemi (blue).

**B. Dynamical studies**

For the airports of Munich [EDDM] and Salzburg [LOWS] the effect of the Alps on the behavior of cold fronts approaching from northerly directions was investigated. It is observed that cold fronts can either be delayed when approaching the Alps, other systems cross the Alpine Foreland and the Alps without delay, and even acceleration can be observed for fronts passing along the Alpine Foreland (e.g. [15] or [16]). Delayed systems can generate long-lasting (up to a few days) continuous rain or snow fall events. Numerical weather forecast can forecast the behavior on a long term basis. However, nowcasting for a time horizon of one to three hours extrapolation techniques are more favorable because numerical models need some spin-up time. Radar-based extrapolation techniques will fail in case of non-linear propagation speed and direction due to delay or acceleration.

22 cases from the winters (December – March) 2013-14, 2014-15, 2015-16, and 2016-17 [April] were investigated where cold fronts did approach the Alps in the Munich/Salzburg region. To increase the number of samples both situations with rain and snowfall at ground were considered. In about half of the cases the fronts did pass the Alpine Foreland without noticeable delay [cf. Fig. 11], whereas the other cases showed considerable delay of the frontal motion leading to long lasting precipitation events [cf. Fig. 12]. The duration of the events was between 8 and 46 hours.
Fig. 13 shows the distribution of the events in relation to the approaching direction of the frontal systems. To find relations between flow and behavior the wind profile as measured by the radio sonde München-Oberschleißheim (located in the Alpine foreland about 50 km north of the Alps) was investigated (cf. Fig. 13). However, there is no clear relation between the propagation direction of the fronts and the wind direction at the 850 and 500 hPa level (about 1000 m above the Alpine Foreland and 2 km above the main ridge). This is mainly caused by the fact that during winter when the tropopause is low the Alps act as a major obstacle and cause a considerable distortion of the atmospheric flow. Especially during those conditions which were classified as up-slope or delay often low pressure systems develop in the Alpine region causing long-lasting precipitation and no more distinctive motion characteristics. It also should be considered that on the pre-frontal side the flow is parallel to the front, i.e. a front approaching from North-West.
will have south-westerly flow ahead of the front. A further investigation of propagation direction, motion vectors for those events, and wind field is ongoing.

**FIGURE 13** — Approaching direction of cold fronts in winter for the Munich/Salzburg region. Blue: total number of events; orange: number of events being delayed/blocking by the Alps.

**FIGURE 14** — Wind direction for two levels (top row: 850 hPa; bottom row: 500 hPa) for situations with up-slope/delay and passage (left column: upslope/delay; right column: passage) of 22 cold fronts events during winter. Labels (and colors) indicate the approaching direction of the frontal systems.

**VI. Conclusions**

The four case studies used for verification show that the values of verification scores depend greatly on the weather situation. Hence, no conclusion can be drawn by comparing the skill of forecasts made during different time periods. Based on visual comparison of cases in different weather situations, the Proesmans method was found out to produce the most reliable motion fields. However, the robustness of the method in cases of poor quality of input data (residual clutter) or missing data must be further improved. Verification of results with a statistically representing dataset remains to be made after the improvements have been implemented.

The stochastic ensemble method is clearly our preferred solution for producing probabilistic nowcasts, as it is assessing more sources of uncertainty than the simpler methods. Work is needed to improve the computational performance and to define the hardware requirements to calculate the nowcasts for real-time service.

Many of the methods for converting radar reflectivity to liquid water equivalent, snow depth and visibility introduced in the literature need such knowledge of microphysics which is not available operationally at the airports. We will continue following new scientific articles in the subject.
VII. Future work

In this project, we have focused on radar-based methods due to their outstanding temporal resolution. In the possible follow-up projects, data fusion with other data sources such as numerical weather prediction should be considered, both to extend the valid time and widen the available weather parameters.

VIII. Acknowledgment

The authors would like to thank colleagues at Helsinki University for fruitful co-operation. The PNOWWA project has received funding from the SESAR Joint Undertaking under the European Union’s Horizon 2020 research and innovation programme under grant agreement No 699221.

REFERENCES

MULTI-CRITERIA ENVIRONMENTAL IMPACT ASSESSMENT AND OPTIMISATION OF AIRCRAFT TRAJECTORIES

Minimizing Environmental Impacts during Different Phases of Flight

Sigrun Matthes
DLR Institute of Atmospheric Physics
Münchner Straße 22
Oberpfaffenhofen, Germany
Sigrun.matthes@dlr.de

ATM4E Team
Project Consortium
SESAR2020, Exploratory Project

Abstract—Air traffic management as currently under development by the Single European Sky ATM Research program SESAR has an important role to play in reducing environmental impact of aviation, in addition to the improvements to be derived from new aircraft and engine technologies. Modelling capabilities are required to allow a multi-dimensional environmental impact assessment. This study presents a concept for a multi-criteria environmental assessment of aircraft trajectories as developed within the Exploratory Research Project ATM4E (SESAR2020). In that context we present ideas on future implementation of such advanced meteorological services into air traffic management and trajectory planning by relying on environmental change functions (ECFs). These ECFs represent environmental impact due to changes in air quality, noise and climate impact. In a case study for Europe prototype ECFs are implemented and a performance assessment of aircraft trajectories is performed for a one-day traffic sample. For a single flight fuel-optimal versus climate-optimized trajectory solution is evaluated using prototypic ECFs and identifying mitigation potential. The ultimate goal of such a concept is to make available a comprehensive assessment framework for environmental performance of aircraft operations, by providing key performance indicators on climate impact, air quality and noise, as well as a tool for environmental optimisation of aircraft trajectories. This framework would allow studying and characterising changes in traffic flows due to environmental optimisation, as well as studying trade-offs between distinct strategic measures.

I. Introduction

Comprehensive assessment of the environmental aspects of flight movements is of increasing interest to the aviation sector as a potential input for developing sustainable aviation strategies that consider climate impact, air quality and noise issues.
simultaneously. Consideration of environmental aspects in en-route flight planning is generally not operational practice apart from the economic goal to minimise fuel use and hence to reduce CO₂ emissions. Note that non-operational, i.e., research-related optimization tools include estimates for climate impacts in route optimization [e.g., [1]]. However, only recently climate impact indicators were considered in more detail, which take into account more than mere emission amounts, for example contrail occurrence and ozone changes from NOₓ emissions [2–8]. The reasons for this include a low TRL (technology readiness level) of a flight planning method that considers a multi-criteria environmental impact assessment and remaining uncertainty on strategic metrics of environmental impact to motivate environmental flight planning.

Aircraft trajectory optimisation has already started in the 1960s [9] while during the last decades development of approaches has been strongly supported by increasing capabilities of high performance computing. Optimisation tools exist that incorporate more detailed aircraft performance data, that consider meteorological data, e.g., wind and humidity, and that perform a full 4D optimisation. In common practice, route optimisation is driven by cost minimisation, hence those environmental aspects which translate into cash operating costs (COC), are taken into account. E.g., emissions of carbon dioxide enter into route optimisation as they directly correlate to fuel consumption. Other environmental impacts enter in COC optimisation through charges, e.g., noise or nitrogen oxide (NOₓ) emissions near an airport in case of associated airport charges. However, besides CO₂ climate impact, air traffic contributes to anthropogenic warming also by non-CO₂ impacts which are strongly dependent on the location, altitude, and time of emission. Overall, air traffic emissions contribute to anthropogenic warming by around 5% through CO₂ and non-CO₂ impacts [10,11] including contrail cirrus. Aviation stakeholders, European and national authorities implemented a series of initiatives that comprise in their workprogrammes the intention to make future aviation sustainable, e.g., the European Commission implemented under its Framework Programmes, CleanSky Joint Technology Initiative (JTI), ‘green’ aeronautical projects and SESAR2020 Joint Undertaking (JU). Previous research has shown that changing aircraft trajectories to avoid climate sensitive regions has the potential to reduce the climate impact of aviation [12]. Studies which focus on individual impact types e.g., [2,3,13–15] presented trade-offs between climate-optimised and cost-optimised trajectories for various regions of the earth (cross-polar, North Atlantic, Pacific traffic). More recent studies similarly exploited benefit and costs of contrails avoidance by analysing an aircraft trajectory [16] or tested route optimisation for climate optimisation [17]. Research aims to enhance our understanding of the environmental impacts of ATM operations and how they can be minimized during different flight phases.

The objective of this paper is (1) to present a concept for multi-criteria environmental assessment of aircraft trajectories, (2) to introduce meteorological (MET) data products which represent environmental impact at given location and time, so called environmental change functions, which we consider as advanced meteorological information which should be made available via ATM information infrastructure. Finally, (3) we apply the concept by presenting (a) a trajectory optimisation under cost-optimal conditions, providing environmental performance data for the assessment of aircraft trajectories using prototype environmental change functions (ECFs) and (b) an environmental optimisation of climate impact.

II. Environmental Impact assessment of Aviation

Aviation emissions change the atmospheric concentration of chemical components and hence disturb the radiative balance in the atmosphere and subsequently contribute to climate change. At the same time changes in concentrations of atmospheric components can cause an impact on local and regional air quality. Finally, aviation emits noise which influences noise levels at ground. As aviation emissions undergo complex physical and chemical transformation processes, the specific impact of aviation emissions depends on time and location of emission due to influence of e.g., background conditions, radiation and other meteorological parameters.
A. Climate impact of aviation

An assessment of climate impact of aviation requires knowledge generated by complex chemistry climate models, which simulate comprehensive atmospheric transformation processes and subsequently provide quantitative estimates on changes of the radiative balance and impact on climate. The study presented relies on an integrative measure which directly connects aviation emission to their climate impact [1-2]. This concept was applied to climate impact assessments in earlier studies [18], by using the initial cost function concept [19] which relied on cost functions pre-calculated with the comprehensive general circulation model EMAC [23] in a Lagrangian approach under specific meteorological conditions. Expanding this concept to a set of environmental impacts introduces the term environmental change function (ECF). A more comprehensive overview including a detailed description of how to generate ECFs is provided in [20].

Climate change functions depend on time and location of emission as the synoptical situations plays an important role, due to influence of e.g., background conditions, radiation and other meteorological parameters. For climate impact, one way to generate these ECF is to provide them as an annual mean change function, which are then climatological climate change functions. Another option is to generate them individually for a specific weather situation, or in conjunction with linking specific weather situation to an archetypical weather pattern as done for the North Atlantic Flight corridor within REACT4C, by deriving them from meteorological key parameters. A third option is to derive algorithmic ECFs (aECFs) which estimate the ECFs based on readily available MET info, i.e., temperature, humidity, vorticity, and background concentrations (meteorological key parameters).

In ATM4E we propose and test the applicability of aECFs (Section III.1), as algorithms allow online generation of ECF from meteorological forecast data which is crucial for future implementation. These climate change functions are calculated for aviation emissions having a direct or an indirect climate impact. Carbon dioxide, water vapour, particulate matter and contrail induced cloudiness (CiC) are among those having direct radiative and climate impact. Emissions with indirect radiative impact are nitrogen oxides (NOx) and particles. Hence, these ECFs are varying with location (position and altitude) and time and date of emission. We refer to average temperature response (ATR) as climate metric (Fig. 1), but do not refer to other possible climate metrics, in order to improve readability of the paper. ATR is computed by averaging the surface temperature response during the considered period, assuming sustained emissions with respective routing strategy applied during the whole period, e.g. two distinct periods, 20 and 100 years. However alternative climate metrics can be used in our overall concept in a similar way.

FIGURE 1 — Algorithmic ECFs as Average Temperature Response (ATR) for case study 19 Dec 2015 for water vapour (left), nitrogen oxides (ozone, middle) and contrail formation (right).
B. Local and regional environmental impacts of aviation

Aircraft operations near the ground produce an assortment of gaseous and particulate air contaminants that affect local air quality levels and potentially human health. Atmospheric concentrations found at surface level depend on emission strength but also on synoptic situation and associated physical and chemical mechanisms active in a specific region. In a polluted background atmosphere, aviation can contribute to exceedance of air quality limits, while in an unpolluted background atmosphere aircraft operations will cause less exceedances of air quality limits.

In addition, aircraft operations increase noise levels especially in localities over which aircraft are climbing out of and descending into airports. Noise is recognized from the WHO (World Health Organization) as a threat to human health and is probably the most significant concern for the residents of communities neighbouring airport. Minimizing the number of people significantly disturbed by aircraft noise is one of ICAO’s main priorities and one of the industry’s key environmental goals.

In this study we expand a modelling concept for climate optimisation to additionally comprise local impacts, air quality and noise issues, leading to a multi-dimensional, and multicriteria, environmental assessment and optimisation of aircraft trajectories. In that context we introduce environmental change functions (ECF), as well as an efficient method to derive ECFs from standard meteorology data.

C. Aircraft trajectories optimisation

The concept of environmental assessment presented here relies on trajectory calculation within two distinct trajectory optimisation tools, in order e.g. to study influence of trajectory optimisation on air traffic flows, and to identify mitigation potential of environmental optimisation.

The stand-alone model Trajectory Optimisation Module (TOM) is used for trajectory management and optimisation receiving input data on air traffic (city pairs), standard MET data and algorithmic ECFs on environmental impacts. TOM applies optimal control techniques in order to determine continuously optimised four-dimensional aircraft trajectories. For verification purposes a module for aircraft trajectory assessment and optimisation has been integrated in a global climate-chemistry model working interactively during atmospheric calculations.

Second, for verification purposes this module AirTraf is compared to another trajectory calculation model FAST. AirTraf is a module which has been integrated in a global climate-chemistry model working interactively during atmospheric calculations. AirTraf (version 1.0) [21,22] was developed as a verification tool for climate optimised routing strategies by analysing individual routing options for given city pairs. AirTraf is a submodel of the ECHAM/MESy Atmospheric Chemistry (EMAC) model [23,24] (ECHAM5 version 5.3.02, MESSy version 2.52) and simulates global air traffic (online) which is able to simulate aircraft trajectories under individual optimisation criteria. An aircraft performance model and International Civil Aviation Organization (ICAO) engine performance data [25] are used. A global air traffic plan is used and both short- and long-term simulations are performed taking into account the individual departure times. The Genetic Algorithm optimises flight trajectories with respect to a selected routing option, taking account of the local weather conditions for every flight, and finds an optimal trajectory including altitude changes.

III. Environmental Change Functions for ATM

A prerequisite for environmental assessment and optimization of aircraft trajectories is to develop an interface how to make available environmental impact information during
aircraft trajectory planning (ATM). For this purpose we define a concept how to establish an interface between ATM and environmental impact information, further developing the so-called climate cost function approach presented in [20,26].

A flowchart (Fig. 1) shows how standard MET information is complimented with algorithmic ECFs in order to be made available for trajectory optimisation, as advanced MET information service. Performance assessment of aircraft trajectories then comprises environmental performance data beside performance data, e.g., on fuel and time efficiency.

For the impact function which describes environmental impact of an aviation emission, we use first order approximation in a Taylor series. This mathematical description can be transformed to represent an overall objective function for trajectory calculation in this study by a penalty function approach as shown in [20]. As environmental impact of aviation emissions depends strongly on meteorological conditions, comprising physical and chemical parameters, provision of this advanced information is integrated as MET information service for the specific application of environmental performance.

FIGURE 2 — Flowchart of Environmental Assessment of ATM using ATM4E algorithmic environmental change functions (aECF) concept, elements introduced by ATM4E highlighted in green (from [20]).

A. Algorithmic weather-dependent Environmental change functions

The climate change and environmental change functions show a strong dependence on meteorological situation on a synoptical scale, hence they are weather-dependent. For that purpose high-quality meteorological information can be used for an accurate generation of such weather-dependent ECFs, which then reflect specific meteorological situation. As presented in [27], different approaches to determine these weather-dependent ECFs exist, i.e., either using MET information to classify synoptical situation according to archetypical weather patterns, as performed within earlier studies [28]. Alternatively, algorithmic ECFs can be developed by using directly spatially and temporally resolved standard MET information available, e.g., provided by a System Wide Information Management (SWIM) as implemented within SESAR, to derive environmental change associated to aviation emission as 4-dimensional functions.
Algorithms are used to establish such link between meteorological key parameters and associated environmental impact, which were identified from comprehensive analysis of environmental impact at a specific location and associated prevailing meteorological conditions. Hence, we define the term algorithmic ECF (aECF) in order to describe such algorithms which enable to calculate ECF from basic MET information.

Development of such algorithms require fundamental understanding of atmospheric processes, statistical analysis and high-quality synoptical scale meteorological information, in order to identify and validate robust relationships, e.g., [29], which need to be in a next step integrated as interactive MET information product in ATM tools. Such aECFs rely on meteorological parameters, e.g., atmospheric temperature, relative humidity, geopotential height, potential vorticity, or boundary layer height, combined with e.g., atmospheric concentration and transformation of key chemical species as well as radiation.

B. Verification of algorithm based environmental change functions

Before these aECFs are used for trajectory optimization, a verification process is performed to ensure that the aECFs serve their purposes by comparing results from two distinct calculation procedures for the overall climate impact of an air traffic sample. The EMAC/AirTraf is an appropriate simulation tool since it combines the Earth-system model EMAC with the air traffic simulation model AirTraf. Similarly, aviation emission are integrated as 3 dimensional flux fields to the atmospheric chemistry model [23], affecting the chemical composition of the atmosphere, identified with a specific tagging scheme [30,31] and changing radiative balance, respectively. This verification procedure is performed to ensure that the overall ATR calculated based on the aECFs matches the ATR calculated from the calculated impact in the atmospheric chemistry model, hence allows performing a proof of concept for aECFs.

IV. Case Study: Environmental Assessment and Optimisation of Air Traffic in Europe

We apply above concept for a multi-criteria environmental assessment of aviation operations in a case study for the European airspace, in order to provide environmental performance data and in order to test feasibility working towards environmental optimisation of air traffic operations. Results are shown for a European traffic sample, together with a sensitivity study on environmental optimisation of an aircraft trajectory.

A. Meteorological and Synoptical Information

As ECFs depend to a large extent on synoptical situation, particular focus was given on selecting candidate days in our case studies. For each day, contrail formation regions were identified using infrared satellite imagery and data from the ECMWF ERA-Interim re-analysis [32]. As indicator for photochemical activity in the atmosphere, the ozone production efficiency was determined with the ECHAM/MESSy atmospheric chemistry model. Additionally geopotential height was determined from ECMWF ERA-Interim re-analysis. For initial analysis, a meteorological situation is selected which represents a medium to high complexity of the meteorological environment which ATM is encountering, e.g. 18 Dec 2015. The 18th of December 2015 is selected as specific date for our case study. The day was characterised by a high-pressure ridge over Europe with the jet stream meandering for North.
B. Traffic sample and engine emissions

The air traffic over Europe on the selected day is used as a reference scenario for the optimization task in ATM4E. Further assumptions are made to filter the traffic data for a better processability. As ATM4E focuses on the European airspace, only intra-ECAC (European Civil Aviation Conference) flights are considered and only flights that can be modelled with aircraft performance data from EUROCONTROL’s Base of Aircraft Data (BADA) 4.0 are taken into account. This required simplification reduces the amount of available seat kilometres (ASK) in the data set by only 8–9%, since especially large commercial aircraft representing major parts of ASK are included in BADA 4.0. Lastly, flights which depart before or arrive after 18 December 2015 are filtered out leading to a final dataset of 13,276 flights (from originally 28,337). In our study [20] computation of profiles in a numerical trajectory simulation tool is described in more detail.

C. Engine emission and environmental performance

In order to assess the environmental impact caused by the traffic sample and in order to prepare trajectory optimization, for the described reference flight set the overall performance parameters with respect to gaseous aircraft emissions, the provoked contrail formation and the overall climate impact are calculated.

Four-dimensional (longitude, latitude, altitude, time) emission inventories are generated by simulating every flight in the traffic scenario and determining the corresponding emission distribution. Fig. 3 shows the NOx emission distribution at 12:00 p.m. UTC of the European traffic sample. Regions with potential persistent contrail formation were identified with a method relying on the Schmidt-Appleman criterion [33], and the contrail situation at 12:00 p.m. UTC is depicted in Fig. 3. From this criterion we calculate the distance flown under persistent contrail formation criteria shown in Table 1 by taking into account real weather conditions on that specific day, which corresponds to 5% of air distance in this representative traffic sample.

**FIGURE 3** — NOx emissions *(left)* and persistent contrail formation distance *(right)* per gridcell (0.25 x 0.25) integrated over a time period of 20 s on 18 December 2015, 12:00 p.m. UTC. Single aircraft are represented as black dots (1216 in total) from [20].
TABLE I — Performance parameter of European traffic sample: Cumulated emissions and Distances: 18 Dec 2015; from [20]

<table>
<thead>
<tr>
<th>Performance Parameter</th>
<th>Parameter</th>
<th>Amount</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFF</td>
<td>Air distancea</td>
<td>$1.42 \times 10^7$</td>
<td>km</td>
</tr>
<tr>
<td>EFF ENV</td>
<td>Carbon dioxide CO$_2$</td>
<td>$1.50 \times 10^8$</td>
<td>kg</td>
</tr>
<tr>
<td>ENV</td>
<td>Nitrogen oxides NO$_x$</td>
<td>$7.20 \times 10^5$</td>
<td>kg</td>
</tr>
<tr>
<td>ENV</td>
<td>LAQ Nitrogen oxides (NO$_x$)</td>
<td>$0.52 \times 10^4$</td>
<td>kg</td>
</tr>
<tr>
<td>ENV</td>
<td>Distance contrailing</td>
<td>$6.8 \times 10^5$</td>
<td>km</td>
</tr>
<tr>
<td>ENV</td>
<td>Climate impact ATR$_{20}$</td>
<td>$5.7(4.1-7.0)$</td>
<td>$10^{-3}$mK</td>
</tr>
<tr>
<td>ENV</td>
<td>Climate impact ATR$_{100}$</td>
<td>$16.7(12.1-20.3)$</td>
<td>$10^{-3}$mK</td>
</tr>
<tr>
<td>ENV</td>
<td>Climate ATR$_{20}$ non-CO$_2$/CO$_2$</td>
<td>$20.0(13.9-24.8)$</td>
<td>-</td>
</tr>
<tr>
<td>ENV</td>
<td>Climate ATR$_{100}$ non-CO$_2$/CO$_2$</td>
<td>$5.8(3.9-7.2)$</td>
<td>-</td>
</tr>
</tbody>
</table>

a. Uncertainty of environmental indicators indicated in parenthesis.

Overall performance data of the traffic sample, comprising air distance travelled cumulated emissions as well as the distance in contrail areas are listed for the chosen reference day. Among environmental performance data the overall climate impact has been evaluated for two distinct climate impact metrics. ATRs over 20 and 100 years have been calculated under the assumption of sustained emissions, which means that routing decision is similar on each day over the time horizon. Ratios of climate impacts of non-CO$_2$ versus CO$_2$-impacts are calculated for ATR$_{20}$ with 20.0, and for ATR$_{100}$ with a lower value of 5.8. Uncertainty range provided refers to sensitivity study on seasonal cycle and annual mean ECFs.

For local air quality the increase of atmospheric NO$_2$ concentration is estimated using parametric study to investigate sensitivities assuming moderate advection of trace compounds and low atmospheric loss rate. We present mean and maximum values for several vertical layers, i.e., ground level, up to 3000 and 5000 feet. Mean NO$_2$ concentration is estimated to increase by about 0.3 to 0.4 μg/m$^3$, with maximum increase of hourly values in specific regions in the order of up to 10.6 μg/m$^3$.

**D. Cost-optimal versus climate-optimal trajectory optimisation**

Beside environmental assessment of aircraft trajectories, the framework can also be applied in an environmental optimisation by adapting corresponding objective functions used in TOM. For a flight from London Heathrow (LHR) to Istanbul (IST) aircraft trajectory was optimized under a series of objectives functions, by varying individual weights from fuel optimal case to climate-minimal solution. ECFs used in this optimisation, are prototypes which were calculated from AirClim climatological mean ECFs. The resulting Pareto front is shown in Fig. 4, together with trajectories from three distinct solutions, the reference case, and solutions for 1% and 5% percent fuel increase, resulting in a climate impact mitigation by reducing ATR by 12% and 25%, respectively.
V. Development of Met Products on Environment

The ATM4E approach on environmental flight planning requires that verified advanced MET information are implemented in flight planning, providing the impact of a local emission on climate, air quality and noise. For that purpose ATM4E develops verified aECFs, which allow to provide among MET services both (standard) weather forecast information and advanced MET information, i.e. aECFs [Fig. 1]. This advanced MET information can be distributed with standard MET information, e.g., in a SWIM concept, while taking legislation into account in terms of objective function and flight planning. There is the possibility to implement algorithm-based environmental change functions into national weather forecast models, which then provide advanced information via services to users, allowing for an environmental flight planning, as well as short-term tactical adaptations to the aircraft trajectory. Having available ECFs during flight planning also offers the ability to environmentally assess the actually executed flight and to record the data for e.g., environmental legislative purpose.

It is proposed to use ECFs as interface between environmental expertise [derived from models] and air traffic management tools, in order to represent environmental impact in air trajectory models, instead of code integration in a flight planning tool. An interface (function) has the advantage that, first, complex and comprehensive systems and models, e.g., climate-chemistry model with coupled homogeneous and heterogeneous atmospheric chemistry, is used for ECF generation. Second, any updates due to scientific understanding or political decision, e.g., on time horizon of climate impact metric being considered, is integrated by simply replacing (mathematical formulation) of a specific aECF function.

Such advanced MET information offers the possibility to determine key performance indicators in the key performance area environment (KP05). Quantitative indicators providing information on climate impact, air quality impact and noise level can be derived by implementing aECFs in flight planning.

VI. Discussion

An environmental assessment of aircraft trajectories during ATM can be performed, if algorithmic environmental change functions (aECFs) are developed, which consider actual weather conditions. Synoptical scale pattern determine regions with high and low sensitivity to aircraft emissions, hence determine climate change functions. Making available algorithms which establish linkage between MET information and environmental impact is a pre-requisite for an efficient generation of ECFs. A concept relying on aECFs brings as advantage that consequently environmental assessment of aircraft
trajectories are not limited to match weather pattern which correspond to archetypical pattern, but for each synoptical situation corresponding advanced MET data products on climate and environmental impact can be generated.

Environmental performance data of European traffic sample in the case study (18 December 2015) showed that overall climate impact is composed of both CO$_2$ and non-CO$_2$ impacts, with non-negligible non-CO$_2$ effects about 5–20 times higher than CO$_2$ impacts alone, depending on climate metric calculated. For longer time horizon this ratio tends to lower values, going down from about 20 to about 6, when comparing a time horizon of 20 and 100 years, respectively.

In terms of implementation of such concept described, additional environmental MET information data products need to be made available to ATM. Hence complexity of the ATM environment due to meteorology needs to be transferred via MET information into the ATM infrastructure, making sure that ATM is having available high quality information for efficient flight planning. Within the SESAR 2020 Master Plan such information is made available system-wide via the SWIM infrastructure, where MET is one component in it, as is e.g., AIM information.

VII. Summary and Conclusion

This paper presents overall concept for a multi-criteria environmental assessment framework relying on environmental change functions (ECFs), as is currently under development in the European project ATM4E which is part of Exploratory Research within SESAR2020 research programme. Models and methods required are described, which are used to quantify environmental impacts and plan aircraft trajectories. Concept of ECFs is presented in detail and methods how to generate are described. We present a case study for a traffic sample over Europe which is applied on a candidate day using real weather conditions. Initial findings are presented using prototype ECFs. We provide an estimate for importance of non-CO$_2$ using ATR as climate metric, with a ratio of non-CO$_2$ to CO$_2$ impacts on climate between 6 and 20, for time horizon 20 and 100 years, respectively. From climate-optimisation of a single-flight trajectory, using prototype ECFs, an estimate of climate impact mitigation potential is calculated in the order of 12% and 25%, for fuel increase of 1% and 5% respectively. For LAQ we selected as environmental performance indicator in this initial case study the increase of atmospheric NO$_2$ concentrations, performing sensitivity tests for different air quality indicators, e.g., using either daily or hourly peak concentrations.

The innovative aspect in this study is to present a quantitative assessment of environmental performance indicators for a trajectory optimisation of a European traffic sample, representing a comprehensive framework for a multi-criteria environmental assessment framework, which comprises both climate impacts and local and regional environmental impacts. Such an assessment framework allows to be used for an analysis of overall environmental performance of a set of aircraft trajectories, but also an optimisation under individual objective functions and weighting factors, to support strategic decision making in the sense of a decision support system.

A novel aspect is the combination with an Earth-System model for online verification of algorithmic environmental change functions and proposed routing strategies when minimizing environmental impacts. The concept presented here relies on identification and effective implementation of aECFs in flight planning tools as advanced MET services providing a flexible interface to comprehensive calculation of environmental impacts of aviation emissions. For establishing required set of individual aECFs representing individual effects of aviation climate and environmental impact, comprehensive assessments of atmospheric and environmental impacts are required. Such assessments require suitable atmospheric chemistry and physics modelling tools being applied. They subsequently need to consider and identify key atmospheric parameters in order to
eventually provide mathematical formulation of aECFs, which can then be implemented in an expanded ATM aircraft trajectory optimisation tool.

This concept lays the basis for performing route optimisation in the European airspace using advanced MET information in the light of environmental assessment and optimisation of aircraft movements in Europe. Ultimately, this will lead to a strategic roadmap of how to implement such a multi-criteria and multi-dimensional environmental assessment and optimisation framework into current ATM infrastructure by integrating tailored MET components, in order to make future aviation sustainable.

Acknowledgment

S.M as project coordinator thanks the whole ATM4E project team for valuable discussion and feedback. The ATM4E consortium is composed of DLR, University of Reading, Manchester Metropolitan University, Envisa SAS, Technical University Hamburg and TU Delft. The ATM4E project team authoring this contribution are in particular the authors of reference [20], with additional thanks to (in alphabetical order), Katrin Dahlmann, Volker Grewe, Florian Linke, Benjamín Lührs for supporting this paper. Finally, S.M. also thanks the ATM4E External Expert Advisory Board (Chair: Herbert Pümpel, WMO) for valuable input and constructive exchange.

REFERENCES


Dahlmann, K.; Grewe, V.; Pomater, M.; Matthes, S. Quantifying the contributions of individual NO sources to the trend in ozone radiative forcing. Atmos. Environ. 2011, 45, 2860–2868.


Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. Royal Meteorol. Soc. 2011, 137, 553–597.

PROBABILITY OF SNOW NOWCASTING FOR AIRPORTS

PNOWWA (Probabilistic Nowcasting of Winter Weather for Airports)

Rudolf Kaltenboeck
Meteorology – Development and Innovation
Austro Control
Innsbruck, Austria
first.lastname@austrocontrol.at

Heikki Juntti1, Elena Saltikoff2, Harri Hohti2, Kari Österberg2
1) Weather Service 2) Meteorological Research Finnish Meteorological Institute
1) Rovaniemi 2) Helsinki, Finland
First.lastname@fmi.fi

Abstract—The PNOWWA project produces methods for the probabilistic short-term forecasting of winter weather and enable the assessment of the uncertainty in the ground part of 4D trajectories. 4D trajectory management is a necessary concept to meet future growth in air traffic. Probabilistic forecasts will be used in air traffic management (ATM) applications to support operational planning in surface management and ATM decision making, thereby increasing airport capacity, shortening delays and promoting safety.

PNOWWA demonstrates very short-term (0-3h, ”nowcast”) probabilistic winter weather forecasts in 15-minute time resolution based on an extrapolation of movement of weather radar echoes and improve predictability of changes in snowfall intensity caused by underlying terrain (such as mountains and seas). An extensive user consultation was performed to focus on user needs (parameters and thresholds) and to ensure products which are suitable to be integrated in various applications on the ATM side. A research demonstration was conducted at airports in Austria and Finland in winter 2016/2017.

PNOWWAs probabilistic forecast of winter weather for airports have been successfully demonstrated. Data were provided via webpage as ”live” data.

User feedback has been collected to improve the PNOWWA product for next year demonstration phase. The PNOWWA project helps to assess impacts on operations at airports during disruptive winter weather and illustrates strong potential for probabilistic nowcasting using weather radar data.

I. Introduction

Probabilistic forecasting is nowadays used in meteorology to quantify uncertainty. In contrast with deterministic forecasting, the natural intrinsic variability of weather and the uncertainty in the observations and in the forecast process itself are considered. Probabilistic information can be created by generating an ensemble of forecasts or by applying statistical post processing. Then the user must choose proper probability thresholds, which gives them the correct balance of alert and false alarms for specific applications. Hence, an objective quantity of uncertainty results, which means increasing risk of wrong decision with lower likelihood. These probability forecasts support best the user specific decision-making processes.
While probabilistic forecasting is more common for medium-range (order of days) from numerical weather model output and model output statistics, probabilistic nowcasting methods have been improved during last years, e.g. [1], [2]. Nowcasting (0-3 hour) of precipitation systems is strongly driven by extrapolation of weather radar images, because of the rapid updated high spatial resolution observations, in three dimensions up to ranges of 2-300 km. There, linear translation of a precipitation area is better captured compared with the propagation part. In general, the quality of the forecast decreases with lead time and increasing spatial scale (scale dependent life time of precipitation). As quantitative methods are needed, sectors of increasing catchment areas or decomposition in different scales of snow fall patterns with different behavior are used to create an ensemble of nowcasts from consecutive radar images [2].

The PNOWWA produces methods for the probabilistic short-term forecasting of winter weather and enable the assessment of the uncertainty from the end points (airports) of 4D trajectories. 4D trajectory management, also sometimes called “Gate to Gate concept” is an essential building block of the ICAO and SES concepts [GANP, ATM Master Plan] to meet future growth in air traffic; probabilistic forecasts will be used in ATM applications to support operational planning in surface management and ATM decision making, thereby increasing airport capacity in critical weather situations, shortening delays and promoting safety.

In PNOWWA demonstration campaign very short-term (0-3h, “Now-cast”) probabilistic winter weather forecasts at 15min time resolution based on the extrapolation of the movement of weather radar echoes were delivered to a selected group of end users at different airports. Users were consulted to the most relevant parameters and operationally important thresholds of the selected parameters (e.g. how many centimeters is considered “heavy snowfall”).

II. Mapping the user needs

A. Selection of representative users

User Needs were sought to be obtained from a wide range of aviation stakeholders mainly at airports, ranging from major hubs to smaller regional European airports. These were selected to represent different (and challenging) topographic regions, ranging from Nordic maritime to high Alpine environments to determine the limits of applicability as well as the capabilities of the proposed Now-casting system. Apart from web-based surveys, direct contact was established to a number of representatives of user groups and their views and operational concepts established and compared, leading to the interesting result that any such Now-casting system will have to be highly flexible, scalable and adaptable to meet genuinely diverse user needs. The relevant thresholds or equivalent decision criteria were discussed in face-to-face meetings with different end users at Vienna (LOWW), Innsbruck (LOWI), Zurich (LSZH), Geneva (LSGG), Rovaniemi (EFR0) and Helsinki Vantaa (EFHK) airports. Written feedback of varying detail was received from Oslo-Gardermoen, Munich, Istanbul, and Salzburg.

B. The different needs as expressed by users

Three major groups of users were identified. The runway maintenance needed accumulation of snow in millimeters during each 15 minute step. Thresholds were expressed separately for dry snow, wet snow and slush. In addition, they wanted a probability for freezing rain – something, what a solely weather radar – based algorithm can not express.

The aviation control tower wanted probability of low visibility procedures, LVP. In winter, LVP is related to clouds, fog or snowfall, and solely weather radar – based algorithm can only express the snowfall-related LVP [visibility reduction without ceiling].
The deicing managers at airports used its own Deicing-weather index (DIW) PNOWWA team had experimented with this already in SESAR1. Basic idea of DIW is that the bigger DIW value is the longer time is needed for de-icing of individual aircraft. Thresholds of frost formation causing need of de-icing of planes is based on the experiences of de-icing companies at Helsinki, Oslo and Stockholm airports for conditions when planes will ask for de-icing (interviews during SESAR1 projects 11.02.02. and 06.06.02). The need of individual plane’s de-icing is dependent also from the previous phases of flight and conditions it has experienced in past not only meteorological conditions. [That is why the probabilistic approach is more suitable for user purposes than deterministic.]

<table>
<thead>
<tr>
<th>Weather</th>
<th>Effect on aircraft</th>
<th>DIW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy snow or sleet (visibility &lt; 1500 m)</td>
<td>Snow on plane</td>
<td>3</td>
</tr>
<tr>
<td>Freezing rain/drizzle</td>
<td>Ice on plane</td>
<td>3</td>
</tr>
<tr>
<td>Light or moderate snow or sleet</td>
<td>Snow on plane</td>
<td>2</td>
</tr>
<tr>
<td>Temperature -3…+1°C and humidity &gt; 75%</td>
<td>Frost formation on the plane surface.</td>
<td>1</td>
</tr>
<tr>
<td>Any other weather</td>
<td>No remarkable contamination on plane</td>
<td>0</td>
</tr>
</tbody>
</table>

An extensive user consultation was performed to focus on user needs such as parameters and thresholds (Figure 1) and to ensure products which are suitable to be integrated in various applications on the ATM side.

**FIGURE 1 — Relative number of responses which show the type of winter weather affecting airport operation, and requires early mitigating actions. Total number of respondents is 25.**

Beside the nowcasting lead time of 3 hours (Figure 2), airport operators are interested additionally also in 12 hours, and more dominantly, in 24 hours lead times for tactical planning and pre-emptive actions. In short range forecasting, exact timing is essential, because wrong timing of the adverse weather event might significantly disturb operations planning and subsequently generate substantial delays for air traffic. Respondent from ATM stated that the needed forecast time is also depending on the flight time to another European destination, this means for capacity planning in the time range around 3 hours.
III. Nowcasting tools for demonstration

A. General structure of the tools

The structure of PNOWWA “demonstration engine” is modular, to allow the use of different components at their most suitable maturity level independent of progress in the other work packages.

Then conversion equations from literature and PNOWWA studies were used to express the user-defined parameters in radar reflectivity dBZ (summary see in [2]). The results, as used in the first scientific demo, are shown in tables 2-4.

The extrapolation software was used to create class probabilities of these thresholds expressed in radar reflectivity (dBZ). These answer e.g. the question “what is the probability, that radar echoes seen at this airport will be between 24.5 and 29 dBZ 30-45 minutes from now”.

Then using the conversion tables again, the probabilities were converted to “user parameters”. The resulting probabilities answer e.g. the question “what is the probability, that at this airport 5-10 mm cm snow will be accumulated 30-45 minutes from now”.

Then, time-series were created and expressed as tables using a web-based interface.

B. Nowcasting methods

Three nowcasting methods have been tested in PNOWWA [2].

In the method suggested by Andersson and Ivarsson [3], wind at 850 hPa level is used to describe the movement. The wind is taken from HIRLAM (High Resolution Limited Area Model) numerical weather prediction model. The uncertainty of forecast is growing with time, related to the snow field texture. This approach had been tested in SESAR1, so it was known to provide reasonable results. It was used in the first real-time demonstration campaign.

The new nowcasting method developed in PNOWWA uses motion vector analysis schema based on approach of optical flow [4] and the stochastic ensembles for creating probabilistic output [5]. This method was used for case studies and offline demonstrations [2].

The method operationally used and originally developed at FMI, applies modified correlation-based atmospheric motion vector (AMV) system by EUMETSAT [6], [7]. This well tested method was used as a reference for comparisons, when developing the new method.
C. Data sources

Radar data from Finland is part of FMIs operational data flow. Radar data from Austria comes from Austro Control. Radar data from other parts of Europe, needed for calculation of motion vectors, is coming from EUMETNET OPERA [8].

For temperature and dewpoint, METAR observations are used. Other model parameters, needed for parameters not available from radar data, as well as for wind vectors for Andersson method, the HIRLAM numerical weather prediction model is used [9].

D. Conversion tables

Simple conversion tables were used to express the dBZ values in user-defined parameters (‘what is the probability, that at this airport more than 10 mm snow will be accumulated 30-45 minutes from now’). Tables 2-5 show thresholds as used in the first scientific demo.

To select the right dBZ thresholds, type of snow had to be determined. ICAO has defined the types of snow as follows [10]

- Dry snow – can be blown if loose or compacted by hand, will fall apart again upon release.
- Wet snow – can be compacted by hand and will stick together and tend to form a snowball.
- Compacted snow – can be compressed into a solid mass that resists further compression and will hold together, or break up into lumps, if picked up.

For this application, the snow type was determined based on temperature and dewpoint, read from the METAR.

| TABLE II — Temperature and Dewpoint for discrimination between dry and wet snow |
|----------------------------------|-----------------|-----------------|
| Temperature in °C                | dry snow        | wet snow        |
| ≤ -0                             | ≤ -0 °C and ≤ +3 °C |
| Dewpoint in °C                   | ≤ -1            | ≤ 0             |

| TABLE III — Radar reflectivity |
|-------------------------------|-----------------|-----------------|
| Visibility m                  | dBZ for dry snow | dBZ for wet snow |
| <=600                          | >29.0           | >29.0           |
| 600-1500                       | 24.5-29.0       | 23.5-29.0       |
| 1500-3000                      | 15.5-24.5       | 19.5-23.5       |
| >3000                          | <15.5           | <19.5           |

| TABLE IV — The dependency between liquid water equivalent and radar reflectivity |
|---------------------------------|-----------------|-----------------|
| Liquid water equivalent mm/h    | dBZ for dry snow | dBZ for wet snow |
| >=4                             | >29.0           | >29.0           |
| 2-4                             | 24.5-29.0       | 23.5-29.0       |
| 0.4-2                           | 15.5-24.5       | 19.5-23.5       |
| <0.4                            | <15.5           | <19.5           |

| TABLE V — The dependency between snow accumulation and radar reflectivity |
|-------------------------------|-----------------|-----------------|
| Snow accumulation mm/15 min   | dBZ for dry snow | dBZ for wet snow |
| >10                            | >29.0           | >29.0           |
| 5-10                           | 24.5-29.0       | 23.5-29.0       |
| 1-5                            | 15.5-24.5       | 19.5-23.5       |
| <1                             | <15.5           | <19.5           |
### IV. End user display

Example of end user display (web page) at one of the participating airports ([Figure 3](#)). Blue bars limit sections for different user groups (runway maintenance, de-icing agents, tower). Horizontal axis is time in minutes since the forecast was issued. Vertical axes are severity classes of each phenomena or index, e.g. intensity of dry or wet snowfall, deicing weather index and visibility. The colourful boxes then depict the probability of each class at each moment, largest probabilities coloured in red and yellow. In layman terms: "it is going to snow for 45 minutes more, then it’s dry for at least half an hour, probably even longer, but after 2 hours the probability of snowfall is increasing again."

### V. Experiences of the first demonstration campaign

First scientific demonstration of PNOWWA conducted for Austrian and Finland airports during February and March 2017. There were only limited amount of real winter weather cases in Austria and Southern Finland. In Northern Finland weather was more favorable. In spite of that it was recognized that prototype worked well and it was flexible to tailor it for different users. We were able to collect valuable and positive feedback from users which further helps to assess the applicability of probabilistic nowcasting for disruptive winter weather using weather radar data.

Reference time, automatic update of web page was felt to be necessary character of product. Accumulation of snow expressed as mm/15 min scale as wished by users. Product description and feedback form was included in the webpage, where online feedback form never used by stakeholders. Therefore, individual contact to different stakeholders worked best. Some users felt more comfortable to use traditional material than new product. It would be beneficial to give hands on familiarization to test users during some real winter weather case. That would give us more information about the level of quality of demo product and improvements, that could be done.

Users should also be well informed about the possible limitations of product. In PNOWWA prototype forecasted amount of decrease of visibility caused by snow, only. Mist or fog forecasts were not included. Users were confused with that and in operative service it should be taken into account all type of effects causing reduction of visibility. Also ATM stated the need for ceiling information in nowcasting decision support system. In current PNOWWA demonstrator forecasting of ceiling is not possible from extrapolated weather radar information. It is not enough to develop ways to produce probabilistic weather forecasts, but it is also necessary develop ways how probabilistic weather information could be used efficiently in ATM processes. Cost loss ratios and suitable, impact-related key performance indicators, which combine traffic load, delays, amount of chemicals and workload have further to be developed. At the moment airport operation is always on the safe side. Therefore, stakeholders are concerned about events with low likelihood which leads to no action but results e.g. in snow fall and resulting possible incidents. By the other hand, runway operation stated possible over-interpretation of low probability winter events which might increase the costs. Subsequent, preparation workshops before next demonstration phase will be organized to train users in interpretation of the PNOWWA product.

Feedback from individual discussions and demonstration of high impact “offline” case studies will help to improve the probabilistic application for next winter demonstration campaign.
**FIGURE 3 — Example of end user online web page from 22nd February 2017, issued at 15:15 UTC. Different forecast classes (left in grey) for 3 stakeholder groups are predicted up to 195 min, where likelihoods are color coded (green 0–20 %, yellow 30–50 %, red 60–100 %).**

<table>
<thead>
<tr>
<th>RUNWAY MAINTENANCE (UPDATED 2017-02-22 15:16:00 UTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>accumulation% day snow, mm/h</strong></td>
</tr>
<tr>
<td>0-15 min</td>
</tr>
<tr>
<td>over 10 mm</td>
</tr>
<tr>
<td>5-9 mm</td>
</tr>
<tr>
<td>1.5 mm</td>
</tr>
<tr>
<td>less than 1 mm</td>
</tr>
<tr>
<td><strong>accumulation% wet snow, mm/h</strong></td>
</tr>
<tr>
<td>0-15 min</td>
</tr>
<tr>
<td>over 5 mm</td>
</tr>
<tr>
<td>3.5 mm</td>
</tr>
<tr>
<td>1.2 mm</td>
</tr>
<tr>
<td>less than 1 mm</td>
</tr>
<tr>
<td><strong>prob of freezing rain</strong></td>
</tr>
<tr>
<td>0-15 min</td>
</tr>
<tr>
<td>prob</td>
</tr>
<tr>
<td><strong>prob of freezing wet runway</strong></td>
</tr>
<tr>
<td>0-15 min</td>
</tr>
<tr>
<td>prob</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DE-ICING AGENTS (UPDATED 2017-02-22 15:15:00 UTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DIW class %</strong></td>
</tr>
<tr>
<td>0-15 min</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td><strong>prob of freezing wet runway</strong></td>
</tr>
<tr>
<td>0-15 min</td>
</tr>
<tr>
<td>prob</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TOWER (UPDATED 2017-02-22 15:15:00 UTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VIS decreased by snow</strong></td>
</tr>
<tr>
<td>0-15 min</td>
</tr>
<tr>
<td>VIs less than 600 m</td>
</tr>
<tr>
<td>VIs 600-1500 m</td>
</tr>
<tr>
<td>VIs 1500-3600 m</td>
</tr>
<tr>
<td>VIs over 3600 m</td>
</tr>
</tbody>
</table>
VI. Verification of the first demonstration campaign

To demonstrate the reliability and applicability of PNOWWA product in air service provision, we have to validate and verify results and show positive impacts but also limitations.

Hence, for verification we focus on last year winter events during first demonstration phase. Different weather pattern and different location (Central Europe, Northern Europe, mountains and sea influences as well as flat areas) have been investigated.

Large synoptic systems like frontal band of snow (Figure 4 left) might persist over several hours and therefore, higher probabilities in larger lead times (e.g. 120 minute) occurred in contrast to small scales of snow showers, which have a typical life time of about 60 minutes.

Figure 4 — Different types of weather systems: frontal band of snow (left) and snow showers (right) over southern Finland.

Example of probabilistic forecast performance is given as time series for two different forecast lengths for Innsbruck. In Figure 5 observations are assumed as 15 min forecast, where probability is larger/less than 50 % means snow/dry. Note, that in mountainous areas weather radar coverage is reduced due to shielding effects and high situated radar sites at mountain tops accompanied by missing snow below radar horizon (e.g. in valleys).

For short term forecasts in Figure 5 and Figure 6 periods of high probability of snow correlates well with snowfall amount. Dry periods are predicted well on the left half in Figure 5, while at the end in the right half of Figure 5 weak probabilities < 0.2 are present. For large scale precipitation events even longer lead times correlate well with snowfall (accompanied by high probabilities) and dry areas.

Figure 5 — Time series of probability for snowfall gathered from radar extrapolation for lead time of 30 minutes (green +) compared with observations (15 min forecast, red line). X axis length is 48 hours and every data point reflects 15 min time step.

Figure 6 — As Figure 5 but for lead time of 120 minutes.
For more quantitative verification results, the reader is referred to paper by Pulkkinen et al. [2].

Investigations of aircraft delay minutes didn’t correlate well with snow height accumulation at Vienna airport for winter 2016/2017.

VII. Future work

A second real-time demonstration campaign will be organized next winter, making use of the more accurate nowcasting methods developed in PNOWWA and helps for further development of impact based key performance indicators.

Before that, discussions with end users are continued, and change to concept of “exceedance probabilities” instead of “class probabilities” is introduced. At first glance, it would feel natural to forecast, that “is the intensity of snowfall between a and b”, and this is what the users asked for. Our experience shows, that forecasting “probability that intensity of snowfall is at least a” is more useful.

Additional winter weather forcing due to topographic influences was investigated. This forcing can strongly affect the weather radar extrapolation technique for nowcasting. Results show, that the forecast quality is lower for precipitation systems arriving from the sea, and north of the Alps frontal delays and upslope enhancement have been observed [2].

In this project, we have focused on radar-based methods due to their outstanding temporal resolution. In the possible follow-up projects, data fusion with other data sources such as numerical weather prediction should be considered, both to extend the valid time and widen the available weather parameters. It should be underlined, that this is S2020 exploratory research, so real-life user applications, such as mobile apps, are not within the scope of this project.

Acknowledgment

The authors would like to thank all the professionals at participating airports for their time in defining the “user needs” and providing valuable feedback.

The authors gratefully acknowledge the EU grant #699221 provided by the Single European Sky Air Traffic Management Research (SESAR) Joint Undertaking (SJU) in cooperation with the EU/H2020 research program.

Additional thanks are extended to the three anonymous reviewers for their valuable comments, which improved the quality of the paper.

REFERENCES


Modelling and simulation
A NOVEL FRAMEWORK TO ASSESS THE WAKE VORTEX HAZARDS RISK SUPPORTED BY AIRCRAFT IN EN-ROUTE OPERATIONS

Marc Melgosa and Xavier Prats
Department of Physics - Aerospace Division
Technical University of Catalonia
Castelldefels (Barcelona), Spain

Sergio Ruiz, Joan Tarragó and Javier Busto
GTD Sistemas de Información
Barcelona, Spain

Meiko Steen
Institut fuer Flugfuehrung
Technische Universitaet Braunschweig
Braunschweig, Germany

Abstract—This paper presents the simulation environment developed within the framework of R-WAKE project, funded by SESAR 2020 Exploratory Research. This project aims to investigate the risks and hazards of potential wake vortex encounters in the en-route airspace, under current and futuristic operational scenarios, in order to support new separation standards aimed at increasing airspace capacity. The R-WAKE simulation environment integrates different components developed by different partners of the R-WAKE consortium, including simulators for weather, traffic, wake vortex phenomena, wake vortex interactions and different tools and methodologies for safety and risk assessment. A preliminary example is presented in this paper, in which 200 historical trajectories were simulated to show that the novel framework works properly. A WVE encounter has been detected in such first scenario, however with no significant safety effect on the follower aircraft. A second controlled scenario has been then run to force the detection of a severe wake encounter under realistic en-route conditions. Such scenario has given evidences that confirm the safety relevance of the underlying research concept.

I. Introduction

Wake vortex issues in terminal maneuvering areas [TMA], especially in the final approach and initial take-off segments, are well known and have received a particular attention in the last decades [1], [2], [3]. In the en-route phase, however, wake vortex encounters are unlikely and so far, are still considered rare events (although...
few significant accidents have occurred in the recent years, such as the encounter reported in [4]). Current knowledge on wake vortex encounters (WVE) hazards, and the corresponding separation standards, is strongly based on studies and data collection at low altitudes, mostly at the vicinity of airports. Few projects have tried to investigate WVE at typical cruise altitudes and most of the research is based in simulation, since data collection at high altitudes (above FL200) is a difficult task. Recent research has shown that current separation standards might not be enough for protecting aircraft against WVE hazards, while in other cases they might be over-conservative [5], [6], [7], [8], [9]. Hazardous WVE en-route might become a serious issue in the near future if we take into account, for instance, a) forecast for higher volumes of traffic in certain areas; b) a more heterogeneous and diverse traffic fleet (different aircraft sizes and performance, introduction of unmanned aerial vehicles, etc.); c) new concepts of operation, in line with SESAR 2020 paradigm [10]; d) more accurate navigation systems (reducing the dispersion of flight tracks); and d) new (or refined) standards leading to reduced separation minima between two aircraft. For these reasons, there is an increasing interest in the air traffic management (ATM) community to assess potential issues related to wake vortex phenomena when refining or proposing new en-route separation standards, aiming to increase airspace capacity. R-WAKE is a SESAR 2020 Exploratory Research project performing an initial risk assessment, with regards to the wake vortex phenomena, of a hypothetical introduction of lower tactical separation standards en-route. For this purpose, a simulation environment is developed within the project activities to perform a safety and robustness analysis and a standards development methodology, thoroughly consistent with reference methodology taken from EUROCONTROL and SESAR [11]. This paper presents the simulation environment developed in the R-WAKE project, which embeds several components with the aim to: a) synthesise trajectories at European scale, according to different concepts of operations and implementing different air traffic control (ATC) tactical separation criteria; b) simulate accurate wake vortex phenomena; c) simulate realistic weather conditions, affecting both the traffic patterns and the behaviour of the vortices; d) detect possible WVE in the simulated scenario; and finally, e) assess the severity of the WVE detected.

II. The R-WAKE Project

The R-WAKE Concept [12] is linked to the following research question: ‘What Separation Minima Reductions can be applied in specific and clearly defined operational conditions keeping the current safety level related to En-Route WVE hazards?’ In order to support the generation of validation evidences and to illustrate the ‘R-WAKE concept’, the ‘R-WAKE system’ has been developed, which is composed of a tailored safety and robustness research methodology and a simulation platform to reproduce a given traffic scenario and the generated wake vortices, and to approximate in a quantitative way the level of risk of the potential WVE hazards (dynamic risk modelling approach). The aim of this project architecture is the achievement of five main outcomes, referred in the Fig. 1 as O1 to O5.
The research goals of the project, i.e., proposal of potential enhancements in the current separation standards to protect against WVE hazards (refer to [13] for details), are approached by the study of a safety and robustness analysis on the three basic components of WVE hazard risks: the severity, the potential frequency, and the level of risk after applying the ATM risk mitigation measures. The methodological approach of the research can be synthesized in the following five activities:

1. To assess and quantify the level of severity of different WVE situations and establish the level of acceptance for each WVE hazard risk.

2. To quantify the probability of finding potentially hazardous WVE in today’s traffic conditions.

3. To analyse how much the level of risk is mitigated due to the ATC separation provision (applied to mitigate risk of collision between aircraft rather than for WVE avoidance).

4. To assess the potential impact on safety after introducing future SESAR concepts of operations (in line with TBO/PBO) and approximate guesses for future traffic demand.

5. To propose a new (possibly dynamic) separation standard, along with mitigation methods to be applied during either the flight execution (tactical separation management) or the 4D trajectory planning (including strategic separation management in a TBO/PBO context).

In order to perform the methodology and the project outcomes, the R-WAKE framework is divided in two steps:

- The Step 1, or ‘micro-analysis’, aims at providing a wake vortex safety baseline in form of a severity matrix and a tolerability matrix. Such outputs of the micro analysis will be used as an input for the simulator system and for the macro model analysis. For this reason, the micro model analysis has to be executed during the implementation phase and before starting the macro model analysis.

- The Step 2, or ‘macro-analysis’, in which current and future traffic situations will be simulated in order to determine if the separation standards are enough to ensure a safety operation of the airspace.

The above research methodology of the R-WAKE project and the relationship between the Step 1 and Step 2 (i.e., micro and macro analysis) can be found in Fig. 2.
III. R-WAKE Micro Analysis Framework

The goal of the micro-scale simulations (or R-WAKE Step 1) is to generate the wake vortex safety baseline. It means that the severity and tolerability matrix used as inputs in the macro model analysis will be the outputs of this step. Given an aircraft pair, a geometry of the crossing, a separation distance (in the horizontal and/or vertical domain) between follower and generator aircraft, and a given set of contextual scenario variables (such as altitude and speed of both aircraft, etc.), the micro-scale simulation will compute the severity of the encounter on the follower aircraft. For this purpose, this simulation is divided in three major phases:

1. Computation of the vortex circulation, generated by the generator aircraft and encountered by the follower aircraft.
2. Computation of the aircraft upset experimented by the follower flight due to the vortex encounter.
3. Assessment of the severity of the upset, based on expert knowledge.

IV. R-WAKE Macro Analysis Framework

Once the potential WVE hazards and the severity of their potential consequences is well-understood, the purpose of the Step 2 research approach of the project is to assess the level of risk for each of the identified hazard categories that may be present in the European ATM context. Different frequency/risk analyses will be performed with the Step 2 R-WAKE
simulation framework under the consideration of different ECAC-wide traffic demand patterns and different ATM mitigation measures applied. The main workflow of the R-Wake system [14] for macro analysis is described below by functionality (different functionalities might be provided by a same software tool used in R-WAKE project) and showed in Fig. 3.

**FIGURE 3 — R-WAKE framework [14].**

1. **Weather Simulator (WXS):** The WXS provides historic weather data to the Traffic Simulator (TRS) and to the Wake Vortex Simulator (WVS), in order to have realistic weather conditions during the trajectory and wake vortex simulations in European ECAC airspace, and to perform statistical simulation based studies to obtain results that are statistically significant during the hazards risk evaluation process. The proposed Weather Simulator concept is based on the background system SIMET, a realistic simulator of atmospheric conditions developed for the evaluation of new generation Flight Management Systems that take into account weather conditions for trajectory optimization.

2. **Traffic Simulator – Traffic and Trajectory Planner (TRS.TTP):** Generates and simulates traffic scenarios based on real or future traffic demand and considering weather data fed by the weather simulator. The output trajectories feed the traffic planner with realistic trajectories. The traffic planner will apply the corresponding ATM constraints according to the concept of operations modelled (current ATM or SESAR 2020+) and the ATM layers activated (airspace, ATFM, Extended ATC Planner, ATC or none). As mentioned before, the Traffic and Trajectory Planner (TTP) module consists of two components working together somehow act as a kind of ATM simulator: 1) the trajectory planner submodule, which can be understood as the component that simulates the airspace users (AUs) and generate the traffic demand in form of 4D trajectories subject to the existing ATM constraints, and 2) the ATM model (or traffic planner), which will be useful to represent the basic ATM mitigation/separation layers in charge of ensuring the required safety performance in the ECAC sky.

3. **Traffic Simulator – Wake Encounter Region Finder (TRS.WERF):** This sub-module identifies regions of airspace (volumes) in which potential wake vortex encounters could occur. Since the simulation of precise wake vortices for all the ECAC-wide flight trajectories requires a high computational burden, the simulation of WV will be limited to those regions that have some likeliness of hazardous WV encounter therein. Therefore, this module acts as a filter to reduce computational burden and the output will feed to the Wake Vortex Simulator and Traffic Planner modules with the regions of risk and with the segments of flight trajectories crossing such hazardous regions.

4. **Wake Vortex Simulator (WVS):** This module simulates realistic wake vortexes given the flight parameters of each trajectory (aircraft mass, speed, path, etc.) and the weather for the airspace region of interest. As an output for feeding the WEPS system, this module will generate a simplified macro-model of the vortexes in which the stochastic behaviour of the vortex (position, size and strength) can be represented as a 4D tube. Such 4D tube will be modelled to still capture all the relevant information for an effective wake vortex prediction process.
5. **Traffic Simulator – WV Encounter Prediction System (TRS.WEPS):** This sub-module receives the discrete model of the 4D tubes from the WV simulator and the trajectory segments from the WERF system and then crosses all the information to perform a probabilistic analysis and predict potential encounters. If an encounter is detected, the system will obtain the expected strength of the vortex and assess the severity of the vortex in relation to the parameters of the affected flight (aircraft, speed, geometry of the encounter, etc.) and other contextual conditions (e.g., surrounding traffic, excess thrust, etc.), as identified in the Step 1 of the R-WAKE research approach. The event and corresponding hazard severity will be recorded for the safety analysis post-process.

6. **Safety & Robustness Analysis (SRA for Step 2):** This module represents a process rather than a simulator. A risk analysis will be performed with the inputs coming from the other modules, and new knowledge will be generated from the different scenario simulations. The insights obtained will be used to report and refine the tool and next scenarios. As part of the knowledge generated will be an evidence-based proposal of new separation standards and methods.

### V. Integration Test

In order to check the framework, an initial test scenario with 200 flights was carried out. The first module TTP was configured with the initial conditions showed in Table I. The airspace and traffic historical data needed by the TTP was obtained from the Demand Data Repository (DDR2) provided by Eurocontrol. The Traffic and Trajectory Planner generated 200 trajectories. The Wake Encounter Region Finder was used to detect aircraft pairs that were close enough (less than 10 NM) to have a potential wake vortex encounter. The same module identified which aircraft were the follower and the generator, respectively. In this scenario, 13 potential encounters were identified.

**TABLE I — TTP initial conditions**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of flights</td>
<td>200</td>
</tr>
<tr>
<td>Aircraft type</td>
<td>A320</td>
</tr>
<tr>
<td>Crossing area</td>
<td>France</td>
</tr>
<tr>
<td>Date</td>
<td>28/07/2016</td>
</tr>
<tr>
<td>Weather</td>
<td>Only vertical</td>
</tr>
<tr>
<td>ConOps</td>
<td>Structured route</td>
</tr>
<tr>
<td>Horizontal Separation Standard</td>
<td>5 NM</td>
</tr>
<tr>
<td>Vertical Separation Standard</td>
<td>1000 ft</td>
</tr>
</tbody>
</table>

After this process, the Wake Vortex Simulator used the information from the generator aircraft to simulate the wake vortex of all these flights. WEPS was used to find which of these potential encounters could be considered actual encounters. It was found that only one follower flight was in the actual influence area of the generator’s wake vortex, therefore being a wake vortex encounter susceptible of causing a potential hazard. Then, this encounter was simulated in order to calculate the upsets suffered by the follower due to the encounter. To determine the severity of the previous encounter, the maximum absolute values of altitude change, bank angle, rate of climb/descent and airspeed change are considered. The values of this encounter are summarized in the Table II.

**TABLE II — Maximum upsets obtained**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude change</td>
<td>0.0036741 m</td>
</tr>
<tr>
<td>Bank angle</td>
<td>2.9445e-05 rad</td>
</tr>
<tr>
<td>Rate of Climb/Descent</td>
<td>0.0022483 m/s</td>
</tr>
<tr>
<td>Airspeed change</td>
<td>0.00049465 m/s</td>
</tr>
</tbody>
</table>
Using the severity matrix based on expert knowledge, these upset values can be categorized as severity level 1, which means that ‘No significant safety effect’ was found in such particular encounter. In order to show the potential hazard of the wake vortex encounters in the en-route phase, a new controlled scenario was configured with a climbing aircraft (generator) and a leveled one (follower), both separated 9.12 NM and 1000 ft. as shown in Fig. 4.

**FIGURE 4 — Relative position the new trajectory.**

The WV Encounter Prediction System was used in order to calculate how the wake vortex affects to the follower aircraft. Fig. 5(a) shows the lateral (\(\Delta y\)) and vertical (\(\Delta z\)) changes of the follower aircraft with respect to the nominal trajectory as well as the velocity changes. After 10 s, the aircraft descended 40 m and was deviated 65.48 m. No significant changes in the airspeed were observed.

**FIGURE 5 — Results of the controlled scenario**

---

**Relative Position**

<table>
<thead>
<tr>
<th>Time [s]</th>
<th>(\Delta x_{\text{max}})</th>
<th>(\Delta y_{\text{max}})</th>
<th>(\Delta z_{\text{max}})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.05 m</td>
<td>-65.48 m</td>
<td>40.00 m</td>
</tr>
<tr>
<td></td>
<td>-0.01 m</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Flight Speed**

- Nominal velocity: 381.99 kts
- Actual velocity: 440.97 kts

**Attitude**

- \(\alpha_{\text{max}}\): 5.66°
- \(\beta_{\text{max}}\): -3.19°
- \(\gamma_{\text{max}}\): 120.8°
- \(\psi_{\text{max}}\): 100.9°

**Rotational Acceleration**

- \(\alpha_{\text{max}}\): 25.31 m/s²
- \(\beta_{\text{max}}\): -6.89 m/s²
- \(\gamma_{\text{max}}\): 4.47 m/s²
- \(\psi_{\text{max}}\): -6.41 m/s²

**Linear and Rotational Acceleration Changes**

- \(\alpha_{\text{max}}\): 0.71 m/s²
- \(\beta_{\text{max}}\): 0.42 m/s²
- \(\gamma_{\text{max}}\): 9.13 m/s²
- \(\psi_{\text{max}}\): -9.50 m/s²
- \(\alpha_{\text{max}}\): 4.69 m/s²
- \(\beta_{\text{max}}\): -5.76 m/s²

---

*Note: \(\Delta\) represents the change value.*
The attitude change, corresponding to pitch angle ($\theta$), bank angle ($\phi$) and yaw angle ($\psi$), and the rotational velocity change, represented by $p$ (x-axis), $q$ (y-axis) and $r$ (z-axis), is showed in Fig. 5(b). Important changes in the bank angle were found. A maximum turn of 38.20° was achieved. In addition, limited changes in the pitch and relevant changes in the yaw angle were identified. Furthermore, the follower aircraft underwent major changes in the x-component of the rotational velocity. The transitional acceleration behavior can be described as an oscillation in the lateral acceleration. On the other hand, the rotational acceleration response shows an important value in the roll acceleration. Both transitional and rotational accelerations changes can be found in Fig. 5(c). With respect to the angle of attack, sideslip angle, flight path angle and flight path azimuth angle changes, Fig. 5(d) shows how this variable changes due to the wake vortex. The angle of attack ($\alpha$) is slightly affected during the first 3 seconds after the encounter. However, a dynamic oscillation appeared in the Sideslip angle ($\beta$). The flight path angle ($\gamma$) changed to negative because the aircraft started to descend and the flight azimuth angle ($\chi$) changed significantly because the heading changed as well. The maximum upset values of this encounter are summarized in the Table III. As shown, a change in the bank angle was the main effect (characteristic effect in case of coaxial encounters). Such encounter was categorized as having a potential “Major safety consequence”, therefore being an actual severe hazard.

The results of this scenario in which a WVE has been forced, while the two aircraft were respecting the current standard separation, show that severe hazards due to wake vortex encounters cannot be discarded in the en-route operations, and therefore the topic deserves further attention.

### Table III — Maximum upsets obtained in the coaxial encounter

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude change</td>
<td>40 m</td>
</tr>
<tr>
<td>Bank angle</td>
<td>0.66675 rad</td>
</tr>
<tr>
<td>Rate of Climb/Descent</td>
<td>12.617 m/s</td>
</tr>
<tr>
<td>Airspeed change</td>
<td>1.9492 m/s</td>
</tr>
</tbody>
</table>

### VI. Conclusion

The R-WAKE project (a SESAR 2020 Exploratory Research project) proposes an advanced simulation framework to assess the risks and hazards of potential wake vortex encounters in the en-route airspace. This can support the definition of new separation standards and the generation of evidences for the corresponding safety case. An initial integration test for this framework has been presented in this paper. A first set of 200 historical flights was simulated applying the current separation standard defined by a horizontal separation of 5NM and vertical separation of 1000 ft. Such first test has been useful as a validation exercise of the R-WAKE framework, showing that the macro-scale framework is ready to be used and all its modules are working well together. No significant wake encounters have been found from the point of view of safety, possibly due to the fact that the traffic sample is still not fully representative of the entire traffic demand patterns in the ECAC area. A second scenario in which a wake encounter was forced with the aircraft separated horizontally and vertically 9.12 NM and 1000 ft, respectively, has shown that severe encounters with major consequences for either the crew, the aircraft, or both, can actually happen in the en-route environment. Future work will include simulations traffic data sets that are more representative of the actual ECAC demand, and the hazard risk will be explored and benchmarked with the application of different separation standards, to analyze their effect in the safety performance of the entire European ATM. A new separation standard will be defined and proposed to reduce over-conservative separations and to protect better the flights in some cases, if it is found necessary.
REFERENCES


ASSESSING ATM PERFORMANCE WITH SIMULATION AND OPTIMISATION TOOLS

The APACHE Project

Xavier Prats and Cristina Barrado
Physics Department - Aerospace Division and Computer Architecture Department
Technical University of Catalonia (UPC)
Castelldefels, Barcelona (Spain)

Andrija Vidosavljevic and Daniel Delahaye
École Nationale de l’Aviation Civile (ENAC)
Toulouse (France)

Fedja Netjasov and Dusan Crnogorac
Faculty of Transport and Traffic Engineering (FTTE)
University of Belgrade
Belgrade (Serbia)

Abstract—This paper describes the objectives and methodology of the APACHE project, a SESAR Exploratory Research project proposing a new framework to assess European air traffic management (ATM) performance. This framework integrates an ATM simulator prototype used to synthesise scenarios for preops performance assessment, but also needed to compute some novel performance indicators, which require from optimisation or simulation capabilities. This simulator embeds a trajectory planner; an airspace planner; a traffic and capacity planner; and finally, a performance analyser module. An illustrative example is given, showing the successful integration of all these modules, where an initial performance assessment is done for a realistic data set of 24h of traffic over the FABEC airspace.

I. Introduction

At present, the European air traffic management (ATM) is evolving in a coordinated manner aiming to improve the overall efficiency of air navigation services across several key performance areas (KPAs). The International Civil Aviation Organization (ICAO) launched in 2003 a worldwide initiative to ensure that the future global ATM system is performance based [1], [2]. Worldwide support to the ICAO initiative is also given by CANSO (Civil Air Navigation Services Organisation) [3]. In line with these initiatives, current ATM performance assessment is addressed in Europe through the Single European Sky (SES) Performance Scheme, which establishes an agreed methodological framework for performance targeting, measuring, baselining and benchmarking in ATM [4].

The work presented in this paper was partially funded by the SESAR Joint Undertaking under grant agreement No 699338, as part of the European Unions Horizon 2020 research and innovation programme: APACHE project (http://apache-sesar.barcelonatech-upc.eu/en). The opinions expressed herein reflect the authors view only. Under no circumstances shall the SESAR Joint Undertaking be responsible for any use that may be made of the information contained herein.
In [5] a comprehensive review is given, comparing the performance frameworks proposed by ICAO, CANSO, the SES performance review unit (PRU) and SESAR 2020; identifying over 150 PIs for performance management/monitoring in 11 different KPAs. Similarly, in NextGen (the North American ATM modernization programme counterpart to SESAR) numerous PIs have also been proposed to measure the performance of the programme deployment[6], [7].

Despite the evident lack of harmonisation, some of the PIs currently in place show some important limitations, mainly due to the lack of availability or quality of the input data required; or because the implementation of too simple models in the PI computations. In many occasions performance is assessed by using proxy indicators, which in some cases difficult drawing clear conclusions.

Moreover, the SESAR target concept of operations [8] introduces new paradigms, such as TBO and PBO (trajectory based operations and performance based operations); where a more dynamic optimisation and allocation of ATM resources is foreseen, in order to enable the airspace users (AUs) to fly with the minimum amount of constraints. It is expected these new concepts will bring a significant positive impact in ATM performance. Current performance frameworks and PIs, however, might not be able to properly capture ATM performance in this future operational paradigm [5].

It is also worth noting that a very important aspect in ATM performance management is balancing between various KPAs by including their interdependencies into the analysis. So far, all relevant organizations observe KPAs independently. To the best of our knowledge, there is no performance scorecard to track achievements versus goals, such that also captures the effects of promoting one PI versus other PIs belonging to different KPAs, or even to the same KPA.

Aiming to assess these open questions, the APACHE project aims at providing advanced simulation, optimisation and performance assessment tools with the objective to better capture ATM performance and assess the complex interdependencies among KPAs. New (or enhanced) PIs were already proposed [5], which not only aim to improve current ATM performance assessment, but also are expected to better capture performance in a future ATM paradigm.

A key element in APACHE is the development of a novel ATM simulation system, which is used with two different purposes. On one hand, to synthesize traffic and airspace scenarios, simulating different operational contexts and enabling in this way, the possibility to perform what-if assessments (“Pre-ops” ATM performance assessment). On the other hand, to provide advanced models and optimisation tools that can support the implementation of novel and more accurate PIs, which can be used for “Pre-ops” but also for “Post-ops” (monitoring) purposes. This paper presents this simulation System and shows some illustrative examples obtained after different software integration and validation tests.

II. The APACHE Project

APACHE (assessment of performance in current ATM operations and of new concepts of operations for its holistic enhancement) is a project funded by the first wave of SESAR 2020 Exploratory Research. The APACHE consortium is formed by UPC (Coordinator), ALG, ENAC and UB-FTTE. The project covers the activity SESAR-11-2015 (ATM performance), started in May 2016 and will run for 2 years.

In this Project a new framework to assess ATM performance based on simulation, optimization and performance assessment tools is proposed. Thus, it is expected to fill some gaps of current state of the art methodologies in ATM performance assessment, aiming to capture the performance impact of ATM operations on different stakeholders taking into account a wide range of KPAs in a holistic approach. The specific objectives of the Project are:
1. to propose new metrics and indicators capable of effectively capturing European ATM performance under either current or future concepts of operation, fostering a progressive performance-driven introduction of new operational and technical concepts in ATM in line with the SESAR 2020 goals;

2. to make an (initial) impact assessment of some SESAR 2020 solutions (PJ06, PJ07-01, PJ08 and PJ09) using the new APACHE Performance Scheme along different KPAs; and

3. to analyse the interdependencies between the different KPAs by capturing the Pareto-front of ATM performance, finding the theoretical optimal limits for each KPA and assessing how the promotion of one KPA may actually reduce (and in which proportion) the performance of other KPAs.

Validation and example case studies are expected to be performed at EU-wide and/or functional airspace block level.

A. Research Approach

APACHE revolves around a novel system that is expected to generate optimal trajectories, considering of the business models of the airspace users; optimal airspace configurations, considering ANSP needs and constraints; and integrate both of them into an advanced air traffic flow management (ATFM) scheme. The same system can be configured to reproduce different modes of operation, representative of current ATM, or simulating some future SESAR 2020 Solutions.

Fig. 1 shows the overall concept of the APACHE framework. First, several scenarios to be studied are defined, setting up different options regarding the demand of traffic, airspace capacities and eventual restrictions; the SESAR solution(s) to be enabled; and the level of uncertainty to be considered. The APACHE-TAP (trajectory and airspace planner), which could be seen as a small prototype of an ATM simulator, has a double functionality in this Project:

► To synthesize traffic and airspace scenarios representative enough of current operations; or emulating future operational concepts in line with the SESAR 2020 ConOps (i.e. one or more SESAR solutions enabled).

► To support the implementation of novel ATM PIs, which require from some advanced functionalities (such as optimal fuel trajectories considering real weather conditions, optimal airspace opening schemes, large-scale conflict detection, etc.)

FIGURE 1 — The APACHE framework

Then, the performance analyser (PA) module implements all the PIs of the APACHE performance framework, including as well some indicators from the current performance scheme for benchmarking purposes.
B. Scope of the Research

Taking into account the exploratory nature of Project and its short duration, ATM performance at ‘pre-ops’ (planning) level will be only performed for a reduced set of SESAR 2020 Solutions. For benchmarking purposes, some scenarios will enable only certain SESAR solutions (but not all of them at the same time) [9]. Moreover, several assumptions and limitations are present in the implementation of the APACHE-TAP. The most relevant ones are summarised below:

- Only the en-route airspace structure is considered: only the airspace above FL195 (and flights cruising above it).

- The APACHE-TAP does not simulate tactical operations: only AUs trajectory planning, strategic airspace management and ATFM processes (pre-tactical layer) are considered. Moreover, interactions with airports are not taken into account (neglecting delays due to tactical airport operations). Moreover, all delay attributable to AUs [such as maintenance issues] is also neglected.

- Only IFR (instrumental flight rules) traffic is considered.

- Flexible use of airspace (FUA) or advanced FUA concepts are not modelled.

- Remotely piloted aircraft systems (RPAS) and unmanned aircraft systems (UAS) operations are not considered.

Fig. 2 wraps up the scope of the research done in the APACHE project, in terms of ATM performance assessment:

**FIGURE 2 — Scope of the ATM performance assessment planned in the APACHE project**

- "Post-ops" analysis (monitoring): the APACHE framework is able to compute a set of PIs for historical data. Ideally, these data should come from the Network manager and/or the different ANSPs. In the context of the APACHE project, this data is taken...
from Eurocontrol’s demand data repository 2 (DDR2) [10]. It contains trajectories according to the last filed flight plan by the AUs (M1 files); regulated trajectories (M2 files); and trajectories as captured in the enhanced tactical flow management system (ETFMS) after the flight has been operated (M3 files). It is expected that this data will be accurate enough to demonstrate the usefulness of the APACHE System.

- *‘Pre-ops’ analysis (planning)*: Besides being used to support the computation of some PIs (as in the ‘Postops’ analysis), the APACHE-TAP is used here to generate (synthesise) the scenarios to be studied. This allows for benchmarking current operations with future concepts and also to capture the Pareto front of ATM performance. Since simulating the tactical layer of ATM is out of the scope of the Project, the APACHE-TAP simulates shared business trajectories (SBT), by modelling the airspace user’s behaviour; and reference business trajectories (RBT), by modelling the DCB negotiation process via the Network Manager.

### III. The APACHE Framework

As shown in Fig. 1, the APACHE framework consists of the integration of several software modules: those included in the APACHE-TAP and those in the performance analyser (PA).

#### A. The APACHE-TAP: traffic and airspace planner

This module is in charge of simulating trajectories and airspace configurations, either for synthesising hypothetical scenarios, recreating historical data or supporting the PA with essential information to compute certain PIs. All modules composing the APACHE-TAP can be configured either to simulate operations in the current ATM paradigm, or to simulate future operations in line with some SESAR 2020 Solutions.

1. **Trajectory Planner (TP)**

The TP component generates and simulates traffic scenario (4D trajectories) based on real or future traffic demand and weather data. The computed trajectories can be optimised according to different optimisation objectives and constraints, configured in the definition of the scenario. To give a couple of examples, the TP is able to compute the most preferred trajectory for the airspace user in a structured route environment, considering real weather and airspace route charges; or to generate the most environmentally friendly trajectory, needed by the PA to compute certain PIs.

This module, developed by UPC, decouples the optimisation of the lateral and vertical profile, implements a module to model aircraft performance (such as fuel flow and aerodynamic drag magnitudes) and a module to process and model weather data [9], [11]. The principal modes of operation of the TP are:

- **Current operations**: the module is configured to use currently published airways (structured routes) and free route areas (FRA). Data from Eurocontrol’s DDR2 is taken. In the vertical domain, current flight level allocation and orientation schemes are used.

- **Full free route**: taking SESAR 2020 solutions PJ06 (trajectory based free routing) and PJ07-01 (AU processes for trajectory definition) [8] to their theoretical limit, this mode of operation will assume that airspace users can freely optimise their trajectories from the origin airport to the destination airport.

- **Continuous Cruise Climb**: the TP can, eventually, simulate hypothetical future operations where flight level allocation and orientation schemes are removed,
allowing continuous cruise climb operations en-route (not a SESAR solution per se, but useful as baseline for maximum fuel efficiency flights).

2. **Airspace planner (ASP)**

The main objective of this component is to simulate the airspace management service of the current and future ATM environments. Airspace management services aim to improve airspace design and utilisation in order to ensure delivery of the performance targets for the ATM system. For a given traffic sample and airspace structure with operational limitations, the ASP component finds an optimal sector opening scheme, i.e. an optimal list of airspace configurations or optimal grouping of the Sector Building Blocks (SBB) for each period of time, depending on ATM environment.

This module, developed by ENAC, has two principal modes of operation, as summarised below:

- **Static sectors**: mode according to the current ATM concept of operations, where for each period and for each Air Traffic Control Centre (ACC) one airspace configuration is selected, from the list of predefined set of configurations, consisting of one or more elementary/collapsed sectors. Sector grouping/ungrouping principles are respected by constraints on the airspace configurations that are selected in two consecutive periods. Since, nowadays, each ACC works independently the problem is separable and it’s modelled as shortest path problem and solved using dynamical programming method.

- **Dynamic sectors**: simulation of SESAR solution PJ08 (Management of dynamic airspace configurations) allowing airspace to be managed as a continuum in order to make optimum use of available airspace resource. In this mode of the ASP, existing elementary sectors are taken as SBB and grouped into controlled sectors not previously defined and not taking into account ACC borders. To keep stability of airspace configuration to an acceptable level, distance between SBB groupings for two consecutive periods is measured. This problem is modelled as a multi-period graph partitioning problem and solved using evolutionary algorithms ([12]).

More details of the implementation of this component can be found in [9], [11].

3. **Traffic and Capacity Planner (TCP)**

This module, developed by UPC, is in charge of network optimisation by balancing demand and capacity. It has also two main modes of operation:

- **Computer Assisted Slot Allocation (CASA)**: configuration of the module according to current ATM concept of operations, where DCB problems are solved by delaying aircraft on ground following a ration-by-schedule principle ([13]).

- **Advanced DCB (ADCB)**: simulation of advanced DCB measures enabling some degree of collaborative trajectory planning close to the execution phase aiming at simulating SESAR solution PJ09 (Advanced demand and capacity balancing). In this mode, the TCP will also consider alternative routes, previously proposed by the AUs (ATFM re-routing), and/or flight level capping, and/or linear holding strategies ([14]; as alternatives to ground holding for DCB purposes.

4. **Performance Analyser (PA)**

The Performance Analyser (PA) module receives, on one hand, the outputs from the APACHE-TAP in order to compute several PIs for different KPAs; and, on the other hand, might provide some feedback regarding the intrinsic risk of the simulated scenario (traffic patterns and proposed sector opening schemes).
1. **Computation of PIs**

The PA implements the computation of a wide set of PIs, providing also some visualisation mechanisms to improve the user experience when assessing the results of the different case studies. The PIs implemented in the PA component can be used for ‘Pre-ops’ assessment, ‘Post-ops’ assessment, or both.

In [5] a total of 40 new (or enhanced) Performance Indicators (PIs) were proposed, along with 18 possible variants for some PIs, covering a total of 11 KPAs. Taking into account the scope, resources and time-frame of the APACHE Project some of these PIs are not finally implemented in the PA, either because some of them require very complex and mature models, and/or due to the lack of data required to implement them. Nevertheless, they are candidates for inclusion in future evolutions of the APACHE framework. Taking this into account, the APACHE framework finally implements a total of 25 new (or enhanced) PIs and 17 PI variants. Moreover, and for benchmarking purposes, 5 performance indicators of current Performance Framework used by the SES/PRU, and reported regularly in their annual Performance Review Reports (PRRs), will be also computed by the APACHE framework. For further details, the reader is referred to [11].

Note that the APACHE framework could also be set up to monitor and target performance in real-time, or at different time-frames regarding the different traffic and airspace planning phases. These real-time capabilities could contribute to the effective implementation of Performance Based Operations (PBO) in a future ATM in which air traffic and airspace will be planned collaboratively and dynamically in order to adapt the KPA performances of the operations to the uncertain changing conditions of the ATM and weather.

2. **Risk Assessment (RA)**

The RA component, developed by UB-FTTE, has two main objectives: to provide safety feedback on traffic pattern and sectorization provided by APACHE-TAP; and to compute safety PIs. The module is composed by 1) a separation violation detection module; 2) a TCAS activation module; and 3) a risk of conflict/accident assessment module.

The first module compares the separation between two aircraft with a given separation minima (both in horizontal and vertical). Once a conflict is detected, this module calculates its duration and severity. If the situation worsens, the TCAS model is activated, which counts Traffic Alerts, Resolution Advisories, as well as Clear of Conflict warnings.

The risk of conflict assessment module is based on the calculation of ‘elementary risk’, which is defined as the area between the surface limited by the minimum separation line and the function representing the change of aircraft separation. The risk of conflict is then defined as the ratio between the ‘elementary risk’ and the observed period of time. Apart from the risk between specific aircraft pairs, an assessment of the total risk in a given sector is also considered.

The conflict/accident risk between aircraft pairs and the total conflict/accident risk depends on airspace geometry, traffic demand, aircraft velocities, spatial and temporal distribution of air traffic in the airspace as well as the applied separation minima. As such, the risk value taken as a safety feedback could suggest changes in flight trajectories and/or changes in sector boundaries, i.e. sector geometry.

More details of the implementation of this component can be found in [15], [16].

IV. **Illustrative Examples**

Aiming at showing the capabilities of the APACHE System and the successful integration of all components, this section presents some intermediate simulations and preliminary results. It is worth noting that at the moment of finishing this paper the APACHE System
was in its last stage of development and testing. According to the APACHE Project schedule, scenario and case studies assessment will take place in the last quarter of 2017.

The test case shown here corresponds to a “pre-ops” analysis done with the traffic demand taken from February 20th 2017, during 24h, and only considering those flights crossing the FABEC airspace. Demand data has been obtained from Eurocontrol’s DDR2, including the aircraft type, departure time and origin/destination airports. Since the APACHE Project focuses in the en-route phase, all flights with a requested flight level below FL195 were discarded for the simulations. Moreover helicopter and piston engine aircraft were also discarded, leading to a total of 14,034 scheduled flights analysed in this test case.

Airspace data, consisting of elementary/collapsed sector and airspace configurations definition, as well as, capacities of the sectors; were also taken from the AIRAC data from the DDR2 supplemented by French national data repository.

A. TP results

In order to run the TP, weather data for the same day of and region of study was gathered from the Global Forecast System (GFS), a weather forecast model produced by the National Centers for Environmental Prediction (NCEP) and provided in GRIB formatted files. Aircraft performance data, for each aircraft type, was obtained from Eurocontol’s BADA v4.2.

As explained before, for “pre-ops” analysis the TP is also used to synthesise the trajectories needed to recreate the scenario to study. For this purpose, each flight has been simulated with a random cost index (CI) and landing mass, following a normal distribution. Depending on the aircraft model, the distribution of the CI has been set assuming that the majority of flights operate at long range cruise (LRC) setting. The normal distribution for the payload mass is centered to 90% of the maximum landing mass, with a standard deviation of 10%.

In order to illustrate the TP capabilities two sets of traffic were synthesised (see Fig. 3): current operations, with structured routes and some FRA; and full free route operations from origin to destination airports. As expected, the spatial distribution for the full free route scenario (Fig. 3(b)) is larger than for the structured route case (Fig. 3(a)), showing also more direct and efficient trajectories (see performance assessment in Sect. IV-D1 below).

FIGURE 3 — Synthesised trajectories crossing FABEC airspace
B. ASP results

Fig. 4 shows the distribution of the number of opened positions in the FABEC airspace during the morning peak of the day of study. The ASP results (blue) are compared with those obtained from NEST ICO tool (red)\(^1\) for validation.

**FIGURE 4 — Number of open positions for FABEC airspace**

As seen in the figure, there is a high matching between ASP and ICO results. However, due to the higher flexibility of the ASP algorithm airspace configuration is better adapted to the traffic resulting in the lower number of the open positions.

Fig. 5 shows the distribution of the load (overload or underload) of the active sectors, expressed as variation percentage of entry count from capacity value. Fig. 5(a) shows, for each period, a five-number summary of the load distribution, including the Interquartile range - IQR (grey bars), for opening schemes provided by the ASP (blue) and ICO (red). Lower IQR in the opening scheme proposed by ASP signifies smaller dispersion, i.e. more even distribution of the load among active sectors. As shown, load median (black circle in 5(a)) in the ASP opening scheme is always closer to optimal (zero) value that implies higher capacity utilisation and explains lower number of open positions compared to ICO results.

A more detailed analysis is given in Fig. 5(b), showing the distribution of the sector load for a single period of the same example. For the sake of compactness, only sectors from the French airspace are shown, organized in 5 ACC having 11 clusters in total. Blue bars represent ASP results, while red bars represent ICO results. Green lines in the figure represent the mean value of the load. Fig. 5(b) confirms, once more, that ASP opening scheme shows more even distribution of the workload among controllers, represented by smaller deviation of sector loads from the mean value.

---
\(^1\) NEST, the Network Strategic Tool from Eurocontrol, is a stand-alone desktop application for airspace structure design and development, capacity planning and post-operations analysis, the organisation of traffic flows, the preparation of scenarios for fast time simulations and ad-hoc studies at local and network level.
C. TCP results

Once the ASP has generated an optimum sectorisation, trying to better allocate airspace capacity, the TCP is responsible to regulate the demand, avoiding to exceed the maximum capacity in any sector. The illustrative results shown here correspond to the 24h test benchmark described above, but focusing only in the French airspace. Both TCP modes of operation have been tested (current CASA algorithm and advanced DCB).

With the CASA algorithm, 2,510 aircraft were subject to delay with a total system delay of 406,042 min, leading to 55 min of average delay. The authors acknowledge that these delays are higher than what is commonly seen in the ECAC. The results shown here assume that a sector is regulated when the demand exceeds the published nominal capacity plus a 10% of sector overload allowance. In real operations, however, each hotspot is carefully analysed by the corresponding flight management position (FMP), having different strategies and criteria to finally decide whether a regulation should be applied or not (and which is the sector overload allowance, if any). This detailed behaviour will not be modelled in the APACHE simulator. Nevertheless, results from this simplified CASA algorithm could still be valid for benchmarking purposes.

Besides delay, the advanced DCB functionality allows for pre-tactical re-routings or flight level capping as a possible solution to solve the demand-capacity imbalance problem. The extra set of trajectories (avoiding hotspots laterally or vertically) are provided by the TP, emulating a DCB negotiation process with the network manager. Then, the DCB
algorithm finds the system-wide optimal solution that minimises the total cost for the airspace users (considering the cost of fuel and an estimated cost of delay).

Table I shows the results, detailing the number of trajectories and delay for each of the three cases. With the advanced DCB algorithm the total departure delay has been significantly reduced (821 min with 301 flights performing delay), but at the expense of allowing less efficient trajectories (re-routings or level capping). This is one of the trade-offs that are subject of study in the APACHE project.

**TABLE I — Advanced DCB algorithm results (French airspace)**

<table>
<thead>
<tr>
<th></th>
<th>Original trajectory</th>
<th>Re-routing</th>
<th>Level capping</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total flights</td>
<td>6,057</td>
<td>284</td>
<td>419</td>
<td>6,760</td>
</tr>
<tr>
<td>Delayed flights</td>
<td>227</td>
<td>20</td>
<td>54</td>
<td>301</td>
</tr>
<tr>
<td>Non Delayed flights</td>
<td>5,830</td>
<td>264</td>
<td>365</td>
<td>6,459</td>
</tr>
</tbody>
</table>

**D. Performance Assessment**

As illustrative example of the Performance Analyser module, some PIs of the environmental impact and safety KPAs were computed using as input data the set of trajectories following current published structured routes and FRA synthesised by the TP, but filtering for those flights crossing only the French airspace.

1. **Environmental indicators**

Fig. 6 shows the PIs ENV-1.1: Strategic ATM inefficiency on the horizontal track and ENV-2.3: Strategic ATM inefficiency on the trip fuel [see [5] for details].

**FIGURE 6 — Strategic ATM inefficiency (24h traffic over French airspace)**

Essentially, ENV-1.1 compares the horizontal flight distance of the trajectory under assessment (in this case those from Fig. 3(a)) with a *environmentally optimal* baseline trajectory, which in this case was a full-free route trajectory with the Cost Index set to zero and optimized taking into account weather conditions (and therefore, different from the orthodromic trajectory between origin and destination airports). As seen in Fig. 6(a), the average horizontal route inefficiency is approximately 40 NM with particular flights that can reach up to 100 NM of horizontal inefficiency.

ENV-2.3, in turn, compares the fuel consumption of the trajectories under assessment with the baseline trajectory. As seen in Fig. 6(b), three different baseline trajectories are used (all of them taking into account weather conditions and the corresponding optimal vertical profile):

1. **FR-CI0**: An optimal full free route trajectory with the Cost Index set to zero.
2. **SR-CI0**: An optimal trajectory, constrained with current published structured routes and FRA, and with the Cost Index set to zero.
3. **FR-CIAU**: An optimal full free route trajectory with the same Cost Index as simulated in the assessed trajectory.
If the SR-CI0 trajectory is used as baseline, this PI captures the impact of flying at not optimal altitudes (from an environmental impact point of view), due to the fact that the AU has planned the trajectory at a Cost Index higher than zero [as explained above, following a normal random distribution in our simulations]. As seen in the figure, fuel inefficiencies are relatively low (below 100 kg in the majority of flights).

The FR-CI0 variant captures the impact on the environment of flying on structured routes, but also due to the fact that the AU has planned a Cost Index higher than zero. Conversely, the FR-CIAU variant isolates the inefficiency in fuel only attributable to ATM (due to the structured route network). In this case, an average of 350 kg of fuel could be saved if aircraft were allowed to fly full free optimal routes.

2. Safety indicators

A risk assessment was done for the selected day of study and for both sets of trajectories generated by the TP but filtering for those flights crossing only the French airspace. Table II shows the results of some safety PIs, as computed by the RA module. The minimum separation values (for SAF-4) were set to 5NM in the horizontal plan and 1000 ft in vertical. Moreover, the simulation time increment was set to 10s. As expected, those indicators are lower for the full free route scenario since potential trajectory crossings are more geographically spread.

<table>
<thead>
<tr>
<th>PI</th>
<th>Description</th>
<th>Current network</th>
<th>Full free route</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAF-1</td>
<td>Number of Traffic Alerts</td>
<td>300</td>
<td>131</td>
</tr>
<tr>
<td>SAF-2</td>
<td>Number of Resolution Advisories</td>
<td>73</td>
<td>1</td>
</tr>
<tr>
<td>SAF-3</td>
<td>Number of Near Mid Air Collisions (NMACs)</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>SAF-4</td>
<td>Number of Separation Violations</td>
<td>1376</td>
<td>939</td>
</tr>
<tr>
<td>SAF-7</td>
<td>Risk of conflicts/accidents</td>
<td>$5.5 \times 10^{-3}$</td>
<td>$3.0 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

Fig. 7 shows the geographical location of the closest points of approach (CPA) that were below 5NM for the example of study. One should keep in mind that CPAs shown are aggregated for 24h, which means that each dot represent a conflict point between different pair of aircraft, at different altitudes and in different time during the day. Also note that even if the test flight set corresponded to flights crossing the French airspace during 24h, CPAs could be located outside this airspace, since the full trajectory was taken into account.

FIGURE 7 — Location of conflicts in French airspace (CPA below 5NM)

This RA assessment could also be used to easily identify these geographical locations with higher frequency of potential conflicts, which could eventually be used as a safety feedback to amend sector boundaries in the ASP component or to be taken into account for advanced DCB purposes.
V. Conclusion

Air traffic management is progressively transitioning to a performance based system, with many key performance areas (KPAs), which show, in the majority of cases, complex and still not well understood interdependences. Properly assessing ATM performance, and conveniently capturing these trade-offs among KPAs, is still a research challenge for the ATM community and a need for correctly deploying novel operational and technical concepts, such those proposed by SESAR.

By combining simulation, optimisation and performance assessment tools, the APACHE framework aims at generating knowledge on the theoretical optimum for each KPA and to assess their Pareto optimality. This will allow a better understanding of the ATM performance drivers and, besides “post-ops” and “pre-ops” analysis, it might be useful for targeting and base-lining in future performance reference periods (RPs).

Acknowledgement

The authors would like to thank Mr. Ramon Dalmau, Mr. Marc Melgosa and Mr. Yan Xu, from UPC, for their support in the development of the TP and TCP modules; Mr. Vladimir Coca, Ms. Inês Costa and Ms. Georgina Ansaldo, from ALG, for they contributions when consolidating the research methodology and defining the validation scenarios; and Ms. Bojana Mirkovic, Mr. Goran Pavlovic, Prof. Obrad Babic and Prof. Vojin Tosic, from UB-FTTE, for they contributions in the definition and implementation of new PIs.

REFERENCES

SIMULATING THE RISK OF BIRD STRIKES

Isabel C. Metz*, Joost Ellerbroek†, Thorsten Mühlhausen*, Dirk Kügler* and Jacco M. Hoekstra†
*German Aerospace Center
Institute of Flight Guidance
Braunschweig, Germany
†Delft University of Technology
Faculty of Aerospace Engineering
Delft, The Netherlands

Abstract—This paper presents a fast-time simulation environment for assessing the risk of bird strikes in aviation. An existing air traffic simulator was enhanced in order to simulate air and bird traffic simultaneously and to recognize collisions between birds and aircraft. Furthermore, a method was developed to generate bird movement information from different radar sources. The resulting set-up represents the first simulation environment to perform fast-time simulations including air traffic and bird movements. A verification with real data revealed that approximately thrice as many bird strikes occur in the simulation as in reality. When considering bird reaction to approaching aircraft, which is not covered in the simulation as well as unreported strikes, this implies an adequate result. For this reason, the simulator can serve as valuable tool to analyse the risk of bird strikes and to evaluate new Air Traffic Management concepts to reduce the number of these events.

I. Introduction

Collisions between birds and aircraft, so called bird strikes, represent an ongoing threat to aviation safety [1]. To mitigate the risk for these events, airports maintain a bird / wildlife strike programme as required by the International Civil Aviation Organization (ICAO). These programmes aim at excluding wildlife in general and especially birds from the airport grounds, for example with habitat modification or harassment [2]. These measures have already led to a reduction of bird strikes at many airports [3]. However, bird strikes are not limited to the airport area. Aircraft taking off and landing have an increased risk of colliding with birds up to an altitude of 3000 ft (ca. 1000 m). Hence the endangered area spreads much wider than the airport perimeter where the wildlife strike programmes are effective. Concepts to further reduce the risks of bird strikes would involve the pilots and Air Traffic Control (ATC). [1], [4]. For example, based on bird movement information from the area, a take-off could be delayed to prevent a probable bird strike during departure (cf. e.g. [5]). The introduction of such a concept has the potential to prevent bird strikes that nowadays would be inevitable. However, delaying of traffic would lead to a reduced runway capacity which is especially critical for airports with high traffic loads. So far, the consequences on the safety and capacity of an airport when implementing such a bird strike prevention system for ATC and the pilots have not been studied. In the present work, a fast-time simulation environment which allows the analysis of these effects as well as the comparison of different options for the implementation of such a system is described. The applicability of the simulation environment as a research tool for assessing the risk for collisions between aircraft and birds was then verified with real bird and air traffic. This paper describes the generation of bird movement information for the simulation, the underlying model for the detection of bird strikes in the simulation as well as the resulting set-up. This is followed by an analysis of the simulation results.
II. Method

To develop a simulation environment for the analysis of bird strikes, an underlying simulation platform is required. This study relies on the BlueSky Open Air Traffic Simulator developed by Delft University of Technology. This simulator enables real- and fast-time simulation of air traffic [6]. A key advantage for the work presented here lies in the simulator’s open character: BlueSky can freely be downloaded and modified. Thus, modules for bird traffic and collision detection between birds and aircraft could thus be integrated without any restrictions. The resulting simulator set-up facilitates the simultaneous simulation of bird movements and air traffic as well as the recognition of bird strike occurrences.

To run simulations in this set-up, input data for bird movements and air traffic is required. The following paragraphs describe, how these were obtained and processed for the simulation. Subsequently, the developed conflict-detection algorithm to identify bird strikes in the simulation is presented. Finally, the simulation set-up is summarized.

A. Bird Movements

For this study, bird movements in the extended airport area are relevant. This includes local movements at the airport itself as well as migrating patterns of birds in higher altitudes in the arrival and departure corridors. Avian radars offer high-resolution information about tracks of individual birds [7]. This seems ideal for this study. However, the range of the chosen avian radar is limited due to its range capability as well as radar shadowing from ground objects. Hence, an additional source for bird movements in the arrival and departure corridors was required for this study. Weather radar was selected for this purpose. Due to their ability to recognize birds, weather radars are widely used for the quantification of bird movements [8], [9].

By combining data from weather and avian radar, the full picture of bird movement in the extended airport environment can be visualized. While weather radar data has previously been used to visualize migration movements (cf. e.g. [10]), this work represents the first study to combine avian and weather radar data to gain movement information about local and migrating birds. Data was available for the area around Eindhoven airport in the Netherlands, namely from an avian radar stationed at that airport as well as from a weather radar in De Bilt. The avian radar serves as input for bird movements from ground to 200 m (ca. 660 ft), the weather radar covers the altitude band from 200 m to 1000 m (ca. 3000 ft). The lateral range of 25 km of the weather radar in De Bilt does not reach Eindhoven airport. However, the broad-fronted bird migration patterns over both location are strongly comparable [Hans van Gasteren, Royal Netherlands Air Force, personal communication, 4/12/2016]. Hence, the bird densities recorded by the weather radar were projected to the airport area of Eindhoven. The following paragraphs describe, how information from the two radar sources was extracted and prepared for the simulation.

1. Avian Radar

The obtained avian radar data contains time-stamped positions of moving objects connected to tracks by a Kalman-Filter. Every track is assigned to an id as well as an object type. For this study, the data was filtered for the object types small bird, medium bird, large bird and flock. Moreover, to gain representative tracks, only birds with at least 20 timestamps were selected. Due to the radar’s turning frequency of 0.75 Hz, this corresponds to a minimum track duration of ca. 27 seconds. This filtering reduces the number of tracks considered and thus is expected to slightly decrease the number of bird strikes in the simulation.

The avian radar at Eindhoven is a horizontal X-band radar providing latitude and longitude information of bird positions. Due to the low elevation resolution, the bird’s altitude is not...
Because the beam size of the radar increases with distance, the range of potential vertical positions rises as well. When crossing a lateral distance of ca. one kilometre from the avian radar, the beam exceeds the altitude of 200 m. As weather radar data is used from 200 m upwards, an overlap of the two sources occurs as Figure 1 visualizes: Area 3 is considered by both radars. To avoid double counts, the number of birds tracked by the avian radar was set as follows: Assuming that birds fly at constant height once airborne, all birds flying within the range of area 1 during one time step at least were selected. For the outer range (areas 2 and 3), birds were filtered corresponding to altitude distributions determined by Shamoun-Baranes, van Gasteren and Ross-Smith [11]. They conclude that 48% of all birds fly below 200 m during daytime. At night, 35% fly below 200 m. Consequently, two out of three birds were removed from areas 2 and 3 to gain a conservative estimate of the number of birds. Area 5 is not covered by neither of the radars.

**FIGURE 1 — Areas covered by the avian and weather radar (not to scale). Area 1 and 2: avian radar. Area 3 and 4: weather radar. Area 5: no coverage**

### 2. Weather Radar

The data of the chosen C-band weather radar in De Bilt contains information about bird reflectivity per km² and in altitude bins of 200 m. For this study, the altitude bands from 200 m to 1000 m were considered. To convert reflectivity to density in \( \frac{\text{birds}}{\text{km}^2} \), the methodology described by Dokter et al. [12] was applied. Velocity and direction of birds was obtained from the Northern and Eastern speed components given in the weather radar data. As the study performed by van Gasteren et al. [13] revealed, recorded bird velocity is underestimated by the radar. Hence, it was increased by \( 3.44 \, \text{m/s} \) as suggested by van Gasteren et al. [13]. To consider the standard deviations, individual birds were assigned to a velocity in a range of \( 12 \, \text{m/s} \) around the average velocity. The applied standard deviation for bird direction amounts to 45° [13], (Hans van Gasteren, Royal Netherlands Air Force, personal communication, 11 October 2016). As for the avian radar data, altitude information was assigned randomly within the respective altitude band.

### 3. Processing of bird movements

After extracting bird movement information from the two radar sources, it was made available for the simulation by storing it in bird movement plans per simulated day. This reduces calculation effort during the simulation and allows a reproducibility of simulations. For birds covered by the avian radar, the bird movement plan contains all time-stamps and the corresponding track data. The last time-stamp is marked with a trigger for the simulation to remove the corresponding bird.

Birds covered by the weather radar are stored with information about their initialization and removal. The calculation of these steps is based on a preprocessing method.
described subsequently. In the first time-step covered by the weather radar, as many birds as represented in the input data’s reflectivity are created at random positions in the designated airport area. In every subsequent update step, the birds’ positions are extrapolated based on their speed and direction and compared to the boundaries of the airport area. Birds that left the area are marked to be deleted in the bird movement plan. They are replaced with new birds in order to keep the flow constant. As the weather radar birds mainly represent migrants, the general flight direction of individuals is very similar. For this reason, they leave the area in the same direction. As a consequence, the birds to replace the fly-outs are initialized at the opposite boundary. The number of these fly-ins is corrected for potential changes in reflectivity between time steps. Furthermore, the number of birds remaining in the area is kept corresponding to the reflectivity. If the reflectivity increases, birds are randomly generated over the entire area. In case of a decrease, birds are randomly deleted. Every initialization and every removal calculated in the preprocessing is stored in the flight movement plan. During the simulation, the bird’s actual position is interpolated (avian radar) respectively extrapolated (weather radar) from the given data in the bird movement plan.

Birds do not always fly alone but also in flocks [14]. Therefore, every bird in the simulation represents one or multiple individuals. This is relevant for later evaluations of the risk for damage resulting from bird strikes. Birds from the avian radar are already grouped in individuals and flocks. However, no information about the flock sizes is available from the input data. From the weather radar data, information about the total number of birds, but not about their distribution in birds flying individually and birds flying in flocks can be obtained. To generate this missing information for the bird movement plans, data from a multi-year-study on bird migration over the Netherlands was used [15]. Despite its age – the report dates from 1985 – this is the most complete source available, including data about the flock and altitude distribution of the most representative bird species in the Netherlands. Furthermore, it contains detailed information for birds flying below and above 200 m which perfectly corresponds to the boundary between avian and weather radar in this study.

Two flock-size distributions were calculated: The distribution for the avian radar birds includes birds flying in flocks. The distribution for weather radar birds also considers the data of individual flying birds per species to receive information about the share between individual and group flyers.

To gain flock information about birds within the avian radar range, the top 15 species reported to fly below 200 m were chosen, representing 89% of all birds in this category. Regarding birds flying above 200 m, 14 species representing 97% of all birds of this category were considered. From the mean flock size per species obtained from the study by Lensink and Kwak [15], weighted flock size averages were calculated for the two radar sources. For the weather radar birds, the distribution between birds flying alone and in groups was calculated in addition. The flock sizes were finally determined by applying a Poisson distribution as suggested by Lensink and Kwak [15] by using the weighted average as expected number. Poisson distributed values include one and zero. To obtain valid flock sizes with a minimum of two members, results smaller than two were increased accordingly.

Bird migration patterns differ significantly between day and night. During daytime, birds mainly migrate in groups, while they fly individually or with large distances between flock neighbours during the night [14], [16], [17]. Hence, the described distributions are only valid for diurnal migration. For nocturnal migration, which the study of Lensink and Kwak [15] does not cover, the flock size distribution was obtained from Hüppop et al. [18] and assigned to the species Lensink and Kwak [15] had observed. The designator for applying the flock distributions for nocturnal or diurnal migration in the bird movement plan is civil twilight.

Next to the number, also the size of birds involved in a bird strike influences the risk of aircraft damage [2]. The avian radar data only contains size information for birds flying individually, but not for flocks. From the weather radar data, this information
cannot be retrieved at all. To assign birds to a size, the chosen species from the study of Lensink and Kwak [15] were categorized based on their weight into the classes small, medium or large as defined by the aviation authorities [19]. Corresponding to the species distributions, these size classes were assigned to the birds in the movement plans.

B. Air Traffic

To get realistic flight plans for air traffic, scenarios based on real traffic were generated. The availability of bird movement information set the simulation area to Eindhoven airport in the Netherlands. This airport has a very low traffic volume [20]. To evaluate the impact of various traffic intensities on the risk of bird strikes, flight plans from additional airports were generated and transferred to the airport of Eindhoven to cover high, medium and low traffic volumes as well. For comparability and to facilitate the integration into the simulated airport area – Eindhoven has one runway – traffic from airports with one operational runway were selected. With regard to their ranking considering number of flights in the 2015 Airports Council International (ACI) traffic report, London Gatwick (UK) for high, Geneva (CH) for medium and Birmingham (UK) for low intensity were chosen. In addition, a scenario covering Eindhoven traffic was generated. The flight plans were generated based on data from one representative day per airport in 2016 (source: European Organization for the Safety of Air Navigation (EUROCONTROL) database, accessed via Bruno Nicolas, Statistics Specialist, Eurocontrol, personal communication 5 April 2017 & 4 August 2017). Figure 2 visualizes the selected traffic volumes.

The flights of the considered traffic intensities were mapped to Eindhoven airport. For all simulations, runway 03 was active. Landing aircraft were initialized at 3000 ft at one
of the initial approach fixes MITSA or RUSAL and performed Instrument Landing System (ILS) approaches. Departing aircraft used the Standard Instrument Departures (SID) and were deleted once reaching 3000 ft.

C. Conflict Detection Algorithm

To detect bird strikes in the simulation protected zones were defined around birds and aircraft. A bird strike occurs, when a bird and an aircraft penetrate each other’s protected area. The definition of the protected zones is described subsequently.

1. Protected Zone Birds

The protected zone of birds depends on two parameters: The bird’s size and the number of birds represented by one bird object. The protected zone around individual and flocks of birds was modelled as disc to minimize the impairment of the collision detection algorithm on the simulation’s runtime performance. The shape of flocks, which varies amongst different species [22], was thus simplified. Due to the small height of birds, especially in comparison to the height of aircraft, it was decided to disregard this parameter and keep the protected zone of birds two-dimensional. The diameter of the protected zone of individual birds directly refers to the wingspan of the category it belongs to. For each of the bird categories *small*, *medium* and *large*, a weighted average for the wing span was calculated based on the species considered from [15] and their distribution.

To model the protected zone for flocks, the theory of dense packings of congruent circles in a circle [23] was used as a base. This theory describes how the radius of a circle increases with rising number of circles within that circle. Considering comprising circles, corresponding to protected zones of flocks, containing up to 20 circles which represent birds within the flock, functions for the radii of the protected zone for each bird category were developed. Thereby, the neighbouring distance between the individual birds was not considered. These distances have so far only been analysed for some species (cf. e.g. [16]). Moreover, for migrating birds, which are mostly relevant in context of this study, it is most efficient to fly adjacent or even with slightly overlapping wing tips [24]. Hence, this parameter was set to zero. Table I summarises the developed functions. Figure 3 visualizes them.

**FIGURE 3 — Increase in flock size radius with rising number of flock members**

![Graph showing increase in flock size radius with rising number of flock members](image)
### TABLE I — RADIi FOR DIFFERENT FLOCK SIZES

<table>
<thead>
<tr>
<th>Bird Category</th>
<th>Wingspan [m]</th>
<th>Flock Radius</th>
<th>Standard Deviation from Flock Size [m] in [23]</th>
</tr>
</thead>
<tbody>
<tr>
<td>small</td>
<td>0.32</td>
<td>$\sqrt{\frac{n_{\text{birds}}}{2}} + 0.06$</td>
<td>0.030</td>
</tr>
<tr>
<td>medium</td>
<td>0.68</td>
<td>$\sqrt{\frac{n_{\text{birds}}}{2}} + 0.16$</td>
<td>0.038</td>
</tr>
<tr>
<td>large</td>
<td>1.40</td>
<td>$\sqrt{\frac{n_{\text{birds}}}{2}} + 0.41$</td>
<td>0.050</td>
</tr>
</tbody>
</table>

### 2. Protected Zone Aircraft

The basic shape of the protected zone of aircraft corresponds to an upright cylinder. To consider the major aircraft types, the aircraft from the flight plans were categorized into the groups *wide body*, *narrow body* and *regional*. The parameters required for the definition of the protected zone per category were obtained for the aircraft with the largest wingspan in each group: The Airbus A380-800 represents *wide bodies*, the Boeing B757-300 *narrow bodies* and the Bombardier Dash 8-400 *regional aircraft*.

The protected zone’s diameter corresponds to the aircraft’s wing span. Because of their small front surface, an aircraft’s rudder and elevator experience almost no bird strikes [25], [26]. Thus, the tail section is cut from the protected area. Its arc length depends on the wing’s sweep.

An aircraft’s height strongly varies along its wingspan. Hence, if setting the protected zone’s height to the aircraft’s largest vertical expanse, the number of bird strikes would be strongly overestimated. Therefore, an average height was determined from the heights of the aircraft’s front surfaces prone to bird strikes: the wings, engines and the fuselage [27]. This average height represents the height required to be multiplied with the aircraft’s wingspan to obtain a rectangle corresponding to the aircraft’s relevant front surface. It is calculated by adding the front surfaces of the aircraft’s components given in Equation 1. The resulting protected zone is visualized in Figure 4. The key parameters to determine the dimensions of the protected zone per aircraft category are given in Table II.

#### FIGURE 4 — Top and front view of an aircraft’s protected zone (Airbus A380-800)

![Figure 4](image)

#### TABLE II — Parameters defining an aircraft’s protected zone

<table>
<thead>
<tr>
<th>Aircraft Category</th>
<th>Reference Aircraft Type</th>
<th>Radius [m]</th>
<th>Height [m]</th>
<th>Sweep [°]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Widebody</td>
<td>Airbus A380-800</td>
<td>79.75</td>
<td>1.99</td>
<td>33.5</td>
</tr>
<tr>
<td>Narrowbody</td>
<td>Boeing 757-200</td>
<td>38.0</td>
<td>1.01</td>
<td>25.0</td>
</tr>
<tr>
<td>Regional</td>
<td>Bombardier Dash 8-400</td>
<td>28.4</td>
<td>1.35</td>
<td>0.0</td>
</tr>
</tbody>
</table>

$$S_{\text{front}} = \left( b - 2 \cdot r \right) \cdot h_{\text{f}} + n_{e} \cdot r^{2} \cdot \pi + r^{2} \cdot \pi$$  

where $b$ represents the wingspan, $r$ the radius of the respective components and $h$ the height, all in metre. $n_{e}$ is the number of engines.
3. Conflict Detection

During simulation, the protected zones of birds and aircraft are constantly tested for overlaps. Every overlap of a bird’s and an aircraft’s protected zone leads to a bird strike. As a consequence, the bird hit is removed from the simulation and a bird strike is counted.

D. Simulation Set-Up

After describing the relevant input parameters for the simulation, the resulting set-up is summarized here. The number of bird strikes occurring at an airport strongly depends on the season: During migration as well as in summer, when many young and inexperienced birds fly, more strikes take place than in winter [1], [2]. To include seasonal effects within this study, bird movement plans were created for an entire year. It was decided to simulate one week per month in the period from October 2015 to September 2016 where radar data was available. This allowed to keep the simulation effort at a reasonable level while all seasons were covered and the number of days \( n = 84 \) per airport was representative. The weeks were chosen based on radar availability and weather. The reason for the latter criteria lies in the radar’s decreasing tracking ability with increasing precipitation [7]. By choosing weeks with little precipitation, a high detection rate and as such representative bird movement information was ensured.

Every bird movement plan was combined with the flight plans representing high, medium, low and very low traffic volume to study the effect of different traffic intensities on the bird strike risk. Depending on the airport, the traffic volume varies throughout the year [20]. This variation is implicitly considered by providing flight plans for different traffic intensities. The combination of bird data from 84 days and flight plans covering the four traffic intensities led to a total of 336 simulated traffic days.

By simulating the described scenarios, two goals were pursued. First, a verification, if the developed simulation environment appropriately reflects the risk of bird strikes at an airport, took place. Here, it was expected that more bird strikes would be counted in simulation than in reality. The main reason is, that the simulated birds are not modelled to show reactions to aircraft whereas in reality, birds often manage to perform last-minute escapes when an aircraft approaches. Furthermore, not all bird strikes are recognized or reported, especially ones with very small birds or strikes that did not damage the aircraft involved. Even with the slight reduction of simulated bird strikes due to the filtering of the avian radar data (cf. section II-A1), the number of bird strikes within the simulation should be higher than in reality. The second goal of the simulation campaign is to acquire data for a baseline scenario for further research involving new Air Traffic Management (ATM) procedures to avoid bird strikes.

III. Results and Discussion

The goal of this work was to develop a simulation environment to model the risk of bird strikes. The resulting setup enables fast-time simulations of bird and air traffic movements. Collisions between birds and aircraft are registered and counted as bird strike occurrences. To verify the set-up and to generate a baseline-scenario for further simulations, 336 days were simulated for the airport of Eindhoven, where the input data for bird movements originates from.

To evaluate the outcomes of the simulations, the bird strike rates were calculated for the four considered traffic intensities high, medium, low and very low. Additionally, the bird strikes occurrences were categorized by altitude band as well as by month of occurrence. Finally, the correlation between bird volume and number of bird strikes per season was determined.
The bird strike rate of an airport is generally given in number of bird strikes per 10,000 flights [28]. The average ratio of all bird strikes at Eindhoven airport amounted to 12.33 between 2007 and 2016 (source: Bird Strike Database, Royal Netherlands Air Force. Hans van Gasteren, 3/8/2017, personal communication). The bird strike rates resulting from the simulations are a little higher: They amount to 21.59 for high, to 19.48 for medium, to 21.78 for low and 15.07 for very low traffic volume. Due to last-minute escapes in reality which are not modelled in the simulation as well as because of unreported strikes, a larger offset would be expected. However, birds were filtered for duration of stay in the altitude band covered by the avian radar [ct. section II-A1]. For this reason, the risk for bird strikes is reduced in this altitude band. This is reflected in the altitude distribution of strikes in the simulation as visualized in Figure 5. As statistics from across the world consistently suggest, the number of bird strikes decreases exponentially with increasing altitude [26], [29], [27]. With regard to the simulation results visualized in Figure 5, a significant decrease can only be found from 201 m on – the altitude from which the weather radar data serves as source for bird movements. Between zero and 200 m, where the data from the avian radar was used, the number of bird strikes is only slightly higher than in the altitude band above. By comparing the number of all birds including the ones filtered out to the number of birds considered for the simulation, the theoretical sum of bird strike occurrences was calculated. To ensure the accuracy of scales, the number of birds was weighted with their average duration of stay. The determined values (top bars in Figure 5) increase the bird strike rate to a reasonable level. Table III presents the comparison between the simulated and theoretical bird strike rates. In contrast to the other scenarios, the rise of the bird strike rate when considering all birds present is relatively small in the very low-scenario. The main reason is that the offset between the number of all birds present within the opening hours and the number of birds selected for the simulation is much smaller than in the other scenarios.

**FIGURE 5 — Bird strike altitude distribution (0-200 m: avian radar, 200-1000m: weather radar) for the chosen traffic volumes**
TABLE III — Simulated and theoretical bird strike rates [number of strikes per 10,000 flights]

<table>
<thead>
<tr>
<th>Airport</th>
<th>Simulated Strike Rate</th>
<th>Bird Theoretical Bird Strike Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>21.59</td>
<td>41.43</td>
</tr>
<tr>
<td>Medium</td>
<td>19.48</td>
<td>42.53</td>
</tr>
<tr>
<td>Low</td>
<td>21.78</td>
<td>38.25</td>
</tr>
<tr>
<td>Very low</td>
<td>15.07</td>
<td>25.62</td>
</tr>
<tr>
<td>Eindhoven (reference)</td>
<td></td>
<td>12.33</td>
</tr>
</tbody>
</table>

Figure 6 visualizes the seasonal distribution of bird strikes (Figure 6(a)) and number of birds (Figure 6(b)). Additionally to the number of strikes in the chosen scenarios, the average number of strikes per month for the period 2007 - 2016 is given for the reference airport in Eindhoven. It has to be noted that the data from Eindhoven reflects all strikes that happened within one month. In contrast, the simulation results cover the number of strikes for one week per month. These numbers differ between the scenarios. This is mostly correlating with the air traffic volume. However, in five months, at least the same number of strikes occurred in the low-scenario as in the medium-scenario. This can be explained by the temporal distribution of flights: Between 6 and 7 a.m., where internationally most bird strikes are recorded [25, [29], more flights depart in the low-scenario (cf. Figure 2). In this period, 24% of all strikes happen in the low-scenario with 8% of the daily air traffic movements. In the medium-scenario, only 2% of all strikes occur with 4% of the daily traffic.

FIGURE 6 — Number of strikes and number of birds per month

The comparison of the number of strikes in the simulation with real data reveals two main differences: First, in the summer months, the number of strikes at Eindhoven are relatively high in comparison to the simulation results. This could be related to increased flight activity in Eindhoven during the summer months, while the scenarios consider average traffic volumes. The second deviation can be found in the month of March, where the simulation results increase significantly, while the number of bird strikes remains at a relatively low level at Eindhoven airport. When comparing the bird strike occurrences in the simulated scenarios with the number of birds (cf. Figure 6(b)), the peak in March is reflected in both statistics. Obviously the majority of spring migration...
took place in March in the year considered for the simulations. In contrast, the data from Eindhoven is averaged over ten years. In this period, the exact timing of spring migration could have shifted between the years. This probably led to a wider distribution of bird movements – and thus bird strikes – over the spring months. Overall, when comparing the seasonal trends between the simulation and reality, a high similarity can be found: in autumn, the number of bird strikes is relatively high. During the winter months, fewer bird strikes occur. In spring, the number increases again and has a maximum in June when many inexperienced juvenile birds fledge. With the named exception of March, the seasonal trends seem to be well reflected within the simulation. This is supported by the number of birds present in the simulation as shown in Figure 6(b). Most birds fly during migration in autumn and spring. In winter, there is very small bird activity while more birds fly in summer. The only offset between number of birds and number of strikes can be found in June: With regards to the number of birds flying, a large number of bird strikes occurred. This could be attributed to very high activity of juvenile birds. Due to a lack of experience, they cause significantly more strikes than adult birds. [30]. The peak in bird strike occurrences at Eindhoven airport in this month supports this assumption. The correlation between number of strikes and number of birds in the simulation was calculated for all months and for all months excluding June. The Spearman correlation was applied for this purpose as not all of the considered values are normally distributed. Table IV summarizes the results. It becomes clear that the exclusion of the values for June, where the high number of bird strikes is not related to a rise in number of birds, notably increases the correlations. Regarding these values, the high-scenario shows a strong significant correlation \( r(10) = 0.89, p < .001 \) the low-scenario a moderate significant correlation \( r(10) = 0.71, p < .01 \). The medium and very low scenarios do not correlate significantly \( r(10) = 0.44, p = 0.088; r(10) = 0.26, p = 0.281 \). This is most likely connected to the opening hours of the airports where the scenarios originate from: Scenarios with longer airport opening times have higher correlations between number of birds and bird strike occurrences (cf. Figure 2). Moreover, scenarios with longer opening times have a higher simulated bird strike rate (cf. Table III). The different opening hours also cause the higher number of birds in the low scenario compared to the medium scenario (cf. Figure 6(b)). The sample size for all the airports was twelve simulated months, which is very small for statistical evaluation. To gain more robust correlation results, Monte-Carlo experiments will be performed. The analysis of these simulations will include a detailed evaluation of the coherence between time of day and number of bird strikes.

With regard to the discussed results considering bird strike rate, altitude distribution as well as the seasonal course of bird strikes, the risk for collisions between birds and aircraft is modelled adequately in the developed simulation set-up.

### IV. Conclusions

The aim of the presented work was to develop and verify a fast-time simulation environment to analyse the risk of bird strikes in the arrival and departure corridors of an airport. For this purpose, information from two radars was -merged for the first time to generate bird movement information and to receive the full picture of birds flying in the extended airport area. By combining these bird movements with air traffic, the risk of bird strikes can be simulated in fast-time. Up to the author’s knowledge, this simulation set-up is unique. The verification of the set-up revealed that the simulated bird strike
rate is 2 to 3.5 times higher than in reality. Due to last-minutes escapes often occurring in reality but not modelled within the simulation, this conforms to the expectations. The correlation between number of bird strikes and number of birds seems to depend on airport opening hours and has to be addressed in future research. Especially if respecting all birds present in the lowest altitude band, the altitude distribution of bird strikes reflects international statistics appropriately. The seasonal effects on the bird strike risk are covered adequately as the comparison to real data from Eindhoven airport visualizes. In conclusion, the verification with real data demonstrates that the developed simulation environment reflects the risk for bird strikes decently. For this reason, the simulation environment and the results from the simulated scenarios are suitable to serve as baseline for future research evaluating ATM concepts for reducing the bird strike risk.

Acknowledgements

We thank the Royal Netherlands Air Force, the Royal Netherlands Meteorological Institute, Robin Radar and Eurocontrol for providing radar data and flight plan data. We are grateful to Hans van Gasteren for constructive discussions and valuable advice. We appreciate his and three anonymous reviewer’s feedback on this paper.

REFERENCES


Performance
INITIAL IMPLEMENTATION
OF REFERENCE
TRAJECTORIES FOR
PERFORMANCE REVIEW

Enrico Spinielli*, Rainer Koelle*
*EUROCONTROL
Performance Review Unit
Brussels, Belgium
enrico.spinielli@eurocontrol.int
rainer.koelle@eurocontrol.int

Massimiliano Zanin†‡, Seddik Belkoura†
†Technical University of Madrid
Madrid, Spain
massimiliano.zanin@ctb.upm.es
‡INNAXIS Research Institute
Madrid, Spain
sb@innaxis.org

Abstract — A reproducible and fair reporting of (European) Air Navigation Service Providers’ performance needs to rely on open, curated and methodologically sound sets of aviation data. The work by EUROCONTROL Performance Review Unit (PRU) to combine Automatic Dependent Surveillance – Broadcast (ADS–B), Correlated Position Report (CPR), Airport and Network Manager (NM) data will allow for the definition of common open datasets, in particular flight trajectories, for research and post-ops analysis.

I. Introduction

Aviation and air traffic management related information is no longer a “behind closed doors” phenomenon. In fact this data becomes increasingly ubiquitous through crowdsourcing and community efforts. Throughout the recent years, the adoption of dependent surveillance technology in aviation and inexpensive receiver sets combined with ubiquitous and cheap broadband internet connectivity resulted in a variety of community sharing networks by aviation enthusiasts. These networks are making flight position reports massively available to industry, academia, the media and the general public. This development has lead to a number of non-commercial and commercial flight tracking applications (e.g. Flight Aware, Flight Radar 24).

At the same time, other flight information (airline, schedules, airspaces & routes, aircraft types, etc.) is similarly becoming more and more available. For example airlines offer respective web services, schedules (i.e. arrival and departure times) are obtainable from webpages via web scraping, or this information is readily collected and prepared by aviation enthusiasts.

While the information becomes readily available, there is a lack of practical implementations of an open data approach. In particular, this data can be fused with data processed by air traffic management in combination with surveillance data collected by
Performance

This paper addresses the conceptual building blocks of such an open data approach to establish a reference trajectory for operational performance review purposes. Identifying the need for change, the architecture of the proposed Reference Trajectory Dataset infrastructure is developed.

II. ANSP performance analysis

The evaluation of Air Navigation Service (ANS) performance in Europe is not a fundamentally new topic. EUROCONTROL initiated an independent performance review system, governed by the Performance Review Commission, in 1997 [1]. The PRC is supported by the Performance Review Unit (PRU). The PRU is responsible for the day-to-day activities of the PRC work programme, including the regular preparation of performance data products. Next to the yearly performance review reports, the PRU publishes performance related data on a monthly basis (c.f. https://ansperformance.eu). Since the very beginning, the PRC’s performance review mandate has been to impartially draw the attention to (and try to explain) trends of excellence that showcase and drive upwards safety levels, operational and financial efficiency.

In 2004 the European Commission (EC) developed the legal framework of the Single European Sky (SES) initiative and adopted four Regulations (i.e. SES I package) covering the provision of air navigation services (ANS), the organisation and use of airspace and the interoperability of the European Air Traffic Management Network (EATMN). Later in 2009, these were revised and extended to establish a mandatory regulatory performance scheme for SES Member States (the SES II package) [3, 4]. The SES performance scheme – currently in its second reference period (2015-2019) – applies performance metrics developed by the PRC and EUROCONTROL performance review system. Similar to the PRC, the Performance Review Body (PRB) supports the European Commission in the execution of the performance scheme by providing policy advice and recommendations.

Both performance review systems are intended to drive economic, operational and societal – in particular safety and environment – improvements in the European aviation system.

Next to the European efforts, ICAO promotes a performance based approach on a global level. Throughout the recent years, regional efforts have been integrated into a wider framework under the Global Air Navigation Plan (GANP) [5]. At the time being, there are 16 so-called proposed ICAO GANP key performance indicators. These indicators are based on the European experience of PRC and build on a core set of indicators regularly utilized for the US/Europe operational ANS performance benchmarking.

---

1. **Air Navigation Service (ANS)** refers to the totality of services provided in order to ensure the safety, regularity and efficiency of air navigation and the appropriate functioning of the air navigation system.

2. Performance review is carried out for EUROCONTROL’s 41 Member States.

3. The **Performance Review Commission (PRC)** was established in 1998 by EUROCONTROL’s Permanent Commission. It provides objective information and independent advice to EUROCONTROL’s governing bodies on European Air Traffic Management (ATM) performance, based on extensive research, data analysis and consultation with stakeholders. Its purpose is “to ensure the effective management of the European air traffic management System through a strong, transparent and independent performance review,” as stated in Article 1 of the PRC Terms of Reference and Rules of Procedure.

4. The **Performance Review Unit (PRU)** supports the PRC by running EUROCONTROL’s Performance Review System and executing the associated PRC work programme.

5. **Single European Sky (SES)** applies to EU’s 28 Member States plus Norway and Switzerland. See [2] for further details.
In summary, performance review by PRC and PRB/EC aims to:

- assess the status of the system against what was planned (PRB/EC);
- highlight trends and some relevant explanatory variables (PRC and PRB/EC);
- allow for comparative analysis and showcase best in class performers others can learn from (PRC) and financially reward above threshold results (PRB/EC).

A. Reference trajectory performance data

Quality of performance related data is one of the key factors that impact the quality of overall performance review and analysis performed at EUROCONTROL PRU. One of the guiding principles of PRU’s approach is transparency in the data processes and performance indicator calculation. The ultimate goal is to enable stakeholders to reproduce the numerical performance results.

The PRU is working hard to make available curated flight data that will allow, for example, development of a reference trajectory data set which will in turn:

- foster an open and collaborative approach to performance review for its Member States and stakeholders,
- facilitate the production of studies in collaboration with International Partners (e.g. Brasil, Singapore, China, Japan)
- define a foundation for comparative studies between different world regions (e.g. EU-USA).

B. The need for change

Trajectories are the building block for a variety of operational performance metrics reported by the PRU. They are used to find airspace (sectors, FIR) intersections in order to count the number of flight at various time intervals, to calculate CO2 emissions and to assess horizontal/vertical flight efficiency or traffic complexity.

The need for transparency, openness to scrutiny and reproducibility of performance indicators calls for the publication of, not only, methodological approaches [6, 7, 8, 9, 10] and final results [11, 12, 13] but also of the relevant data used as input for the calculation of performance metrics.

Based on the aforementioned principles, the goal is therefore to establish an infrastructure to support the performance related data processing by external stakeholders. This includes the access and availability of the underlying data (e.g. reference trajectory, relevant aeronautical information), respective performance algorithms, and results.

---

6 The Performance Review Body (PRB) is an advisory body to the European Commission. It assists the Commission and National Supervisory Authorities (NSA) in the implementation of the performance scheme for air navigation services.

7 A Flight Information Region (FIR) is a specified region of airspace in which a flight information service and an ALeRtng Service (ALRS) are provided.
Current PRU metrics use trajectories as assembled by the Network Manager\(^8\) (NM) both at the pre-tactical, FTFM\(^8\) or Model 1, and tactical stage, CTFM\(^10\) or Model 3 and CPF\(^11\) profile [see [15] 14.3 and 14.4 for NM’s Tactical Flight Models (TFM) and handling of CPRs.] These profiles are devised and calculated by NM to fulfill its ATFCM mandate: slot (and hence delay) allocation and sector load monitoring. Hence the way a flight, an airspace, an aerodrome or a route are modeled is driven by the above goals and the need to keep the NM systems design robust/performant (memory consumption, CPU load) and maintainable (logically simple without bloated requirements yet useful.) For example SID\(^12\) and STAR\(^13\) concur very simply in the calculation of NM trajectories, i.e. the impacted portion of trajectory is calculated as a straight segment from point fix to the runway rather than a curved line. Similarly the left/center/right (when is the case) runways and eventually the relevant marked positions (threshold, touchdown area) are not used in the construction of the flight profiles, see for example Figure 1).

**FIGURE 1 —** Comparison of CTFM, CPF, and ADS-B trajectories for flight SWR563, LFMN (Nice, France) - LSZH (Zurich, Switzerland), on 2017-07-15. The arrows show artifacts in CPF due to position reports from 2 overlapping radars.

Furthermore the low CPR\(^14\) rate (approximately 1 every 30 seconds, see Figure 2) and reduced geographical coverage, can suffice for NM’s operational purposes but limits the granularity and extent of performance analysis.

---

\(^8\) The Network Manager [NM] administers air traffic management network functions (airspace design, flow management) as well as scarce resources (transponder code allocations, radio frequencies.) The European Commission nominated EUROCONTROL as the Network Manager in July 2011 [14] with a mandate that runs till December 31, 2019.

\(^9\) The Filed Tactical Flight Model (FTFM) or Model 1 is a flight trajectory constructed (by the ETFMS system of NM) from the last filed flight plan.

\(^10\) The Current Tactical Flight Model (CTFM) or Model 3 is a flight trajectory constructed (by the ETFMS system of NM) to tactically represent a flight being flown. It refines the previous Tactical Flight Models when CPRs show a significant deviation (1 min in time, more than 400 feet in en-route phase, more than 1000 feet in climb/descent phase or more than 10 NM laterally) and/or upon message updates from ATC (IDCT, level requests, FPL update), see 14.3.1 [15].

\(^11\) The Correlated Position reports for a Flight (CPF) is a trajectory constructed (by the ETFMS system of NM) from CPRs (and ADEP/ADES.)

\(^12\) A Standard Instrument Departure (SID) route is a standard ATS route identified in an instrument departure procedure by which aircraft should proceed from take-off phase to the en-route phase.

\(^13\) A Standard Arrival (STAR) route is a standard ATS route identified in an approach procedure by which aircraft should proceed from the en-route phase to an initial approach fix.

\(^14\) Correlated Position Report (CPR) is a radar position report from Air Traffic Control which contains information about the flight it is associated to.
Especially around airports and within the terminal airspace a higher fidelity of the reference data is required, see for example Figure 3. An increase in positional information benefits current Performance Indicators (PI) like the additional ASMA time or new ones like vertical flight efficiency [8, 9] for which the identification of holding patterns, point merge procedures, level flight segments, etc. can help to better characterize operational performance.

FIGURE 2 — CPR reception map - May 2017 (source NM)

FIGURE 3 — Comparison of CTFM, CPF, and ADS-B trajectories for flight SWRS63, LFMN (Nice, France) - LSZH (Zurich, Switzerland), on 2017-07-15. Note the higher density of black dots (position reports) from FR24. Also see how CTFM *misses* the side of takeoff and the holding over Zurich.
Actual flown trajectory reconstruction using ADS-B\textsuperscript{15} data is being pursued by other researchers, e.g. [17]. However, augmenting ADS-B with both CPR’s and airport data will allow to fill the ‘last miles’ gap and correctly link the enroute part of the trajectory with the departure and approach portions.

In addition, ground ADS-B positions and aircraft movement data as reported by airports (around 130, see Figure 4) can provide essential information to further model and analyse the surface operations from take-off/touchdown to the relevant terminal/gate/stand in order to better characterize taxi-in/out times and delays.

Moreover, the wider geographical coverage of ADS-B and higher rate (up to 1 position report every 5 seconds for FR24 live feed, Figure 5) complement the generally more accurate CPR’s further enhancing both en-route, arrival and departure portions of the trajectory.

While the application of such a reference trajectory can be immediately linked with operational performance analyses, further data analytical applications will benefit from this approach as well. For example, similar to what is happening in the media [18, 19, 20], the availability of the reference trajectory data could trigger innovative analysis by both industrial actors and academia.

\textbf{FIGURE 4 — Airports providing flight information}

\textsuperscript{15} Automatic \textbf{D}ependent \textbf{S}urveillance – \textbf{B}roadcast (ADS–B) is a surveillance technology in which an aircraft determines its position via satellite navigation and periodically broadcasts it [16].
III. Methodology

For the purpose of this work, data fusion and mining of data from FR24 and other ADS-B providers, NM and airports is performed in order to synthesize a gate-to-gate reference trajectory, c.f. Figure 6.

Significant time instants are also computed for later use in performance metrics calculation:

- $t_{ob}$: off-block time,
- $t_{to}$: take-off time,
- $t_{toC}$: top-of-climb time,
- $t_{toD}$: top-of-descent time,
- $t_{td}$: touchdown time.
The resulting reference trajectory is made of 5-sec synthetic position reports. The 5-sec interval is chosen to address data fidelity requirements.

The production of a synthesized reference trajectory for performance analysis can be abstracted as a data-analytic process. Such a process extends from the data sources, through a series of processing steps, to the exploitation of the results (e.g. dissemination of metrics, access to data). The implementation of this process requires a dedicated Reference Trajectory Dataset (RTD) infrastructure (c.f. Figure 7.)

![RTD infrastructure logical block diagram](image)

This infrastructure is composed of four parts:

1. data feeds processing;
2. trajectory synthesis;
3. repository;
4. dissemination.

### A. Data

The first block, data feeds, is where raw source data is collected, stored and an initial batch of sanity checks is performed. This is where the system verifies that all required data is available for each day, logging and eventually triggering alerts for missing and/or partial and/or corrupted data.

The current feeds used by the PRU are:

- FlightRadar24’s live feed, geographically covering all EUROCONTROL Member States;
- CPR’s, as received and processed by the Network Manager;
- Airport reported movements, as part of EUROCONTROL PRU & CODA\(^\text{16}\) data collection process.

With a view to augmenting the data, PRU is contacting various groups of enthusiasts in order to extend, diversify and complement the above mentioned feeds. The inspiration for this approach comes from Natural Earth (NE) \(^{21}\) and OpenStreetMap (OSM) \(^{22}\) where crowd-sourcing has been successfully applied to establish high quality data sets.

The goal is to aggregate different sources under a crowdsourcing model where enhancements, contributions, and updates can be managed in a distributed and non centralized way. Work is on-going to identify sources for the extension of the trajectory data with items such as airspaces (administrative [FIR] and operational [Elementary Sectors]), airports (especially aprons, gates/terminals), aircraft (airframe info, owner, registration, etc.)

---

\(^{16}\) The Central Office for Delay Analysis (CODA) provides policy makers and managers of the ECAC air transport system with timely, reliable and comprehensive information on the air traffic delay situation in Europe.
B. Trajectory synthesis

The second process block in the RTD framework is the kernel of the system. In this step, ADS-B position reports are combined with CPR’s, flight information and aircraft data is matched, and further airport flight movements are checked to enhance the trajectory data. The cloud in Figure 7 represents the data fusion process which is driven both by rules and by heuristics on the input data elements and relationships. The final stage/step samples the fused trajectory at 5 seconds intervals in order to build the reference trajectory.

C. Repository

The third block deals with the storage of the data sets. This capability allows saving (version controlled) releases of the data sets as well as data sets that are under development (e.g. missing data, new combinations of data sources.)

One or multiple relational databases with spatial capabilities are the candidates for this stage. Additional data sets stemming from RDT, e.g. airspace intersections, can be produced in this block.

As for previous blocks, PRU is investigating the use of cloud based offerings for both raw data storage and processing, the output data sets storage and version control, and data processing.

D. Dissemination

The final block deals with the sharing aspects of the project. Inspired by NE and OSM, PRU aims at providing Github repositories for the released versions of the data sets.

Other outreach possibilities are web services, Amazon Web Services (AWS) Public Dataset\textsuperscript{17} (AWSPD) and Google Public Data\textsuperscript{18} (GPD).

IV. Results and discussion

A. Feasibility Study

In 2016 PRU launched a feasibility study contract aimed at exploring the opportunities offered by cloud providers for aviation performance monitoring analysis. In particular the goals of the project were to investigate solutions for:

- raw data storage and management;
- controlled and monitored sharing with external stakeholders of data sets
- computational resources;
- provision of datasets via web API.

\textsuperscript{17} large datasets made publicly available on AWS can be analyzed using AWS compute and data analytics products, see https://aws.amazon.com/public-datasets/.

\textsuperscript{18} Public Datasets on Google Cloud Platform are freely hosted and accessible using a variety of data warehouse and analytics software, from open source ones to Google technologies, see https://cloud.google.com/public-datasets/.
The proposed use case for the project was the provision of trajectory intersections with airspaces given CPRs, ADS-B data, airport movements and airspace definitions.

The contract was awarded to and performed by INNAXIS Foundation & Research Institute. The work started in the last and finished in first few months of 2016 and 2017 respectively.

**Architecture**

The proposed PaaS, Amazon Web Services (AWS), has deemed the best one for the moderate load needs of the study. Other providers such as Google Cloud or Microsoft Azure have not been selected on the ground that they are targeting more high load applications.

The solution for storage has been as follows:

- **processed data storage and collection**: AWS RDS node (T2 medium instance, 2 virtual CPU and 4 GB of RAM, auto resizable SSD performance.) This is the node hosting a clustered Amazon Aurora database (based on MySQL) with InnoDB as storage engine.

The solution for the data processing and analysis has been as follows:

- **AWS EC2** (T2 burstable large size instance, 2 virtual CPUs, 8 GB of RAM, general purpose SSD of 300 GB.) This is the node hosting the installed OS (Linux), libraries and tools (Python: numpy, pandas, shapely, geopandas, flask) for data analysis and web API.

Other services used are AWS API Gateway for API version management, AWS IAM for user profile and authorization management and AWS S3 for initial storage and audit of raw data.

**Algorithms**

The development of the use case considered the merge of the ADS-B data and CPRs with a point pruning strategy aiming at reducing the global trajectory error, $E$.

Given a trajectory, $T$ composed of $n$ nodes (i.e. position reports), where $n_i$ is the $i$-th node, the following greedy algorithm is applied in order to prune noisy position reports:

(A) $E = total\_error\ (T)$
(B) for $i$ in $T$
    remove node $i$ from $T$
    $E_i = total\_error\ (T)$
(C) if $(\min \ (E_j) < E)$
    permanently delete node $j$ from $T$
    $E = E_j$
    goto (B)

Figure 8 shows for 9 trajectories the evolution of the error with respect the number of point removed (without stopping at the local minimum.)
The error, $E$, for a trajectory is calculated as

$$E = \frac{1}{t_{2} - t_{1}} \sum_{i} \left( \hat{\Theta}_{i}(t) - \hat{\Theta}_{i}(l) \right) \frac{t - t_{i}}{t_{i} - t_{i-1}}$$

where $t$ represents the time, and spans between the time stamps of the first and last trajectory points; $\Theta_{i}(t)$ and $\Theta_{i+1}(t)$ are two functions respectively yielding the offset of the trajectory points defining the segment in which $t$ is contained; $\hat{\theta}$ is the vector of the error at each point of the trajectory.

The error $\hat{\theta}_{i}$ at each point $i$ in the trajectory is calculated as follows:

$$\hat{\theta}_{i} = \min \left( \hat{\theta}_{i} - \frac{\gamma}{2} (\hat{\theta}_{i+1} + \hat{\theta}_{i-1}), 0 \right)$$

where $\gamma$, in the range $[0, 1]$, modulates the error propagation from the two neighboring points; with $\gamma = 0$ adjacent points’ errors have no influence.

Procedurally $\hat{\theta}_{i}$ is calculated as follows:

1. sort all points by $e_{i}$
2. calculate $\hat{\theta}_{i}$
3. (if $\gamma \neq 0$) repeat step 2. until $\hat{\theta}_{i}$ converges.
Figure 9 shows how is computed the estimation of the error $e_i$ between the position report $p_i$ and probable position $H$ when it is assumed that $v_{i-1} = v_{i-2}$ and $v_{i+1} = v_{i+2}$, i.e. that the aircraft has maintained a constant dynamic in the short time windows $[t_{i-2}, t_i]$ and $[t_i, t_{i+2}]$ respectively.

The depicted case is for the when the two circles of radius $\hat{d}_{i-1} = \hat{v}_{i-1}(t_i - t_{i-1})$ and $\hat{d}_{i+1} = \hat{v}_{i+1}(t_{i+2} - t_i)$ intersect in $H$, otherwise the intersection point is assumed to lie somewhere in between $p_{i-1}$ and $p_{i+1}$ at:

$$x_H = x_{i-1} + \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}} (x_{i+1} - x_{i-1})$$

$$y_H = y_{i-1} + \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}} (y_{i+1} - y_{i-1})$$

The proposed procedure has been studied with synthetic trajectories (with addition of noise both in lon/lat/elevation and time) to validate its soundness and theoretical characteristics.

Observations

The whole pruning procedure is nicely filtering out noisy points and able to reconstruct real trajectories, but it can be quite computationally onerous given it has a cost proportional to $O(n^2)$ for $n$ points in the trajectory: every point has to be checked for possible deletion via calculating $E$ which has a linear cost in $n$.

B. Additional uses and cross-fertilisation

The availability of RTD, still at an early stage of inception, is an essential and useful starting point for governmental, industrial and academic groups. An open data set of reference trajectories makes it possible to reproduce analyses, e.g. to review ANS performance by PRU, to research new solution on a consistent foundation of aviation data, to assess different solutions proposed in different research project, etc.

Future developments will research alternative fusion and reconstruction procedures in order to reduce the computational cost.

Further investigation would also consider the use of trajectory predictive models (for example SESAR-funded DART project) to fill the inevitable gaps after the fusion of the various input data feeds described above. For example, the (predictive) model built from machine learning from past flight trajectories from/to the same aerodrome could be able to contribute the missing portions of the trajectory that needs to be reconstructed.

V. Conclusions

The need within PRU for reference trajectories stems from the intention to better characterize performance at European scale especially in the terminal area for topics such as vertical flight efficiency or the effect of holdings on additional ASMA time whereby existing trajectories do not support finer granularity analysis.

---

19 the Single European Sky ATM Research (SESAR) project was launched in 2004 as the technological pillar of the SES. It is the mechanism which coordinates and concentrates all EU research and development (R&D) activities in ATM.

20 Data-Driven Aircraft Trajectory Prediction Research (DART) aims to present to the ATM community an understanding on what can be achieved today in trajectory prediction by using data-driven models, also accounting for network complexity effects. For more details see http://dart-research.eu/the-project/.
But the usefulness of a reference trajectory dataset goes beyond the needs of PRU: just looking at SESAR projects the same task of collecting, cleaning and validating flown flights data is repeated over and over again for each project, with all the limitations of having access to a limited temporal and/or geographical area (typically only the one covered by the partner ANSP) and with the difficulty of not being able to share such data as soon as the project partners’ composition changes in the followup or subsequent phase of the research. Another example of application of a reference trajectory dataset is the possibility to compare and evaluate different (research, industrial, operational) solutions given the same traffic.

This paper reports on conceptual and initial work of the PRU in developing a data analytical process chain for the production of a reference trajectory data set. The production of such a trajectory dataset is based on the augmentation of air traffic management data with open source and crowd-sourced data. The resulting data set shall be made available to Member States, stakeholders, academia and the public.

To adhere to the goals of transparency and openness the input data sets will be referenced and made available when the data provider’s licence allows for it.

This paper reflects the design and specification stage of a project within PRU. Work is on-going to implement and refine the described processing infrastructure. Additionally, PRU is establishing contact with a variety of open/crowd sourced projects to investigate the modalities of data fusion and sharing through the described RTD framework. Next to this fundamental capability building steps, there is a need to provide documentation for the generation of the synthesized trajectory, the respective data cleaning and verification steps. The overall goal is to make the reference data sets openly available and to maintain the underlying infrastructure and sharing processes.

Acknowledgment

The authors would like to thank their colleagues, in particular Tamara Pejovic, for the vibrant exchange of ideas, constructive feedback and support.

REFERENCES


USER-CENTRIC COST-BASED FLIGHT EFFICIENCY AND EQUITY INDICATORS

Javier López-Leonés, Manuel Polaina Morales
Boeing Research and Technology Europe
Madrid, Spain
{javier.lopezleones, manuel.polainamorales}@boeing.com

Pablo Sánchez Escalonilla, Damían Ferrer Herrer, Marcos Sanz Bravo, Fernando Celorio Cámara, Ángel Martínez Mateo
CRIDA, ATM R&D Reference Center
Madrid, Spain
{psescalonilla, dfherrer, msbravo, fccamara, amartinezm}@e-crida.enaire.es

Abstract—Flight efficiency is a generic term that varies depending on the agent’s viewpoint. Whereas Air Navigation Service Providers (ANSPs) take a wider look at efficiency, considering components such as sector capacity, air traffic controller’s interventions, emissions and noise; airlines are mainly concerned on costs, i.e. fuel consumption and schedule adherence. It is important to bring these two agents’ viewpoints together in new Key Performance Indicators (KPIs) in order to capture airspace users’ needs without leaving out the inefficiencies of the entire net.

The current implementation of efficiency measurement (as defined in the SES Performance Scheme) affects the ANSPs view on efficiency since the ANSPs have to report on specific KPIs to evaluate their performance and management of the air traffic. This implementation takes into consideration only the horizontal portion of the flight, measuring the excess horizontal en-route distance compared to the orthodromic. This approach lacks of important information for airspace users’ objectives since it leaves out the vertical component of the flight or wind conditions.

In order to introduce the airspace users’ objectives into the global net efficiency measurement, it is key to develop advanced metrics that consider fuel consumption, schedule adherence and cost of flights. These new efficiency metrics require the design of user-preferred trajectories as the main references for performance comparison purposes. Additionally, airspace users are claiming for equity metrics showing how these inefficiencies are distributed between them in certain areas such as FIRs or city-pairs.

This paper presents the methodology followed for the design of advanced user-centric cost-based efficiency and equity indicators as well as a flight efficiency and equity assessment of the European traffic flow in two particular days in February 2017 taking into consideration the airspace users’ perspective.

This research was conducted under the AURORA project (Grant 699340) supported by SESAR Joint Undertaking under European Union’s Horizon 2020 research and innovation programme. AURORA aims to propose new metrics to assess the operational efficiency of the ATM system and to measure how fairly the inefficiencies in the system are distributed among the different airlines.

The AURORA consortium is formed by Centro de Referencia I+D+i ATM (CRIDA), Boeing Research and Technology Europe (BR&TE), Centre for Applied Data Analytics Research (CeADAR) and Flight Radar 24 (FR24) with the support of Iberia, Air Europa, KLM and Turkish Airlines as members of the AURORA’s Airspace Users Group.

KEYWORDS
- component
- Airlines
- ANSP
- Flight Efficiency
- KPI
- Air Traffic Management
- SESAR
- ADS-B
I. Introduction

The Single European Sky [SES] Performance Scheme [1, 2] is designed to drive and steer the continuous improvement of European Air Traffic Management (ATM) performance. This Performance Scheme establishes EU-wide targets for four Key Performance Areas [KPAs]: Safety, Cost-Efficiency, Capacity and Environment. These overall targets, which are reviewed and updated periodically, are transposed into binding national/FAB (Functional Airspace Block) targets that are incorporated into national/FAB performance plans. The Performance Scheme, established by the Performance Review Unit (PRU), defines a set of Key Performance Indicators (KPIs) for each of the KPAs. These metrics, which are obtained through air traffic-related data [5] [15] [16] [17], allow evaluating the aggregated performance of the European ATM services and their impact on airspace users without explicitly taking into account their requirements [20].

This set of KPIs is not thought out to be static. New indicators and techniques are being continuously researched as means to improve the understanding of the ATM system. Following this trend, EUROCONTROL, on behalf of the European Union (EU), invests on researches that will allow further improvement on the system measurement [3, 4]. In addition to the research made in this area, the EU publishes reports with analysis and recommendations for the ATM system on a particular year [5]. Joint reports with the Federal Aviation Administration (FAA) are also published to compare both systems and to identify best practices for the optimization of the ATM performance [6]. Being able to better understand how these new practices are really addressing the real airlines’ interests is essential.

AURORA (Advanced User-centric efficiency metRics for air traffic perfORmance Analytics – www.aurora-er.eu) project is addressing the need to explore promising new performance indicators for operational efficiency, based on aircraft operators’ needs. Its scope is to investigate new indicators for flight efficiency, equity and fairness as well as to explore innovative methodologies to calculate them, not only by using historical data but also using real-time data for the on-line monitoring of efficiency. This paper presents those metrics that quantify costs of the flights together with the methodology used for their calculation based on historical ADS-B surveillance and flight plan data. Additionally, a case study is analysed, showing the potential of using ADS-B data as a mean to assess the global (origin to destination) efficiency of a flight.

II. Background

Flight efficiency indicators are currently monitored and reported by the SES Performance Scheme [8, 9] as part of the Environmental KPA defined by ICAO [1, 2]. This monitoring is done both in the U.S. and Europe [5-7] as well as in other countries such as Australia.

Flight efficiency is a generic term that can refer to different concepts and definitions. Nevertheless, flight efficiency is always considered as a relevant area under study due to the direct economic and environmental impacts it has according to well-known studies [3, 4, 10-14]. As a consequence, efficiency indicators’ monitoring is continuously growing to allow for a better understanding of the drivers of ATM flight efficiency.

Today’s mandatory KPI used by the SES Performance Scheme is the “Horizontal Flight Efficiency”. This KPI limits the calculation of flight efficiency to the horizontal component of the flight, and considers the geodesic route as the most efficient reference.

The method to calculate this indicator is named “the Achieved Distance Methodology” [15]. This methodology calculates the average En-Route additional distance with respect to the Achieved Distance, which is an apportionment of the most direct route between two airports (between the ASMA exit point of the departure airport and the ASMA entry point of the arrival airport), named the Great Circle Distance.
Some studies performed by EUROCONTROL [16, 17] have shown that this approach for the calculation of flight efficiency, based only on the horizontal component of the flight, doesn’t capture the “optimum” trajectory when considering meteorological factors or the airspace users’ operational objectives. FAA’s researchers have studied the possibility of introducing the wind as a parameter for the optimum trajectory calculation [18]. On the other hand, European ANSPs are also trying to improve the representativeness of flight efficiency indicators. As an example, NATS has developed the 3Di metric that may provide a good measure of the ATM influence on fuel efficiency [19]. BR&TE and CRIDA began exploring an innovative direction in a collaborative study using real operation data; as a result, a new methodology was explored to construct an Enhanced Flight Efficiency indicator that better captures the fuel consumption [20]. All previous studies showed that the existing Horizontal Flight Efficiency methodology based on the Great Circle Distance trajectory does not fully capture the optimum or more efficient trajectories, which are the cornerstone for the calculations.

These findings open a new way for investigation on optimum trajectories, considering factors such as fuel consumption, flight time costs or schedule adherence. AURORA’s study takes as starting point the previous research to overcome the gaps of the today’s most common flight efficiency indicator.

III. Methodology

The evaluation of flight efficiency indicators requires the definition of several types of trajectories, each of them accounting for a loss of efficiency due to different factors. The definitions below follow the nomenclature and framework used in [20, 27] and are a subset of the reference trajectories used in AURORA:

- **Optimal Distance Trajectory (ODT)**. This is the shortest distance trajectory, the one that follows the Great Circle from origin to destination. The ODT does not consider the impact from other traffic or from any airspace structure restrictions. This trajectory is aligned with how efficiency is currently measured by SES Performance Scheme through the Achieved Distance methodology, as explained in [15] [17] [20].

- **Optimal Cost Trajectory 1 (OCT1)**. This trajectory goes from origin to destination in free flight conditions and minimising costs of fuel and flight time. It does not take into consideration any airspace or ATC restrictions. Although air navigation fees are not considered in the calculation of this trajectory, they will be considered in the cost-based indicators.

- **Optimal Cost Trajectory 2 (OCT2)**. The OCT2 differs from the OCT1 in the fact that it takes into consideration the airspace structure since it follows the horizontal path given in the flight plan. The flight plan provided by the airlines is thought to be the optimal horizontal path taking into consideration the airspace structure and air navigation fees since it comes from powerful flight planning tools used by the airlines to plan their route according to their business strategy, although, in some specific cases, airlines may file a flight plan knowing that it will not be flown beforehand and a new one could be filed once airborne.
Flight Plan Trajectory (also Procedure-Optimal Trajectory) [FPT]. This trajectory corresponds to the filed flight plan and contains all procedural constraints. This trajectory would be flown by the aircraft if no ATC tactical interventions took place.

Actual Flown Trajectory [AFT]. This trajectory corresponds to the true trajectory flown obeying objectives specified in the filed flight plan, but also considering ground delays, tactical ATC interventions and weather diversions. All these factors contribute to the actual flown trajectory being different to what was planned [the FPT].

The methodology and process followed in the calculation of AURORA’s efficiency indicators presented in this paper is summarized in Fig. 2.

**FIGURE 2 — Process followed for the calculation of new efficiency indicators**

AFT is calculated from surveillance information [ADS-B track data]. National Oceanic and Atmospheric Administration (NOAA) weather forecasts is used as the weather model and Base of Aircraft Data (BADA) 3.10 is used as aircraft performance model [24]. This process, which is named Trajectory Reconstruction, enables the acquisition of the full state vector of the aircraft, including variables that are not explicitly included in the surveillance data and are needed to analyse the efficiency of the flight, such as fuel burnt.

ODT, OCT1, OCT2 and FPT are also calculated for each flight through the Trajectory Generation process. These are trajectories never flown by the aircraft, but used as references for comparison purposes. Each indicator is then obtained by selecting and comparing the proper variables of Actual Flown Trajectory with those of the User-preferred trajectories.

In the study presented here, both processes were carried out using PERCEPT [Predictive assessment of the impact of new air traffic concepts on current operations], which is a flexible air traffic modelling tool proprietary of BR&TE [20, 21]. In PERCEPT, Trajectory Reconstruction and Generation processes rely on a common Trajectory Computation Infrastructure (TCI) that produces a trajectory using as input the initial conditions (latitude, longitude, altitude, mass, time and speed) of the flight and an aircraft intent expressed using the Aircraft Intent Description Language (AIDL). Details on the AIDL and the TCI used can be found in [21, 22, 23, 25, 26]. The main idea behind the concept of Trajectory Reconstruction using PERCEPT is to find an instance of AIDL that fits the ADS-B track and then feed the resulting aircraft intent to the TCI that integrates the full trajectory. In the Trajectory Generation process, the AIDL instance that is fed into the TCI to obtain the aircraft trajectory is created depending on the trajectory that is sought after. The AIDL instance comes from flight intent information and initial conditions. Flight intent information condenses all the restrictions and objectives that affect a particular flight that have a direct impact on the resulting trajectory. For the same origin and destination, depending if the final trajectory needs to comply with the

---

21 Aircraft intent is the information that describes how the aircraft is to be operated within a certain time interval. An instance of aircraft intent defines the aircraft behavior that has an impact on the aircraft trajectory.

22 Flight intent can be seen as a generalization of the concept of flight plan. Details on the flight intent can be found in [25].
operational flight plan or should follow an optimal profile, different instance of AIDL will be created. The complete explanation of the processes of Trajectory Reconstruction and Generation, including the optimization process used for the creation of the optimal profiles, are explained in detail in [31].

A. Definition of Efficiency Indicators

The following list presents a subset of all the indicators defined [32] and evaluated [33] in AURORA. The first indicator [KEA], which is equivalent to the one currently used by the PRU in their efficiency analysis and reports, is calculated for comparison purposes. The new indicators were identified through workshops and questionnaires completed by the airlines participating in the project. It is important to highlight that, by definition, positive higher values of all indicators imply higher inefficiencies.

▶ KEA quantifies the horizontal deviations of the Actual Flown Trajectory (AFT) in comparison with the Optimal Distance Trajectory (ODT), i.e. the geodesic trajectory.

$$\text{KEA} = \left( \frac{L_{AFT}}{H} - 1 \right) \%$$  \hspace{1cm} (1)

Where \( L_{AFT} \) is the horizontal distance flown by the aircraft, i.e. AFT horizontal distance, and \( H \) is the Great Circle Distance between origin and destination, i.e. ODT horizontal distance.

▶ CEA_CW1 quantifies the extra-costs of the Actual Flown Trajectory (AFT) in comparison with the Optimal Cost Trajectory (OCT1), which optimizes flight time and fuel costs in free route conditions.

$$\text{CEA}_{\text{CW1}} = \left( \frac{C_{AFT}}{C_{OCT1}} - 1 \right) \%$$  \hspace{1cm} (2)

$$C = p_{\text{FUEL}} \cdot (\Delta m + CI \cdot \Delta t) + RC$$  \hspace{1cm} (3)

Where \( C_{AFT} \) and \( C_{OCT1} \) are the total costs of the AFT and OCT1 respectively, both given by [3]. With \( CI = \text{cost index} \), \( \Delta t = \text{flight time} \), \( p_{\text{FUEL}} \) is the average fuel price as given in [28], \( \Delta m \) is the fuel consumption and \( RC \) are the route charges, calculated using the formula given by EUROCONTROL in [29][30].

▶ CEA_CW2 quantifies the extra-costs of the Actual Flown Trajectory (AFT) in comparison with the Optimal Cost Trajectory (OCT2), which optimizes flight time and fuel costs following horizontally the flight plan.

$$\text{CEA}_{\text{CW2}} = \left( \frac{C_{AFT}}{C_{OCT2}} - 1 \right) \%$$  \hspace{1cm} (4)

Where \( C_{OCT2} \) is the cost of the OCT2 given by [3].

It is important to clarify that, in the study presented in this paper, all the indicators are calculated from origin-destination. This implies that the calculation of KEA differs from the

23 A nomenclature was developed for a better understanding of these indicators. This nomenclature is composed of five letters, the first letter is for the variables being compared (K for distance, F for fuel, C for cost); the second letter (E) is used to indicate that these are efficiency or environment indicators; the third and fourth, separated by underscore, identify the trajectories being compared (A for actual, P for planned and D for geodesic, F for minimum fuel and C for minimum cost); the last letter is used to indicate if a weather model is considered (e.g. CEA_CW means Cost Efficiency indicator of the Actual Flown Trajectory versus the Optimum Cost Trajectory considering Weather).

24 The costs considered in this paper are those corresponding to fuel, time and air navigation fees, not considering explicitly the true cost of delay. The cost of time is only considered through the cost index, which is extracted from publicly available documents [34], [35] and [36].
current implementation indicated by the PRU to ANSPs, where the portion of the flight in an area of 40NM around the airports (ASMA) is excluded from the evaluation of the indicators [15] [17] [20]. The airlines involved in the study mentioned their interest to understand the efficiency of their flights by considering the whole trajectory, including the ASMA.

It also relevant to remark that the calculation of route charges for the different trajectories is not based on the route charges associated to the flight plan (current way to calculate navigation fees) but the route charges associated to the actual trajectory (calculated using the geodesic distance between the entry and exit point to each airspace which is crossed by the trajectory), which will be the future way to pay charges.

**B. Equity indicators**

_Equity_ metrics tend to capture how the inefficiencies of the system would be distributed between all airspace users within a certain context (e.g. ECAC region, airport, airline type or airspace crossed). In the context of AURORA five equity metrics were defined [32] and evaluated [33] and the one focused on costs is presented in this paper:

- **EQ-4** indicates the standard deviation of the mean ratio between the actual costs and the planned costs of all flights belonging to each airline. Below is the expression for calculation of the EQ-4 indicator:

\[
EQ - 4 = \sqrt{\frac{\sum_{j=1}^{n} (\frac{x_{AUj}}{x_{C}})^2}{n-1} - \left( \frac{\sum_{j=1}^{n} x_{AUj}}{N} \right)^2}
\]  

With 

\[
x_{AUj} = \frac{\sum_{\text{flights} \in AUj} C_{AT}}{\text{number of flights} \in AUj}
\]

\[
x_{C} = \sum_{j=1}^{N} x_{AUj} / N
\]

Where 

\(C_{AT}\) and \(C_{PT}\) are the cost of the AT [actual trajectory] and PT [planned trajectory] respectively, as expressed in [3], \(n\) is the total number of flights in the context under study and \(N\) is the total number of airspace users in the context.

**C. Scenario Description**

This study analyses actual ADS-B equipped flights whose whole track remains inside the European Civil Aviation Conference (ECAC) area. ADS-B data to apply the proposed methodology were needed in time intervals of less than 5 seconds. Traffic samples with the required granularity were generated starting at the beginning of 2017. Two days without major disruptions, i.e. without abnormal ATC regulations or delays, were selected: 2017 February 20th and February 24th. February 24th had higher magnitude and different predominant wind direction than February 20th.

Constraints in time processing of the reference trajectories made necessary to focus the data sets. The study considered flights departing from 12:00 to 14:00 as these are the main peak hours of the selected days. Additionally, all flights operating several city pairs along the 24 hours of the two days were also included in the data sets. These city pairs, which were identified by the members of the AURORA’s Airspace Users Group, are: London Gatwick – Madrid Barajas, London Gatwick – Barcelona, Frankfurt – Madrid Barajas, Paris Orly – Toulouse, Paris Orly – Lisbon, Istambul – Amsterdam, Roma Fiumicino – Amsterdam and Barcelona – Brussels.

This adds up to 1,583 trajectories for the 20th and 1,692 trajectories for the 24th. Fig. 3 depicts a sample of the trajectories analysed for February 20th. The analysis of flight efficiency indicators was performed with both data sets. The analysis of equity indicators was performed with February 20th set exclusively.
IV. Results

A. Flight efficiency

Table I summarizes the mean values, the standard deviation and the coefficient of correlation between the assessed indicators for the two ECAC traffic samples. Linear correlation is obtained as an indication of up to which point the behaviour of two indicators is similar. High correlation values between the new cost-based indicators with current efficiency indicator, i.e. KEA, imply that this easy-to-obtain indicator could be representative enough to estimate how cost-efficient a flight is and there is no need of defining more complex indicators.

<table>
<thead>
<tr>
<th>Days</th>
<th>Ind.</th>
<th>Mean value</th>
<th>Standard Dev.</th>
<th>Linear Correl. with KEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>20/02/2017</td>
<td>KEA</td>
<td>9.7%</td>
<td>7.6%</td>
<td></td>
</tr>
<tr>
<td>24/02/2017</td>
<td>CEA_CW1</td>
<td>9.3%</td>
<td>6.4%</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.0%</td>
<td>6.4%</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>CEA_CW2</td>
<td>4.6%</td>
<td>5.3%</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2%</td>
<td>5.1%</td>
<td>0.37</td>
</tr>
</tbody>
</table>

KEA and CEA_CW1 mean values are similar for both days. Thus, the ECAC view of efficiency in terms of costs deviations of the actual trajectory with respect to the optimal cost-based trajectory, i.e. OCT1, is similar to the deviations in distance with respect to the geodesic trajectory, i.e. ODT.

In spite of this, KEA is not properly representing how good is the actual trajectory with respect to the optimal cost-based trajectories. Testing different relations between the indicators, a linear relationship is taken as the most representative due to the higher value of correlation in comparison with other non-linear relations. The linear correlation between KEA and CEA_CW1 is around 0.70 which is identified as a medium-strong correlation according to Pearson scale. The correlation between KEA and CEA_CW2 is 0.25 and 0.35 which is identified as a small correlation according to Pearson scale. Fig. 4 shows KEA vs CEA_CW1. For similar values of KEA, CEA_CW1 scatter can be observed in the figure.
On the other hand, actual trajectories in the ECAC are more efficient than expected when comparing with the best possible cost-efficient trajectory following the flight plan, i.e. OCT2, as it can be seen in the difference between CEA_CW2 and KEA mean values. In fact, KEA and CEA_CW1 mean values are around 50% higher than CEA_CW2 in the two traffic samples. This indicates that half of the ECAC inefficiencies in terms of costs are due to the constraints of the route design.

Weather (wind, pressure and temperature) is not causing major horizontal deviations of the optimal cost-based trajectories in free route with respect to the geodesic for short and medium-haul flights. Fig. 5 shows a representative example of the horizontal path of OCT1 and OCT2 trajectories versus the AFT and ODT trajectories for flight IBE31RG from MAD to DUS. In this case the horizontal profile of the ODT and OCT1 trajectories are the same, and therefore they cannot be differentiated in the figure. This is also the main reason for the higher correlation between KEA and CEA_CW1 than between KEA and CEA_CW2 as horizontally ODT and OCT1 are very similar for short and medium-haul flights.

Wind is identified as a factor causing changes in the vertical profile, flight time and speed of the Optimal Cost Trajectories, both OCT1 and OCT2, with the subsequent impact on total flight. This effect, which is not captured by the current efficiency indicator, has a clear impact on CEA_CW1 and CEA_CW2 values. Fig. 6 shows an example of two flights, IBE481 and IBE04VM, operating the same aircraft type (A319). IBE481 is flying with tail wind from Oviedo to Madrid in the afternoon (8:00 PM). IBE04VM flies from Madrid to Oviedo in the morning (7:00 AM) with head wind. Both AFTs have the same flight duration because IBE481 AFT increases the speed to cover more distance, and consequently consumes more fuel. On the contrary, IBE481 OCT1 benefits from the tail wind, reducing flight time and maintaining the fuel consumption in comparison with IBE04VM OCT1. In conclusion, IBE481 is less efficient in terms of costs than IBE04VM as it is seen in the difference in CEA_CW1 values.
CEA_CW1 and CEA_CW2 can also change the global picture of local inefficiencies. Fig. 7 represents the inefficiencies in the southwest area of the ECAC for February 24th. For example, flights crossing Romania and Bulgaria from Istanbul have values of CEA_CW1 from 15% to 30%, while KEA values are in the range of 5% to 15%.

B. Equity

This subsection presents equity values calculated using EQ-4 indicator for the traffic sample of the 20th of February. The following figures provide the value for the equity indicator according to (5), and also the associated mean of the set to provide statistical background to the equity indicator. Depending on the context chosen (ECAC level, FIR level or route level), different conclusions can be extracted from the analysis.

Fig. 8 shows the EQ-4 calculation at ECAC level. It can be observed that EQ-4 is 4.05% while there is a mean of 3.65% of ratio between costs of the AFT and the FPT for all the flights considered in the analysis. These values serve as baseline to compare the results per region or city pair.
Fig. 9 shows the EQ-4 calculation for those flights with tracks inside four different ECAC regions: France (LF), Spain (LE), Germany (ED) and Italy (LI). According with these values and for this day, it seems that Germany represents the airspace in which the distribution among the airlines is less equitable. France is on the other extreme, with the lowest values for EQ-4. It can be also observed that its mean value of the cost ratio is not the lowest among the selected ECAC regions. Thus other regions are penalizing less the costs of flights but these other regions are less equitable between airlines.
Fig. 10 shows the equity calculations for eight city pairs, selected by the airlines collaborating on AURORA. As it can be seen, some city pairs have an average mean value of over 5%. However, when comparing IST-AMS and FCO-AMS, it can be observed that equity values are lower for the first city pair (IST-AMS). This means that the deviation in cost are less equitable distributed among the airlines flying FCO-AMS than IST-AMS. Observing the other city pairs, low equity values mean that the differences in cost between planned and actual are distributed equally among the airlines flying those city pairs. When compared with the ECAC level values, these city pairs performance better in terms of equity than the average at the ECAC level.

V. Conclusions

Due to the methodology proposed in this paper, ADS-B data could serve as a reliable source for the performance monitoring at the ECAC level, providing a new paradigm where ANSP’s performance is not only evaluated locally, i.e., at the level of an ANSP area of responsibility, but also globally, i.e., how the actions of an ANSP impact the overall efficiency of a flight and the actions of other ANSPs responsible of that flight. In terms of the value of the KPIs analysed, the following conclusions were extracted:

- CEA_CW1 and CEA_CW2 represent how costefficient are the flights with respect to a future free routing environment and using today’s route design respectively. Results show that half of the inefficiencies in terms of costs are due to the constraints in the route design.
Flight inefficiency in terms of costs is not necessarily aligned with inefficiency in terms of horizontal difference with respect to the geodesic trajectory, i.e., CEA_CW1 and CEA_CW2 values differ from KEA values.

Vertical and speed profiles together with the impact of weather conditions (wind, temperature, and pressure) are relevant factors to be taken into account in order to quantify how cost-efficient a flight is, and this is not considered in today’s indicator, i.e., KEA.

Equity indicators provide an insight on how inefficiencies are distributed among airlines, allowing the detection of regions or routes that present abnormal values comparing with some average ones.

VI. Future work

Based on the results and conclusions obtained from the analysis, some research areas are defined:

1. Perform sensitivity analysis on how the assumed parameters, e.g., initial mass, affect the results.

2. Perform a systematic analysis to test if the different values of the indicators account for different inefficiency sources (i.e., if an indicator value, or combinations between different indicators, can pinpoint the source of the inefficiency to a specific event, such as holding patterns, inefficient speed profiles, etc.).

3. Calculate the indicators per phase, considering the need of introducing the approach and ascent procedures in all generated and reconstructed trajectories to isolate the effects of TMA. It should also consider the need of defining some overlapping between phases due to differences in flight time per phase between the generated and reconstructed trajectories. It is recommended to analyse the applicability of machine learning techniques.

4. Test how to measure the efficiency of flights which are also crossing the ECAC, and not only departing and arriving since this traffic also affects the efficiency of the ECAC traffic.

5. Analyse cause and effect relationships to quantify the impact of Airspace Users’ operation modes on AURORA indicators.

6. Analyse how to include delay costs in AURORA indicators as an additional cost which is considered relevant for the airlines.

REFERENCES

Traffic prediction
DART: A MACHINE-LEARNING APPROACH TO TRAJECTORY PREDICTION AND DEMAND-CAPACITY BALANCING

Esther Calvo Fernández, José Manuel Cordero
CRIDA, ATM R&D Reference Center
Madrid, Spain

Georg Fuchs, Natalya Andrienko, Gennady Andrienko
Fraunhofer Institute IAIS
Sankt Augustin, Germany

George Vouros, Nikos Pelekis, Theocharis Kravaris, Harris Georgiou
University of Piraeus Research Center
Piraeus, Greece

Enrique Casado, David Scarlatti, Pablo Costas
Boeing Research & Technology Europe
Madrid, Spain

Samet Ayhan
University of Maryland
College Park, United States

Abstract—The current Air Traffic Management (ATM) system worldwide is managing a high (and growing) amount of demand that sometimes leads to demand-capacity balancing (DCB) issues. These further impose limitations to the ATM system that are resolved via airspace management or flow management solutions, including regulations that generate delays (and costs) for the entire system. These demand-capacity imbalances are difficult to predict in the pre-tactical phase (prior to operation), as the existing ATM information is not accurate enough during this phase. With the aim of overcoming these drawbacks, the ATM system is moving towards a new, trajectory-based operations (TBO) paradigm, where the trajectory becomes the cornerstone upon which the ATM capabilities rely on. This transformation, however, requires reliable information available in pre-tactical phase or, at least, high-fidelity aircraft trajectory prediction capabilities to reach sufficient levels of confidence in the available planning information.

In this scenario, the DART (Data-driven Aircraft Trajectory Prediction Research) project from SESAR 2020 Exploratory Research aims at reaching this goal, by means of machine learning and agent-based modeling methods in two different use cases: trajectory prediction and demand-capacity balancing. This paper presents the machine learning approach followed, as well as the promising results already achieved by the project.
I. Introduction

A. DART Project description

Within SESAR 2020 Exploratory Research, DART project has the main objective of exploring the applicability of data mining, machine learning and agent-based models and algorithms to derive a data-driven trajectory prediction capability. In addition to the expectation that data-driven techniques will enhance trajectory predictability and thus, will reduce uncertainty factors during the pre-tactical phase, agent-based modeling methods are expected to provide increased levels of accuracy while considering ATM network effects in the prediction process, which have been rarely introduced by current state-of-the-art solutions. For this, the project relies on extensive, high-quality operational datasets which support the data-driven approach.

Machine-learning algorithms with promising results, will be used for predictions in a collaborative trajectory scenario, accounting for delays due to ATM network effects. Towards an agent based modeling approach for collaborative trajectory prediction, DART leverages reinforcement learning techniques to refine predictions based on (a) potential trajectory predictions and (b) contextual information, in a coordinated way, for groups of trajectories.

In combination, the ultimate goal of DART is to demonstrate how machine learning methods can help in refining single trajectory predictions (learned from surveillance data linked to weather data and other contextual information), considering also cases where demand of airspace use exceeds capacity, resulting to hotspots. This is referred as the Demand and Capacity Balance (DCB) problem, which is the testing use case identified but not the only potential application environment of such techniques. In this work we focus on the way trajectories are affected due to the influence of the surrounding traffic (i.e., considering interactions among individual predicted trajectories), taking into account an important aspect of ATM system complexity by determining delays for affected trajectories at the pre-tactical stage in order to resolve DCB problems, so improving trajectory prediction.

So, this paper addresses (i) the DART research approach both in terms of data-driven trajectory prediction (individual) and agent-based collaborative learning applied to DCB environment in pre-tactical phase, (ii) the positive results obtained so far; and (iii) next steps of project research.

II. Background

A. Trajectory Prediction

In the context of this work, the first required step is the determination or common understanding of what a trajectory is. Basically, a trajectory is a chronologically ordered sequence of aircraft states described by a list of state variables. The most relevant ones are airspeeds (True Airspeed, TAS, Calibrated Airspeed, CAS, or Mach Number, M), 3D position (latitude $\phi$, longitude $\lambda$ and geodetic altitude $h$ or pressure altitude $H_p$), the bearing $\chi$ or heading $\psi$ and the instantaneous aircraft mass $m$. A predicted trajectory can be defined as the future evolution of the aircraft state as a function of the current flight conditions, a forecast of the localized weather conditions, contextual information regarding the airspace and a description of how the aircraft is to be operated from this initial state and so on.

Even though there might be available extremely accurate aircraft performance models, such as BADA (Base of Aircraft Data) models released by EUROCONTROL, or weather forecasts, such as those generated by the Global Forecast System (GFS) provided by the National Oceanic and Atmospheric Administration (NOAA), there are intrinsic errors that produce unavoidable deviations between predicted and actual trajectories. Those
deviations are the result of representing a stochastic process (prediction of an aircraft trajectory affected by stochastic sources) by a deterministic approach (formulation of a kinematic or kinetic aircraft motion problem).

The concept of data-driven trajectory prediction used in DART project, does not consider any representation of any realistic aircraft behavior, only exploits trajectory information recorded from the ground-based surveillance infrastructure or by onboard systems (e.g., Flight Recorded Data, FDR, or Quick Access Recorder Data, QAR) and other contextual data that may impact the final trajectory, which constitutes an innovative approach. This decoupled solution from the mathematical formulation of the aircraft motion should capture variations of the trajectory that cannot be derived directly from the filed Flight Plans [FPs], both during the pre-tactical and tactical phases. These discrepancies usually come from Air Traffic Control (ATC) interventions to ensure optimum traffic management and safe operations (e.g., delays added due the effect of adverse weather). If these interventions respond to a pattern, big data analytics and machine learning algorithms might potentially identify them once the proper system features are considered.

Thus, the preparation of available trajectory data is crucial to train the algorithms in accordance to the expected performance. Several solutions aim at predicting some aircraft state variables (Target Times) for a representative scenario. The DART goal is to assess generic prediction methods to be applied in different possible scenarios envisioned in the future Trajectory Based Operations (TBO) environment.

B. Demand Capacity Balancing

The DCB process considers two important types of objects in the ATM system: aircraft trajectories and airspace sectors, and is divided in three phases: Strategic, Pre-tactical and Tactical Phase. The overall objective is to optimize traffic flows according to ATC capacity while enabling airlines to operate safe and efficient flights.

Planning operations start as early as possible - sometimes more than one year in advance. Given that the objective is to protect ATC of overload, this service is always looking for optimum traffic flow through a correct use of the capacity, guaranteed safety, but also potentially considering other dimensions such as better use of capacity, equity, information sharing among stakeholders and fluency.

In DART research, it is considered the demand-capacity balancing process during the pre-tactical phase. Pre-tactical flow management is applied days prior to the day of operations, and consists of planning and coordination activities. This phase aims to compute the demand for the operations day, compare it with the predicted airspace capacities on that day, and make any necessary adjustments to the flight plans. Since DART goal is trajectory predictions and is focused on a TBO environment, this research considers individual predicted trajectories instead of flight plans, in order to determine the delay that should be imposed on them due to traffic.

At this pre-tactical phase, trajectories are sent to the Network Manager who takes into account sector capacities to detect problematic areas. The main objective of this stage is to optimize efficiency and balance demand and capacity through an effective organization of resources, as much as possible given the accuracy of existing information, which will be greatly improved in a TBO environment. This is done by determining delays at the pre-tactical stage in order to resolve DCB problems. Actually, the current work methodology today is based on a collaborative decision making process between the stakeholders resulting to an Air Traffic Flow Control Management Daily Plan (ADP).
III. Methodology

A. Individual (single) Trajectory Prediction

This section details the big data analytics (BDA) and machine learning (ML) algorithms applied to aircraft single trajectory prediction. The potential three candidates chosen to be assessed throughout the execution of DART have been considered as most suitable and promising techniques to tackle with the problem of data-driven aircraft trajectory prediction. The selection of these three main ML-based approaches is based on the current state-of-the-art, as well as the specifications of the problem. These options are briefly described below:

- **Hidden Markov Models (HMM):** one of the most popular and well-known approaches for studying the state transitions of a system, with applications ranging from time series analysis to speech recognition and medical diagnostics [1][6]. The HMM approach models the evolution of a system by a set of states and transitions between them, each one accompanied by a probability that is typically extracted by analyzing historic data. In the context of TP, the flight route and all the associated information are encoded into discrete values that constitute the HMM states. Then, the trajectory itself is treated as an evolution of transitions between these states, using the raw trajectory data of many flights for training, plus spatio-temporal constraints. Some very recent case studies with this approach show that its results on real data are very promising [7].

- **Trajectory prediction via appropriate kernel-based distance metrics for clustering.** Many approaches to data-driven trajectory prediction based on surveillance data makes use of the flight path itself as the feature vector and test its similarity with other tracks. In practice, the input vector can include several other properties associated with any trajectory segment but not necessarily derived from the spatio-temporal data of the trajectory. For example, each trajectory segment could be enriched with weather variables, the type of the aircraft, as well as any other semantic information that is relevant. Similar approaches have been widely used in time series classification, as well as the encoding of local spatial features in image analysis [e.g. see [8]]. In trajectory prediction, k-NN classifiers have been used extensively in similar works with trajectory data [7][9][10].

- **Advanced ML models for non-linear regression.** The current state-of-the-art in regression models for raw-data TP includes various methods from the statistical point of view, as well as some ML-based methods. More specifically, several types of localized linear regression, such as Locally Weighted Linear Regression (LWLR) [11] and Locally Weighted Polynomial Regression (LWPR) [12], have been applied to similar problems. As the scale becomes more and more local, the margin of stochastic effects becomes smaller and the regression becomes more accurate. At the same time, there are numerous robust ML algorithms [11][13]-[16] that are much more efficient than standard linear regression or variants. These include kernel-based approaches like Support Vector Machines (SVM) for regression, Decision Tree methods like Classification and Regression Trees (CART), as well as typical soft-margin classification methods like Neural Networks [10][17] that can also be used for regression of the trajectory at different levels and scales.

In this general context, DART addresses the TP task by combining elements of these three basic approaches, in order to produce innovative solutions that are: (a) purely data-driven, (b) efficient and accurate, (c) scalable to very large amounts of input data when applied in the real world (ATM).

The three main approaches, i.e., HMM, clustering and regression, are being developed in parallel and the main focus of work is currently allocated to designing a hybrid clustering/HMM two-phase algorithm for the single TP task. More specifically, clustering is applied
as a first processing phase for aircraft trajectories, using a rich set of “annotated” trajectories that include flight plans, localized weather and aircraft properties, which enable modeling in a space higher than the typical 4-D spatio-temporal trajectories domain. Clustering is applied using properly designed distance functions that implement similarity metrics for the complete N-dimensional enriched domain, thus providing a more effective matching between “similar” trajectories, not only with regard to their spatio-temporal path but also to local weather, aircraft properties, calendar properties (e.g. weekday), etc. This first phase essentially creates compact groups of aircraft trajectories, typically separating airport pairs (departure/destination), but also differences in takeoff and landing patterns and severe weather deviations even for the same flight route. Then, each group is represented by one median route or medoid, which scales down the complexity of the TP task by at least two orders of magnitude for the next phase (e.g. treating 5-8 medoids instead of 600-800 single trajectories, per month per airport pair).

Next, a hidden Markov model (HMM) is defined and trained for each cluster, using non-uniform graph-based spatial grid and exploiting flight plans as constraints for a parametric model for the HMM emission probability. More specifically, the HMM states are not defined in a uniform grid of typically 3+k dimensions, where k is the number of additional enrichment parameters (e.g. local weather) [7]. Instead, the waypoints of the filed flight plans of each specific flight are used as the reference points for the HMM states. Each of these points can be matched to the closest point of the medoid of the cluster that each flight is assigned to during the first phase (using the properly defined similarity metric). Thus, each of the individual flight plan is matched waypoint-to-waypoint to its assigned medoid and the true 3-D deviation (Haversine distance) between each pair is formulated probabilistically as the HMM emissions. In practice, instead of using the full-resolution medoid as the baseline, the waypoints of the flight plans are used for setting up the states and emissions for each HMM, one for each medoid. As a result, the complexity of the TP task is further scaled down by at least one more order of magnitude, since e.g. a 600-800 point 5-second sampling trajectory (IFS) is processed as a graph of 11-18 vertices and directional single-edge transitions.

This proposed method has been applied in real radar operational tracks and NOAA weather data for a one-month dataset of flights in Spanish airspace. Using parametric Gaussians as the base for the emissions model and confidence interval estimations for the associated errors, the proposed method exhibits exceptionally low HMM complexity and per-waypoint prediction accuracy of a few hundred meters compared to their filed flight plans submitted prior to the flight. Further enhancements are currently being developed, primarily focusing on enhancing the efficiency, scalability and optimal balance between spatio-temporal and enrichment parameters in the design of similarity metrics for the trajectory matching as k-nearest neighbors (k-NN) clustering with k=1 used Dynamic Time Warping Euclidean distance. Additionally, the regression approach is being investigated independently for extending the current state-of-the-art methods on short-range single TP.

**B. Collaborative Trajectory Prediction: Demand Capacity Balancing**

The objective is to demonstrate how agent-based modeling methods can help in trajectory forecasting when planned demand exceeds sectors capacity, taking into account interactions among trajectories, considered as self-interested agents that aim to minimize their delays and resolve demand-capacity imbalances. In this case, regulations of type C (i.e. delays) [18] are applied to the trajectories. This module deals with the trajectories provided by the previous data-driven TP.

Considering the problem specification, let there be trajectories in a set of trajectories $T$ that must be executed over the airspace in a period of $p$ time instants (e.g. hours). The airspace consists of a set of sectors $S$. Time is divided in intervals $\Delta t$, equal to the duration of the Occupancy Counting Period used for measuring demand [19].

Each trajectory is a sequence of timed positions in airspace, which can be exploited to compute the series of sectors that each flight crosses, together with the entry and
exit time for each of these sectors. For the first (last) sector of the flight, i.e. where the
departure [resp. arrival] airport resides, the entry [resp. exit] time is the departure [resp.
arrival] time. Also, there may exist flights that cross the airspace but do not depart and/
or arrive in any of the sectors of our airspace. In that case we only consider the entry and
exit time of sectors within the airspace of our interest.

Thus, a trajectory $T$ is a time series of elements of the form:

$$T = \{ (s_1, entryTime_1, exitTime_1), (s_2, entryTime_2, exitTime_2), \ldots, (s_m, entryTime_m, exitTime_m) \}$$ (1)

where $s_i, i=1,..,m$ are sectors in $S$.

For instance, considering the trajectories $T_1$, $T_2$ and $T_4$ in Figure 1, these are specified
as follows:

$$T_1 = \{ (s_5, 10:00, 10:20), (s_2, 10:20, 10:45) \}$$ (2)

$$T_2 = \{ (s_5, 10:15, 10:30), (s_6, 10:30, 10:34), (s_7, 10:34, 11:00), (s_{12}, 11:00, 11:27) \}$$ (3)

$$T_4 = \{ (s_{12}, 12:00, 12:10), (s_{15}, 12:10, 12:25) \}$$ (4)

**FIGURE 1 — Example of trajectories crossing sectors**

This information per trajectory suffices to measure the demand $D_{sp}$ for each of the
sectors $s_i$ in $S$ in the airspace in any Occupancy Counting Period $\Delta t$ of duration $\Delta t$.
Specifically, $D_{sp} = |\bigcap_{p} T_{sp}|$, i.e. the number of trajectories in $T_{sp}$.

In other words, the demand equals to the number of trajectories co-occurring over of a
period $p$ in the same sector. For instance, considering the trajectories $T_1$ and $T_2$ and
crossing the sector $s_2$ in Figure 1, it holds that $T_{s_2} = \{T_1 \cup T_2\}$ with $p = [10:00, 10:25]$. The
trajectories in $T_{s_2}$ are defined to be **interacting trajectories** for the period $p$ and the
sector $s_2$.

Each sector $i$ has a specific capacity $C$ over a period. The aim is to resolve imbalances of
sectors’ demand and capacity. These are cases where demand $D$ exceeds capacity $C$, for any
period $p$ of duration $\Delta t$ (occupancy count period duration) in $H$, in any of the sectors $s_i$ in $S$.

Subsequently we refer to these cases as demand-capacity imbalance cases, resulting
to hotspots.

In case of imbalances for a period $p$ and sector $s_i$, the interacting trajectories in $T_{s_i}$ are
defined as hotspot-constituting trajectories: one or more of these trajectories must be
delayed in order to resolve the imbalance in $s_i$. Given the exploratory research nature of
DART, at this stage of research no 4D measures are considered for hotspot resolution, just delays. Enhanced context of research foresees 4D measures.

This problem specification emphasizes on the following problem aspects: (a) agents, corresponding to a single trajectory, need to coordinate their strategies (i.e. chosen options to impose delays) to execute their trajectories jointly with others, taking into account traffic, operational constraints; etc... (b) agents need to explore and discover how different combinations of delays affect the joint performance of their trajectories in terms of the DCB process, given that the way different trajectories do interact is not known beforehand. Agents do not know the interacting trajectories that emerge due to own (and others) decisions, and of course they do not know whether these interactions result to new hotspots; and (c) agents' preferences on the options available may vary depending on the trajectory performed, and are kept private.

In principle, a collaborative multi-agent Markov decision problem (MDP) can be regarded as one agent in which each joint action is represented as a single action. However this may result to a huge state-action space and thus to high computational complexity. So, in order to exploit its various advantages, we use the model of collaborative multi-agent MDP framework [20][21] which assumes:

- The society of agents, where each agent $A_i$ corresponds to a trajectory and is connected to a set of agents (denoted by $N(A_i)$) corresponding to interacting trajectories, resulting to a graph $(A,E)$, where $A$ is the set of agents and $E$ the edges between them.

- A time step $t=1,2,...,H$, where $H$ is the total number of time instants considered.

- A local state per agent $A_i$ at time $t$, comprising state variables that correspond to (a) the delay imposed to the trajectory $T_i$, ranging to the sets of options assumed by $A_i$, and (b) the number of hotspots in which $A_i$ is involved in (for any of the sectors and time periods). Such a state is denoted $s_t$. The joint state $s_{t(j)}$ of agents $A_i$ and $A_j$ at time $t$ is the tuple of the state variables for both agents. A global state $s_t$ at time $t$ is the tuple of all agents’ local states.

- The local strategy for agent $A_i$ at time $t$, denoted by $strt_i$ is the action that performs at that specific point: An action for any agent at any time point, in case the agent is still on ground, may be, either impose a delay or not. Thus, at each time point the agent has to take a binary decision. When the agent flies, then it just follows the trajectory. The location (i.e. sector) of that agent at any time point can be calculated by consulting its trajectory. The joint strategy of a subset $Ag$ of agents executing their trajectories at time $t$, is a tuple of local strategies, denoted by $str_{Ag}$. The joint strategy for all agents $A$ at time $t$ is denoted $str^t$.

- The state transition function gives the transition to the joint state $s^{t+1}$ based on the joint strategy taken in joint state $s^t$. It must be noticed that although this transition function may be deterministic in settings with perfect knowledge, the state transition per agent is stochastic, given that no agent has a global view.

- The local reward of agent $A_i$, denoted $Rwd_i$, is the reward that the agent gets by executing its own trajectory in a specific joint state of its peers in the society (i.e. the agents) according to the sectors’ capacities, and the joint strategy of agent involved. The joint reward for a set of agents specifies the reward received by involved agents by executing their actions in their joint state, according to their joint strategy. It depends on the number of hotspots occurring while the agents execute their trajectories according to their joint strategy in their joint state, i.e. their decided delays, and also according to their preferences on the chosen delays while performing jointly.
A **local policy** of an agent $A_i$ is a function $\pi_i: \text{state} \rightarrow \text{strategy}_{Ai}$ that returns local strategies for any given local state, for $A_i$ to execute its trajectory. The objective for any agent in the society is to find an optimal policy $\pi^*_i$ that maximizes the expected discounted future return for each state $s$, while executing its trajectory. This model assumes the Markov property, assuming also that rewards and transition probabilities are independent of time.

The next paragraphs describe three collaborative reinforcement learning methods that take advantage of the problem structure, considering that agents do not know the transition and reward model (model-free methods) and interact concurrently with all their peers.

- **Independent Reinforcement Learners (Ind-Colab-RL):** The independent learners Q-learning variant proposed in [22] decomposes the global Q-function into a linear combination of local agent-dependent Q-functions. Each local $Q_i$ is based on the local state and local strategy for agent $A_i$. Dependencies between agents, and thus the coordination graph, are defined according to the agents’ society specified above. It must be pointed out that these dependencies may be updated while solving the problem. Each agent observes its local state variables. A local $Q_i$ is updated using the global temporal-difference error, the difference between the current global Q-value and the expected future discounted return for the experienced state transition. As opposite to [22], we use the reward received by the agent, taking into account only the joint state and joint strategy of its neighborhood.

- **Edge-Based Collaborative Reinforcement Learners (Ed-Colab-RL):** This is a variant of the edge-based update sparse cooperative edge-based Q-learning method proposed in [1]. Given two peer agents performing their tasks, $A_i$ and $A_j$, the Q-function is denoted succinctly $Q_{ij}(s_{ij}, str_{ij})$, where $s_{ij}$ with abuse of notation denotes the joint state related to the two agents, and $str_{ij}$ denotes the joint strategy for the two agents. The sum of all these edge-specific Q-functions defines the global Q-function. In this case this is approximated using the max-plus message-passing algorithm [2].

- **Agent-Based Collaborative Reinforcement Learners (Ag-Colab-RL):** This is a variant of the agent-based update sparse cooperative edge-based Q-learning method proposed in [1]. As in Ed-Colab-RL method, given two peer agents performing their tasks, $A_i$ and $A_j$, the Q-function is denoted succinctly $Q_{ij}(s_{ij}, str_{ij})$, where $s_{ij}$ denotes the joint state related to the two agents, and $str_{ij}$ denotes the joint strategy for the two agents.

Further details on these methods are reported in [24].

**IV. Training and testing**

**A. Trajectory Prediction**

This section summarizes how the aforementioned BDA and ML algorithms are applied to the data-driven trajectory prediction process based exclusively on raw surveillance data.

As described above, the first phase of the proposed approach is based on clustering. For our task, we adopt the SemT-OPTICS approach proposed in [23]. The dissimilarity between two enriched points is decomposed by two parts, one regarding their spatio-temporal dissimilarity and another regarding their dissimilarity on the semantic components.

**Definition 1 [distance between enriched points $D_{ij}$.**] Given two enriched points $r_i$ and $r_j$, their distance $D_{ij}(r_i, r_j)$ is defined by using the following monotone, ranking function with
respect to Euclidean distance proximity of their points dist\(_e\), and the relevancy of their enriched vectors dist\(_v\):

\[
D_{\text{L} \& \text{R}}(r_i, r_j) = A \cdot \text{dist}_e(r_i, r_j) + (1 - A) \cdot \text{dist}_v(r_i, r_j)
\]

(5)

\[
\text{dist}_e(r_i, r_j) = \sqrt{w_x \cdot (x_i - x_j)^2 + w_y \cdot (y_i - y_j)^2 + w_z \cdot (z_i - z_j)^2 + \frac{w_t}{w_x} \cdot (t_i - t_j)^2}
\]

(6)

\[
\text{dist}_v(r_i, r_j) = 1 - \frac{v_i \cdot v_j}{\|v_i\| + \|v_j\| - v_i \cdot v_j}
\]

(7)

where the distance proximity of the spatio-temporal components dist\(_e\) is the Euclidean distance in the 4-D vector \((x, y, z, t)\). Weights \(w_x\) and \(w_z\) can be defined by the user to weight the spatial versus the temporal dimension. Ratio \(w_t/w_x\) determines the spatial difference that “is equivalent” with one unit time difference [e.g., one second]. This ratio can be estimated by the mean speed of all moving objects. As regarding maxEuclideanDistance(DB) function, it is the coverage in the 4-D spatio-temporal space that acts as a normalization factor. The “semantic” distance dist\(_v\) is measured by Jaccard distance, while \(\lambda \in [0, 1]\) is used to tune the relative importance between the two components.

Based on the Definition above, the distance \(D_n\) between two enriched trajectories is defined as follows:

**Definition 2 (distance between enriched trajectories, \(D_n\)):** The distance \(D_n\) between two enriched trajectories \(R_i\) and \(R_j\) of arbitrary length [i.e., arbitrary number of enriched points], is given by:

\[
D_n(R_i, R_j) = \min \left\{ D_p\left(T(R_i), T(R_j)\right) + D_e(r_i, r_{j,1}), \right. \\
\left. D_p\left(T(R_j), T(R_i)\right) + D_e(r_j, r_{i,1}), \right. \\
\left. D_p\left(T(R_i), T(R_j)\right) + D_e\left(r_{\text{gap}}, r_{i,1}\right), \right. \\
\left. D_p\left(T(R_j), T(R_i)\right) + D_e\left(r_{\text{gap}}, r_{j,1}\right) \right\}
\]

(8)

where \(T(R_i)\) denotes the tail of \(R_i\), namely the enriched points of \(R_i\) after removing the 1-st enriched point of the i-th semantic trajectory \(r_{i,1}\), and gap is a virtual enriched point whose spatio-temporal value is the origin of the 4-D space of the entire dataset, while its “semantic” component corresponds to the zero vector.

**FIGURE 2 — Example of four main clusters (colored) and one cluster of noise & outliers (black) produced in the clustering phase upon the RT (actual routes) using the EDR semantic-aware similarity metric.**

![Example of four main clusters](image)

Subsequently, in the second phase of the proposed approach, the medoid produced for each cluster is used as the base for designing a Hidden Markov model (HMM).

As described earlier, the states and the corresponding state transition matrix for each cluster are defined by the reference points included in the associated flight plans, while the emissions (not to be confused with fuel consumption related emissions) and the corresponding emissions matrix are defined by a probabilistic model of the pair-wise deviations between flight plans and the cluster’s medoid itself.
Typically, emissions are associated with some property or output from the system that is modeled by the HMM, in the sense that the system shifts between states internally and the emissions are the corresponding observations produced with every such transition, since the states themselves are not observed in a HMM. It is common to assume that the HMM emissions follow a Gaussian distribution in each state, if the number of observations allow such a statistical approximation (more than 30 unbiased samples). Thus, in this approach it is sufficient to have clusters of at least 30 member trajectories.

Using the formulation above, this two-phase hybrid clustering/HMM approach was tested in a benchmark dataset of actual flight trajectories (around 1400 flights). One airport pair was considered from the Spain airspace (Barcelona/Madrid) and each direction was modeled separately, as it involves different takeoff/landing approaches. Each direction and pair of airports will be associated with a separate clustering/HMM model, in order to capture the fine details of each case. For other different city-pairs, the process can be straightforwardly applied, although the identified clusters, the related medoids and the associated HMM will be different.

Figure 3 illustrates the per-waypoint means and confidence intervals for Latitude in cluster 1 as described above. The height of each bounding box is directly linked to the uncertainty associated with producing the maximum-likelihood deviation from the HMM emissions in each reference waypoint, i.e., the difference between the flight plan and the aircraft actual route. As expected, most of the waypoints just after takeoff and before landing have the tightest confidence intervals, while sharp turns are the most difficult to predict. Figure 4 illustrates the distributions of the confidence intervals (ranges) of Lat/Lon/Alt and inclusion radius R, providing an overview of the statistical uncertainty per dimension and in 3-D for cluster 1. The height of each box, i.e., the size two central quartiles, is directly linked to the statistical uncertainty in predicting each dimension of the pair-wise deviations between flight plans and the cluster medoid.

**FIGURE 3** — Mean and confidence interval of the Latitude deviations (in meters) within cluster 1 over the minimum common length of flight plans included.

**FIGURE 4** — Distributions of confidence intervals (ranges) of Lat/Lon/Alt and radius of inclusion sphere (in meters) within cluster 1 over the minimum common length of flight plans included.
In this sense, flights in cluster 1 (255/703 members) were predicted with accuracy of roughly 183...234 meters upon each reference waypoint of filed flight plans. In contrast, flights in the much smaller cluster 4 (75/703 members) were predicted with accuracy of roughly 595...736 meters. In practice, these implies that for each reference waypoint of the flights in the cluster, there is 1-\(\alpha\) probability (here 90%) that the pair-wise deviation in Lat/Lon/Alt between the flight plan and the cluster’s medoid will reside within the corresponding confidence interval of the mean (emission output) and the true 3-D distance of this deviation will be at most \(R\) (in meters). In other words, these numbers define how compact is the cluster.

These results demonstrate the robustness and the statistical significance of the proposed hybrid clustering/HMM approach. As described earlier, this method exploits the constraints imposed by the flight plans, i.e., the intended flight path, as well as other “enrichment” parameters such as localized weather and aircraft properties. It should be noted that the proposed method is inherently generic. It does not rely on spatio-temporal grid sizes or resolution, number of semantic parameters or discretization of them. It does rely on pre-flight constraints, more importantly the flight plan that is associated with each actual route.

B. Demand Capacity Balancing

There has been performed a series of experiments in order to test and compare the efficiency of the three collaborative Q-learning methods. The efficiency is measured by means of the resulting number of hotspots, the mean delay achieved and the distribution of interacting flights in Occupancy Counting Periods, in conjunction to the number of learning periods needed for methods to compute policies. Simulation scenarios of trajectories crossing airspace have been used based on actual traffic situations (nominal). The airspace comprises a grid of sectors (and capacities). Parameters used in producing the experimental cases are the following: size of the grid of sectors, sector capacity \(C\), number of flights \(N\), occupancy count period, total time, and maximum delay.

To evaluate the three approaches in cases of varying difficulty we modify the capacity of sectors, and the number \(m\) of sectors that each flight crosses. Results included here are the most challenging cases in the grid considered, where \(m \in \{3, 4\}\). For every capacity value \(C \in \{4, 10\}\), 10 experiments were run. This approach will be extended in a further stage to usual sectors being defined around traffic crossing areas.

Figure 5 shows the mean value and the standard deviation of the final (after learning) number of hotspots, as well as the mean delay for all flights. According to the results, all methods showed a similar behavior in terms of the number of hotspots (Fig. 5.a). A significant improvement in the ‘mean delay of all flights’ criterion is shown in Fig. 5.b concerning the edge-based and the agent-based collaborative RL approaches.

Figure 6 illustrates an example of the received learning curves by each method, i.e. the number of hotspots and mean delay as estimated in the first 1000 episodes during learning. All methods were able to converge rapidly, achieving strategies with zero hotspots to any sector, and with flights’ delay much less than the maximum acceptable delay.

Finally, Figure 7 shows an example of the distribution of interacting flights in terms of Occupancy Counting Periods. This was obtained by measuring the interacting flights to a specific sector in different periods: (a) at the beginning and (b) at the end of learning. As can be seen, the proposed collaborative RL schemes manage to offer strategies with significantly reduced interactions among flight trajectories.
FIGURE 5 — Comparative results: (a) the number of hotspots and (b) the mean delay estimated by each method in terms of various values of sectors’ capacity.

FIGURE 6 — Learning curves received by three methods in a setting considering sectors’ capacity equal to 7.

FIGURE 7 — Example of the distribution of interacting flights.
The final experiment was created using operational data from Spanish airspace, corresponding to one day in January 2016. The main difference here, regarding the parameters, is that the delays applied are no longer a multiple of the occupancy period, but plain minutes. They are the same parameters as above considerably higher values (for instance, number of flights equals to 3195). In this case results are presented for just one method (Independent Learners), but they are representative of those provided by the different methods.

This change brings the experiment closer to a real world situation, but poses an advanced difficulty for two reasons. Firstly, the maximum delay is much bigger than in the previous experiment, which means that every agent has many more states to explore. Secondly, a flight can be delayed for less than one occupancy period, as opposed to the previous experiments.

Figure 8 shows the learning curve received by the Independent Learners (Ind-Colab-RL) method, which converges to a solution with average delay close to 0. The exploration-exploitation policy used was the $\varepsilon$Greedy strategy. The exploration stops at episode 130, where the exploitation begins. Figure 9 shows the initial and final distribution of flights in the sector with two out of seven total hotspots.

**FIGURE 8 — Learning curve received by the Independent Learners**

**FIGURE 9 — An example of the distribution of interacting flights in Occupancy Counting Periods (a) initially and (b). Finally the sector’s capacity is 20**

V. Conclusion

The results achieved by DART project so far in terms of application of machine learning algorithms to both trajectory prediction and demand-capacity balancing problems are already very positive and promising, with still room for refinement in subsequent research stages of the project.
Different approaches have been presented, and tested with actual operational data. Future work will focus in improving the problem modeling to include further operational features that help to explore the benefits that such techniques can bring to the ATM domain. The results presented in this paper have already been shared within an Expert group involving including Network Managers, ANSPs and Airspace Users with positive feedback.

REFERENCES

[17] Le Fablec, Yann, and Jean-Marc Alliot. “Using Neural Networks to Predict Aircraft Trajectories.” IC-AI. 1999
EFFECTS OF REDUCING WIND-INDUCED TRAJECTORY UNCERTAINTY ON SECTOR DEMAND

Alfonso Valenzuela, Antonio Franco, and Damián Rivas
Department of Aerospace Engineering
Universidad de Sevilla, 41092 Seville, Spain
Email: avalenzuela@us.es

Javier García-Heras and Manuel Soler
Department of Bioengineering and Aerospace Engineering
Universidad Carlos III de Madrid, 28911 Leganés, Spain

Abstract—In this paper, a first step to analyse the effects of reducing the uncertainty of aircraft trajectories on sector demand is presented. The source of uncertainty is wind, forecasted by Ensemble Prediction Systems, which are composed of different possible atmosphere realizations. A trajectory predictor determines the routes to be followed by the different flights to reduce the uncertainty of the arrival times. The sector demand is described in terms of entry count, that is, the number of flights entering the sector during a selected time period, which is uncertain because so are the the entry times to the sector. Results are presented for a realistic application, where the dispersion of the entry count is shown to be reduced when the dispersion of the arrival times is also reduced.

I. Introduction

In 2005, the European Commission stated the political vision and high level goals for the Single European Sky and its technological pillar SESAR. Accomplishing the goals of increasing capacity and improving safety requires a paradigm shift in operations through state-of-the-art, innovative technology and research. A promising approach that can improve current prediction and optimization mechanisms towards meeting these goals is to model, analyse, and manage the uncertainty present in Air Traffic Management (ATM).

Weather uncertainty is one of the main sources of uncertainty that affect the ATM system [1]. The limited knowledge about present and, especially, future meteorology conditions, such as wind velocity and direction, fog, snowfall or storms, is responsible for much of the delays and flight cancellations, which negatively affects ATM efficiency and translates to extra costs for airlines and air navigation service providers.

The work presented in this paper shows that the effects of wind uncertainty on the prediction of the demand of an Air Traffic Control [ATC] sector can be reduced when the airspace users plan the route of each individual flight with the objective of increased predictability. The general framework of this work is the project TBO-Met1, funded by the SESAR Joint Undertaking. The overall objective of this project is threefold: 1) to advance in the understanding of the effects of meteorological uncertainty in Trajectory-Based Operations [TBO], 2) to develop methodologies to quantify and reduce the effects of meteorological uncertainty in TBO, and 3) to pave the road for a future integration of the management of meteorological uncertainty into the ATM system.

1 https://tbomet-h2020.com/
Ensemble Weather Forecasting is a prediction technique that allows to estimate the uncertainty in a weather forecast. In this work, the meteorological uncertainty is provided by Ensemble Prediction Systems (EPS). Typically, an EPS is a collection of 10 to 50 forecasts, referred to as members, with forecasting horizons of up to 2-5 days. They consist on running many times a deterministic model from very slightly different initial conditions [2]. Often, the model physics is also slightly perturbed, and some ensembles use more than one model within the ensemble or the same model but with different combinations of physical parameterization schemes. This technique generates a representative sample of the possible realizations of the potential weather outcome. The uncertainty information is on the spread of the solutions in the ensemble, and the hope is that the spread of the predictions in the ensemble brackets the true weather outcome [3].

Uncertain trajectories are obtained during the process of trajectory planning when meteorological uncertainty is taken into account. For each flight, the trajectory predictor computes an ensemble of aircraft trajectories, each one corresponding to a different member of the EPS [4]. Because they are computed for different weather realizations, different flight durations and fuel consumptions are obtained [5]. For each flight, the trajectory predictor developed in TBO-Met [6] determines an optimised route, which minimises a combination of the average and the dispersion of the flight time. The relative importance of each term can be adjusted through a parameter to make the trajectory more or less predictable. The trajectories of all flights, along with the information of the ATC sector, are then used to analyse the sector demand.

In this work, the sector demand is described in terms of the entry count, which is defined as the number of flights entering the ATC sector during a selected time period [7]. This count is obtained from the intersections of the individual aircraft trajectories with the boundary of the sector. Since the aircraft trajectories are uncertain, then the associated entry times are also uncertain and, thus, the entry count is also uncertain. The analysis is then based on the statistical characterization of these times and this count [8].

Results are presented for a realistic application. The demand of an ATC sector is analysed for a whole day when predicted the day before, as would do, for example, the Network Manager when balancing capacity and demand. In this application, it is shown the effect of increasing the predictability of the individual flights on the entry times and the entry count.

II. Methodology for Sector Demand Analysis

The general scheme for the analysis of sector demand is shown in Figure 1, see Ref. [8]. Initially, a scenario is defined in terms of: 1) ATC sector (e.g., geometry and capacity), 2) flights that cross the sector (e.g., origin and destination, departure times, flight levels, and cruise speeds), and 3) weather forecasts (e.g., EPS to be considered, release time, and forecast times).

FIGURE 1 — General scheme for the analysis of sector demand.
The meteorological data provided by weather forecasts need to be processed for its use by the trajectory predictor. For example, the necessary values of wind and air temperature are extracted, and information about convection can be derived from different parameters.

The trajectory predictor computes, for each flight and for each weather prediction, a different aircraft trajectory. The trajectory predictor used in this application and developed in TBO-Met is described in Section III. The computed trajectories, along with the information of the ATC sector, are then used to perform the analysis of the sector demand. The different atmospheric realizations lead to different predicted entry times and, therefore, to different entry counts. In this work, this trajectory predictor is adjusted to obtain more predictable trajectories.

To perform the analysis of the sector demand, the entry times of the flights to the sector and the entry count are statistically characterized. Mean, maximum, and minimum values, and the spread of the times and of the count, measured as the difference between the maximum and minimum values, are examined.

### A. Definitions and general hypotheses

In this work, it is considered that the geometry of the ATC sector is fixed and does not change with time. Therefore, the effects of opening/closing sectors are not analysed.

It is considered that there exist \( m \) different flights and that the EPS is formed by \( n \) different members or atmospheric realizations. The position of flight \( i \) \( (i = 1, \ldots, m) \) for member \( j \) \( (j = 1, \ldots, n) \) at time \( t \), is denoted as \( x_{ij}(t) \). It is given by the longitude \( \lambda \), the latitude \( \phi \), and the pressure altitude \( h \):

\[
x_{ij}(t) = [\lambda_{ij}(t), \phi_{ij}(t), h_{ij}(t)]
\]

The trajectories \( x_{j} \) generated by the trajectory predictor are provided as a list of discrete points and times. A linear interpolation is used to obtain the position of the aircraft at any time.

In this work, it is considered that the trajectory crosses the ATC sector only once; trajectories that cross the same sector multiple times are not considered (for example, flights that return to the departure airport), because this is an uncommon practice in commercial aviation.

### B. Entry time and entry distance

If the trajectory \( x_{j} \) crosses the ATC sector, then there exist an entry time to the sector \( t_{ij,E} \) and the associated entry point \( x_{ij}(t_{ij,E}) \). In this work, since the considered trajectory predictor determines a unique route for all the weather realizations, the entry point is the same for all the members of the EPS. The uncertainty information is on the spread of the entry times.

The entry times are statistically characterized. For flight \( i \), we define the average entry time \( t_{i,E} \) as

\[
t_{i,E} = \frac{1}{n} \sum_{j=1}^{n} t_{ij,E}
\]

and the dispersion of the entry time, \( \Delta t_{i,E} \), as the difference between the maximum and the minimum values for the different atmospheric realizations

\[
\Delta t_{i,E} = \max_{j} t_{ij,E} - \min_{j} t_{ij,E}
\]
For flight $i$, the entry distance is the distance travelled by the aircraft from its origin to the entry point, denoted as $d_{i,E}$. Since the trajectories will be provided as a list of discrete points, the distance between two consecutive points is calculated considering a rhumb line.

### C. Entry count

The entry count for a given sector is defined as the number of flights entering the sector during a selected time period $P_k$. In this paper, time periods with durations $\delta t = 10$, 30, and 60 minutes are considered.

Because the entry times are uncertain, the aircraft may enter the sector in different time periods, thus leading to an uncertain entry count. The larger the dispersions of the entry times and the smaller the values of $\delta t$, the more likely the entry count to be uncertain. For example, in case that the dispersion of the entry time of one flight is larger than the duration of the time period, then this flight may enter the sector in two or more consecutive time periods.

An entry function for flight $i$, for ensemble member $j$, and for time period $P_k$ is defined, denoted as $E_{ij}(P_k)$: it takes the value 1 when the aircraft enters the ATC sector during this time period and the value 0 otherwise

$$E_{ij}(P_k) = \begin{cases} 1 & \text{if } t_{ij,E} \in P_k \\ 0 & \text{if } t_{ij,E} \notin P_k \end{cases}$$

The entry count for ensemble member $j$ and for time period $P_k$, denoted as $E_j(P_k)$, is obtained as the sum of the entries of the different flights

$$E_j(P_k) = \sum_{i=1}^{n} E_{ij}(P_k)$$

From these $n$ values of the entry count, mean, maximum, and minimum values ($\bar{E}$, $E_{\text{max}}$, and $E_{\text{min}}$, respectively) for time period $P_k$ are determined

$$\bar{E}(P_k) = \frac{1}{n} \sum_{j=1}^{n} E_j(P_k)$$

$$E_{\text{max}}(P_k) = \max_j E_j(P_k)$$

$$E_{\text{min}}(P_k) = \min_j E_j(P_k)$$

The uncertainty information is on the spread of the entry count. The dispersion of the entry count, $\Delta E(P_k)$, is defined as follows

$$\Delta E(P_k) = E_{\text{max}}(P_k) - E_{\text{min}}(P_k)$$

Notice that, since the entry of each flight to the ATC sector for time period $P_k$ only depends on the entry time $t_{ij,E}$, see Eq. (4), then the entry count and its statistical characterization are only affected by the uncertainty in this time. Therefore, the uncertainty in the entry count increases when the uncertainty in the entry time increases.
III. Trajectory predictor

The trajectory predictor considered in this work solves the problem of trajectory planning considering uncertain wind fields provided by EPS, see Ref. [6]. For each flight and a given EPS formed by \( n \) members, the trajectory predictor determines \( n \) different trajectories, each one corresponding to a different ensemble member. All trajectories follow the same route, described as a sequence of waypoints. The difference between the trajectories is then the arrival times to the waypoints of the route, because they are subject to different wind fields. Since each member of the forecast is considered as equally probable, then each trajectory is also considered as equally probable.

The route determined by the trajectory predictor for each flight minimises a weighted sum of the average flight time of the \( n \) trajectories and of the flight-time dispersion, measured as the difference between the maximum flight time \( t_{\text{f,max}} \) and the minimum flight time \( t_{\text{f,min}} \). The relative weight of the dispersion is controlled by a parameter denoted as \( p \). By changing the value of this parameter, one can obtain routes that are more efficient on average (they arrive earlier) or routes that are more predictable (they show less dispersion).

In this work, the en-route phase is considered, flown at constant altitude and constant airspeed. The inclusion of variable altitude and speed profiles, and the exploration of other objective functions that include the flight dispersion is left for future work, see for instance Ref. [9].

This problem of trajectory planning is formulated as a deterministic optimal control problem, where the \( n \) trajectories are simultaneously considered. For each flight, the mathematical formulation of the problem is as follows

\[
\min \frac{1}{n} \sum_{j=1}^{n} f(r) + p (t_{f,max} - t_{f,min})
\]

subject to the constraints

\[
\frac{d}{dr} \begin{bmatrix} r \cr t_1 \cr \vdots \cr t_n \end{bmatrix} = \begin{bmatrix} \cos(x_g) \\
\frac{R_e + h}{V_{g,n}^{x_g}} \\
\sin(x_g) \\
\frac{R_e + h}{V_{g,n}^{x_g}} \cos(\phi) \\
\frac{1}{V_{g,n}^{x_g}} \\
\vdots \\
\frac{1}{V_{g,n}^{x_g}} \end{bmatrix}
\]

\[
\begin{bmatrix} V_{g,n}^{x_g} \cos(x_g) \\
V_{g,n}^{x_g} \cos(x_g) \\
\vdots \\
V_{g,n}^{x_g} \sin(x_g) \\
\vdots \\
V_{g,n}^{x_g} \sin(x_g) \end{bmatrix} = \begin{bmatrix} V \cos(x_1) + w_{y_1}(\theta, \lambda) \\
V \cos(x_1) + w_{y_1}(\theta, \lambda) \\
\vdots \\
V \sin(x_1) + w_{y_1}(\theta, \lambda) \\
V \sin(x_1) + w_{y_1}(\theta, \lambda) \end{bmatrix}
\]

where \( r \) is the distance flown along the route, \( r_f \) is the distance flown when arriving to the destination, \( f(r) \) is the flight time at distance \( r \) for ensemble member \( j \), \( x_g \) is the course,
\( R_e = 6371 \text{ km} \) is the Earth radius considered as a sphere, \( V_j \) are the groundspeeds derived from the winds provided by the ensemble member \( j \), \( V \) is the aerodynamic airspeed, \( \chi_j \) is the heading for ensemble member \( j \), and \( w_{x,j} \) and \( w_{y,j} \) are the zonal and meridional components of the wind for member \( j \). Notice that the wind does not depend on the time in this formulation, it is obtained from the forecast closer to the middle time of the flight. An estimation of the arrival time needs to be provided to choose the appropriate forecast beforehand.

The boundary conditions of the problem are

\[
\begin{align*}
(\phi(0), \lambda(0)) &= (\phi_0, \lambda_0) \\
(\phi(t_j), \lambda(t_j)) &= (\phi_i, \lambda_i) \\
t_j(0) &= t_0, \quad \forall j \in \{1, \ldots, n\}
\end{align*}
\]

where \( \phi_0 \) and \( \lambda_0 \) are the coordinates of the departure point, \( \phi_i \) and \( \lambda_i \) are the coordinates of the destination point, and \( t_0 \) is the departure time.

The resolution of this mathematical problem relies on an initialization and wind approximation procedure described in Ref. [10]. It is solved with direct methods, discretizing the trajectory with a trapezoidal scheme and then solving the resulting nonlinear optimization problem with NLP software [see, for example, Ref. [11]].

**IV. Application**

In this application, the demand of the ATC sector LECMSAU is analysed for a whole day, 01 September 2016 (from 00:00 to 24:00), when predicted the day before, 31 August at 00:00. Next, the traffic scenario is described, in terms of ATC sector, flights, and weather forecasts. The results are presented and analysed in Section V for two different values of the parameter \( p \).

**A. ATC sector**

The sector LECMSAU is located in the Northwest of Spain, see Figure 2. It is an en-route sector, ranging from flight level 345 to 460. Its declared capacity (i.e., the maximum number of flights entering the sector per hour) is 36 flights/hour. This information has been obtained from Eurocontrol’s Network Strategic Tool (NEST) for the Aeronautical Information Regulation and Control (AIRAC) cycle 1609.

**FIGURE 2 — Geographical location of ATC sector LECMSAU and the extended area.**
B. Flights

The information of the flights is also obtained from NEST, and it corresponds to the last filed flight plans from the airlines (i.e., initial trajectories, according to NEST nomenclature). Notice that the optimal routes determined by the trajectory predictor of this work are, in general, different from the routes filed in the flight plans, and it may happen that an optimal route crosses the sector whereas the corresponding planned route does not. Therefore, to take into account this situation, it has been decided to consider flights that planned to cross an extended area around LECMSAU. The coordinates of the four vertices of this area are (see Figure 2): [N 47°, W 15° 30'], [N 46° 30', W 2° 30'], [N 40°, W 5° 30'], [N 40°, W 15°].

A total number of 1443 flights is obtained from NEST for the date of analysis and the extended area. However, 51 of these flights are discarded, those flights arriving or departing to/from LEST, LEVX, and LECO airports. The reason is that, under the hypothesis of flying at constant pressure altitude, these flights instantly appear or disappear inside the sector, not crossing the sector boundaries. Thus, a total number of 1392 flights is considered in this application. This traffic is composed of short flights (departing from Portugal, Spain, and France), medium flights (flights from the Canary Islands, the British Isles, the Scandinavian Peninsula, and Eastern Europe), and long flights (from South, Central, and North America).

The trajectory predictor described in Section III requires the following information for each flight:

- coordinates of the departure and arrival airports: obtained from NEST;
- departure time: obtained from NEST;
- arrival time: obtained from NEST, used as a reference time;
- pressure altitude: fixed to 200 hPa (approximately 38600 ft) for all aircraft and the whole flight;
- airspeed: the average cruise Mach provided by Eurocontrol’s Base of Aircraft Data (BADA) 3.13 [12] for the aircraft model that performs the flight is considered for the whole flight (from origin to destination), ranging from 0.63 to 0.85.

C. Weather forecast

The meteorological uncertainty is retrieved from the European Centre for Medium-Range Weather Forecasts. In particular, the weather forecast ECMWF-EPS, composed of 50 perturbed members, is used.

Since the analysis is performed the day before the operation, the forecasts released at 00:00 on 31 August 2016 are considered. According to the flight plans retrieved from NEST, the earliest flight departs at 16:20 on 31 August and, as a reference, the latest flight arrives to its destination at 09:54 on 02 September. Taking into account these times, the forecasts with forecasting horizons of 12, 18, 24, 30, 36, 42, 48, 54, and 60 hours are considered.

In agreement with the coordinates of the route waypoints, the forecasts are retrieved for a coverage area which ranges from 45 degrees South to 75 degrees North, and from 130 degrees West to 50 degrees East. The spatial grid resolution is 0.25 degrees. According to the cruise altitude chosen for all flights, the forecasts are retrieved for the pressure level 200 hPa.

The meteorological variables required by the trajectory predictor described in Section 3.2 are the zonal wind and the meridional wind (winds along the West-East and South-North directions, respectively).
In Figures 3 and 4, the average and the dispersion of the meridional and the zonal winds are shown for the forecast corresponding to the time instant 12:00 on 01 September. The dispersion is measured as the difference between the maximum and the minimum values at each geographic location.

The average meridional wind, see Figure 3 top, ranges approximately between -30 m/s (South direction) and 40 m/s (North direction). High values of the wind are found at the East coast of North America and the North Atlantic Ocean. The average zonal wind, see Figure 3 bottom, is larger than the meridional wind, ranging approximately between -30 m/s (West direction) and 70 m/s (East direction). The zonal wind is therefore the main contributor to the existence of jet streams. The larger values are found again at the East coast of North America, the North Atlantic Ocean, and South America.

FIGURE 3 — Average meridional (top) and zonal (bottom) winds, ECMWF-EPS released at 00:00, 31/08/16, forecasting horizon 36 hours.

The dispersion of the winds, see Figure 4, is rather large, with maximum values above 40 m/s. The geographic areas affected by high uncertainty are approximately the same in both cases. In particular, it can be highlighted the East coast of North America and the North Atlantic Ocean, affecting flights from North America to Europe.
V. Results

Next, results are presented for two different values of the relative weight of the dispersion, $p = 0$ and $p = 20$. Notice that, for a given flight and for each value of the parameter $p$, a different route is obtained which may or may not cross the sector. Trajectories that exit the sector and briefly enter again have been discarded for being not realistic.

The number of trajectories that enter the sector for $p = 0$ is 440 and for $p = 20$ is 624. In this work, only the effect of the reduction of the time dispersion on the sector demand is analysed. For this reason, 328 flights are considered, those flights that cross the sector LECMSAU for both values of $p$. The effect of the varying number of aircraft entering the sector, that is, the displacement of the traffic flows from one sector to another, is left for future work.

A. Entry times

The dispersion of the entry time, $\Delta t_{i,E}$, as a function of the distance to the entry point, $d_{i,E}$, for each flight is presented in Figure 5. Firstly, one can see that, as one could expect, in general the dispersion increases as the distance increases because the uncertainty is accumulated along the trajectories; flights arriving from distant locations present more uncertainty. For example, for $p = 0$, the maximum dispersion is as large as 449 seconds and it is found for 7553 km.

Secondly, it can be seen that there are flights with similar distances but different values of dispersion. For example, for $p = 0$ and for $d_{i,E}$ approximately equal to 1500 km, the dispersion ranges between 115 and 339 seconds. As possible causes of these different values, the following ones can be highlighted:
different routes (flights over regions of the airspace with different uncertainty),

different effects of the same wind uncertainty on different flights (uncertainties in tail/headwinds have a higher impact than uncertainties in crosswinds),

different departure times (the predictions for flights departing later are made with weather forecasts with larger time horizons, thus having a larger uncertainty), or

different speeds (flights with lower values of Mach number and with headwinds are more sensitive to uncertainties in the wind).

Finally, it can be seen that the trajectories obtained for \( p = 20 \) show a lower dispersion in the entry time. The average value of dispersion for all the aircraft (that is, the average value of the points in Figure 5) is 156.4 s for \( p = 0 \), and 125.8 s for \( p = 20 \), a reduction of 30.6 s. However, as can be inferred from the objective function (10), this reduction of the dispersion comes from an increase of the average flight time. In this application, the average flight time increases 382.5 s.

**FIGURE 5 — Dispersion of the entry time vs distance to the entry point.**

![Dispersion of the entry time vs distance to the entry point.](image)

---

**B. Entry count**

The entry count for \( p = 0 \) and three different time intervals, \( \delta t = 60, 30, \) and 10 minutes, is shown in Figure 6. The average entry count is shown as vertical bars, and the minimum and maximum entry count as whiskers. The capacity of the sector is depicted as a red horizontal line, which is 36 flights/hour, and is assumed to be 18 flights/30 minutes, and 6 flights/10 minutes.

For the 60 minute interval, the largest value of the mean entry count is 40.5 flights, found at 07:00-08:00, which exceeds the sector capacity. The entry counts determined for smaller time intervals allow a more precise identification of the traffic peaks. In particular, for the 30 minute interval, the traffic peak takes the value 26.6 and is found in the period 07:00-07:30, and for the 10 minute interval it is 14.2 flights in 07:20-07:30, much higher than the assumed capacities.

The uncertainty on the entry count is on the spread of the number of flights, that is, the height of the whiskers in Figure 6, also represented in Figure 7 for convenience. For example, for \( \delta t = 30 \) minutes, the difference between the maximum and the minimum values of the entry count is as large as 3 flights for a total of 0.5 hours, 2 flights for 5.5 hours (in disjoint periods), 1 flight for 9 hours, and 0 flights for the remaining 9 hours.

When the duration of the time periods is shortened, it can be observed that the maximum values of the dispersion increase, as already noted in Section II-C, but the average values of the entry counts are proportionally reduced; thus, the uncertainty becomes relatively
more important. For example, for the 30 minute interval, the largest dispersion on the entry count is 3 flights and the average entry count is 6.83 flights/period (obtained as the number of flights entering the sector divided by the number of time periods), 44% in relative terms; whereas for the 10 minute interval the largest dispersion on the entry count is 4 flights and the average entry count is 2.28 flights/period, 175%.

For \( p = 20 \), the average entry counts are slightly different to those found for \( p = 0 \), shown in Figure 6, due to differences in the average entry times; they are not shown for brevity. The main difference between the two set of results is found in the dispersion of the entry count, as can be seen in Figure 8 compared to Figure 7. The maximum dispersion can be occasionally larger, for example, for 10 minutes and \( p = 20 \) the maximum dispersion is 5 flights, and for \( p = 0 \) is 4 flights, but on average, the dispersion is significantly reduced: for 60 minutes the average dispersion per period (and for the whole day) is reduced from 0.83 to 0.50 flights, for 30 minutes from 0.99 to 0.69, and for 10 minutes from 0.90 to 0.50.

**FIGURE 6 — Entry count for \( p = 0 \) and \( \delta t = 60 \text{ minutes} \) (top), 30 minutes (middle), and 10 minutes (bottom).**
FIGURE 7 — Dispersion of the entry count for $p = 0$ and $\Delta t = 60$ minutes (top), 30 minutes (middle), and 10 minutes (bottom).
VI. Conclusions

A first step to show the effects of reducing the uncertainty of aircraft trajectories on the sector demand has been presented in this paper. The source of uncertainty is wind, forecasted by Ensemble Prediction Systems. The uncertainty is reduced by a trajectory predictor to be employed by airspace users, which determines the appropriate route to minimise a weighted sum of the average flight time and of the flight-time dispersion. The sector demand is described by the entry count, which is uncertain because the entry times to the sector are uncertain.
In the application presented, it has been found that, because uncertainty is accumulated along the flight, the uncertainty in the entry time increases as the distance travelled by the aircraft to the entry point increases. Thus, sectors with predominance of incoming long-haul flights are expected to be more affected by weather uncertainty. Also, it has been found that the uncertainty in the entry count may be rather large, in particular when small time periods are considered. It is worth noting that, in this application, only wind uncertainties are considered. Larger values of uncertainty are expected in scenarios that consider uncertainties on air temperature and, primarily, convective phenomena.

When the dispersion of the individual trajectories is reduced, the dispersions of the entry times and of the entry counts are also reduced. However, this reduction of the dispersion comes from an increase of the average flight time and, thus, of fuel consumption and operating costs. Both the airlines and the Network Manager will benefit of better predictability. The airlines will know better when the aircraft will arrive to the destination airports, leading to a better fleet scheduling. The Network Manager will know more precisely the demand of the sector, which may allow to improve the Demand-Capacity Balancing process, better identifying the Air Traffic Flow and Capacity Management measures to be applied.

It has been found that when the predictability of individual flights is increased, the trajectories are deviated, modifying the number of aircraft crossing the sector. The displacement of traffic flows from one sector to another is left for future work.

This analysis can be extended to consider other demand indicator as it is the occupancy count, that is, the number of flights inside the sector during a selected time period. For this count, it is necessary to consider also the exit time from the sector, which is also affected by the weather uncertainties that exist inside the ATC sector.

In the immediate future, under the scope of the TBO-Met project, this methodology will be applied to quantify the effects of weather uncertainty on convective phenomena. In this case, the uncertainty will be obtained from probabilistic nowcasts.

Acknowledgements

This work is part of the project TBO-Met. This project has received funding from the SESAR Joint Undertaking under grant agreement No 699294 under European Union’s Horizon 2020 research and innovation programme.

REFERENCES


SHORT-TERM 4D TRAJECTORY PREDICTION USING MACHINE LEARNING METHODS

Zhengyi WANG  
Sino-european Institute of Aviation Engineering  
Civil Aviation University of China  
Tianjin, China

Man LIANG  
Ecole Nationale de l’Aviation Civile  
Toulouse, France

Daniel DELAHAYE  
Ecole Nationale de l’Aviation Civile  
Toulouse, France

Abstract—4D trajectory prediction is the core element of future air transportation system, which is intended to improve the operational ability and the predictability of air traffic. In this paper, we introduce a novel model to address the short-term trajectory prediction problem in Terminal Manoeuvring Area (TMA) by application of machine learning methods. It consists of two parts: clustering-based preprocessing part and Multi-cells Neural Network (MCNN)-based machine learning part. First, in the preprocessing part, Principle Component Analysis (PCA) is applied to the real 4D trajectory dataset for reducing the vector variable dimensions. Then, the trajectories are clustered into partitions and noises by Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method. After that, the Neural Network (NN) model is chosen as machine learning method to find out the good predicting model for each individual cluster cell. Finally, with the real traffic data in Beijing TMA, the predicted Estimated Time of Arrival (ETA) for each flight is generated. Experiment results demonstrate that our proposed method is effective and robust in the short-term 4D trajectory prediction. In addition, it can make an accurate trajectory prediction in terms of MAE and RMSE with regards to comparative models.

I. Introduction

4D trajectory prediction refers to the calculation and prediction of longitude, latitude, altitude and time on the future waypoint sequence based on the existing data. During the development of Trajectory Based Operation (TBO) concepts in Single European Sky ATM Research (SESAR) and Next Generation Air Transportation System (NextGen) programs, trajectory prediction is intended to improve the predictability of air traffic, it is the core element of future air transportation system.

The 4D trajectory prediction can be influenced by several factors, such as aircraft weight, pilot actions, wind and temperature. These uncertainties will not only make it difficult to improve the prediction accuracy, but also will decrease the prediction process efficiency as the prediction time becomes longer[1]. According to the time scale, 4D trajectory prediction can be divided into two categories [2]:

- Short-term 4D trajectory prediction
- Long-term 4D trajectory prediction
1. Tactical (short-term) trajectory prediction: A prediction in a short period within several minutes or even shorter. Since the prediction scale is relatively small, minor change may have great impact on prediction results. Therefore, tactical trajectory prediction require as much information as possible. Flight-related information contained in radar or ADS-B data is usually taken;

2. Strategical (long-term) trajectory prediction: A kind of prediction before departure based on the flight plan, which provides the prediction from a macroscopic view. It is mainly applied to fuel consumption and airspace flow evaluation.

In this paper, we propose a novel short-term trajectory prediction model, which combines the different machine learning techniques to address the problem of 4D trajectory prediction in Terminal Maneuvering Area (TMA). This model can be divided into two main parts: preprocessing part and machine learning part. The preprocessing part contains several steps: data cleaning, filtering, re-sampling, Principle Component Analysis (PCA), density-based clustering and training. In the machine learning part, Multi-Cells Neural Networks (MCNN) technique will be applied to generate the predicted trajectory for different patterns.

II. Literature review

4D trajectory prediction can be mainly classified into aircraft performance models and machine learning models, according to input parameters models [3].

Aircraft performance models belong to physics-based approaches. The model structure is based on kinetic assumptions. The model parameters are determined based on a model of the aircraft performance, the planned flight routes, the predicted atmosphere condition, and the expected command and control strategies given by pilots or FMS (known as Aircraft Intent). The most precise aircraft performance model is Base of Aircraft Data (BADA) Family 4, which provides increased levels of precision in aircraft performance parameters for modelling and simulation [4]. A variety of researches based on BADA and Aircraft Intent have been conducted. In 2008, Lin Xi et al. presented a classified ADS-B-based trajectory prediction algorithm [5]. Based on the state estimation by Kalman filter and intent information captured by a pretreatment and probability method, the aircraft trajectory can be predicted with computation efficiency and less errors. M. Porretta et al. presented a novel aircraft performance model in consideration of effects of wind, for aircraft lateral guidance and a new procedure for speed estimation [6]. The model input includes navigation data and aircraft intent information, based on EUROCONTROL BADA set. Simulation results show that the model is suitable for reliable trajectory prediction. In 2014, J. Kaneshige et al. described the implementation and evaluation of a motion-based trajectory prediction function, which can increase the resiliency and robustness of TBO [7]. Based on the performance index such as the fuel consumption, flight time, the algorithm computes the difference between with trajectory prediction and without trajectory prediction. Although, aircraft performance models have made great contributions to trajectory prediction, most of these models made ideal assumptions, rarely considered the real constraints, human behaviour factors, and the intersection of trajectories.

As a branch of Artificial Intelligence (AI), machine learning has been developed over 30 years, aims to learn from experiences and make predictions. The trend of recent years show that machine learning is widely used in trajectory prediction domain. Compared with those aircraft performance models, machine learning models were constructed with weak assumptions or even without assumptions. In some case, it shows better prediction performance. For example, in 1999, Yann Le Fablec et al. used Neural Networks to predict an aircraft trajectory in the vertical plane. The model is trained by a set of real historical trajectory, where two different method were adopted: in the first method, the input is current altitude, the remaining altitude to reach Request Flight Level (RFL) and $n$ past vertical speeds, the output is the next speed; while in the second method, it is built with
the starting altitude and the remaining altitude to reach, the RFL as input, the first initial speeds as output. Simulation result showed that the Neural Networks give better results than classical prediction functions based on model of aircraft [8]. In 2013, De Leege et al. introduced Generalized Linear Models [GLMs] for trajectory prediction at a prediction horizon of 15NM to 45NM on fixed arrival route. The inputs of the model are aircraft type, ground speed at the Initial Approach Fix (IAF), altitude over the IAF, surface wind and altitude winds. All inputs come from surveillance data and meteorological data [9].

In the view of improving the accuracy in prediction tasks, S. Trivedi et al. carried out a study on the feasibility of utilizing clustering as a preprocessing approach [11]. Their research shows that the improvement on prediction accuracy is significant on large-scale cluster-able datasets by combining the clustering with even some simple machine learning predictors. Under routine traffic situation, in the TMA, the aircraft follows the standard arrival/departure procedure and regular ATC instructions, which makes trajectories cluster-able. Thus, application of machine learning together with clustering for 4D trajectory prediction in TMA is a valuable and interesting research topic. Several efforts on combining clustering with simple machine learning predictors have been investigated. For example, in 2014, K. Tastambekov et al. considered the short to mid-term aircraft prediction problem, namely, the prediction with a horizon of 10-30 min [1]. The model firstly searches similar trajectories in terms of shape and time, then uses wavelet decomposition to solve the linear regression model in the relationship between time and trajectory projection onto one of the three axis X, Y and Z. This method produces efficient results with high robustness. In 2015, S. Hong et al. introduced a new framework for predicting aircraft arrival times by combining the ATC intent information [12]. The training stage of the method contains two steps: trajectory pattern identification and regression models construction for each pattern. The prediction of arrival times can be achieved by applying different regression models for each trajectory pattern of target aircraft.

However, most of the aforementioned existing models still fall short. Some models neglect the prediction steps, directly consider clustering results as prediction results. A majority of trajectory pattern identification approaches are not robust, require high-quality flight data that follow the same departure/arrival procedure. If there are some noise and overflights, the results will be far less effective. In addition, the machine learning approaches that have been used are relatively simple and shallow in structure.

In this paper, we will extend the trajectory clustering method, which is introduced by Gariel et al. in reference [13], to study the short-term trajectory prediction model with machine learning methods. The main contributions of this paper are threefold:

1. A novel hybrid 4D trajectory prediction model based on clustering and MCNN is developed.

2. The proposed model is robust. The preprocessing part of the model can effectively and efficiently process the data, provide the high-quality inputs to the prediction part.

3. It can improve the accuracy of prediction. A comparative study is conducted to demonstrate the effectiveness of our model, compared with Multiple Linear Regression (MLR) model.

III. Methodology

A. Overview

The flow chart of the proposed trajectory prediction approach is demonstrated in Fig. 1. Our novel trajectory prediction approach includes two parts: clustering-based preprocessing part and MCNN-based machine learning part.

The DBSCAN method together with PCA form the preprocessing step. In this part, our model aims to identify the 4D trajectories into different clusters and remove noises in
an efficient way. Each cluster symbolizes that the corresponding trajectories have the similar pattern. Noises contain trajectories with holding patterns, trajectories with large vectoring, the trajectory in special cases and overflight trajectories. After identifying the trajectory pattern and removing noises, the trajectory data quality will be highly increased.

**FIGURE 1 — Proposed 4D trajectory prediction approach**

In the part of machine learning, we apply the MCNN method to process different traffic data. First, for each partition of trajectories, there is a predictor, in which there is an individual NN-based learning cell. Each individual learning cell will be trained with the associated cluster of trajectories. Consequently, each classified partition of trajectories will have its corresponding predicting model. Second, for the new input data, we will classify them into different corresponding clusters, then with our proposed multi-cells predicting model, trajectory prediction of the input data is generated.

**B. Data preparation**

The available dataset includes ADS-B records in July, 2017 over the TMA of Beijing Capital International Airport (BCIA), which is one of the busiest airport in the world, with three parallel runways: 18R/36L, 18L/36R and 01/19.

Since the studied airspace is relatively small, the longitude, latitude and altitude of trajectory points can be transformed into 3D Cartesian coordinates. Each sample of data contains:
1. Type of operation (departure/arrival),
2. Record beginning time $t$,
3. Aircraft number,
4. Position $(X, Y, Z)$,
5. Heading $\Psi$,
6. Horizontal velocity $V_h$,
7. Vertical velocity $V_v$, etc.

Each record with the same aircraft number belongs to an aircraft $i$, and the collection of all records for that aircraft forms the trajectory $T_i = \{t_i|t\in[1, n]\}$, where $n$ is the total number of trajectory in the dataset. Note that, in this paper, only flights that correspond to runways 18R/36L and 18L/36R are taken into consideration. These part of data consist of 36288 flights and 3242384 trajectory points.

Fig. 2 depicts the four traffic patterns in the 18R/36L and 18L/36R configuration, roughly clustered according to route nodes passed. Here, QFU means the magnetic orientation of runway-in-use. QFU 36 is to North, and QFU 18 is to South.

**FIGURE 2 — Runways 18R/36L and 18L/36R traffic patterns in Beijing capital international airport**

C. Clustering-based preprocessing

The preprocessing part can be divided into the following steps:

1. Data cleaning and formatting,
2. Dimensionality augmentation,
3. Principal component analysis,
4. Clustering via DBSCAN.

Data Cleaning and Formatting: Due to the instability of ADS-B data receiver, our collected ADS-B data is not complete. Some trajectories have missing parts. It is necessary to filter them out. To solve this problem, a low pass filter is applied by the following function:

$$\tilde{x}_i^j = x_i^j$$ (1)
\[ \hat{x}_i^l = \alpha x_i^l + (1-\alpha) \hat{x}_i^{l-1}, \quad l \in [2, m_i - 1] \]  

(2)

Where the 3D coordinates and heading of the \( l \)-th point of \( i \)-th trajectory are substituted into \( x_i^l \). \( \alpha \) is a smoothing factor in \([0, 1]\). In this study, \( \alpha \) is set to 0.5 to provide better results without too much delay. \( m_i \) is the number of points in \( i \)-th trajectory.

Trajectories with less than 50 points were eliminated due to statistical insufficiency. In order to make dataset suitable for clustering, each trajectory should be represented as a vector. All the trajectory vectors are re-sampled into the same length, then their distance can be computed. The re-sample method for \( i \)-th trajectory is given as follow:

\[ T_i = \left\{ T_i^j | j = \text{round} \left( \frac{k \cdot m_i}{50} \right), k \in [1, 50] \right\} \]

(3)

**Dimensionality augmentation:** This step aims to augment the dimensionality of dataset. The existing dimensions may not be sufficient and will result in lack of information, which can’t completely reflect the differences between each trajectory. The augmentation of dimensions will help improve the clustering performance. Therefore, the following dimensions will be added into the dataset:

1. Distance from the reference point \( R \), which indicates the convergence degree of trajectory. Due to the runway configuration, we define the reference point \((X_{\text{ref}}, Y_{\text{ref}}, Z_{\text{ref}})\) as \((73.5, 65.5, 0)\). For each trajectory point, \( R_i^l \) is given as:

\[ R_i^l = \sqrt{(X_i^l - X_{\text{ref}})^2 + (Y_i^l - Y_{\text{ref}})^2 + (Z_i^l - Z_{\text{ref}})^2} \]

(4)

2. Distance from the corner point \( D \). According to the dataset, the corner point \((X_{\text{cor}}, Y_{\text{cor}}, Z_{\text{cor}})\) is assigned as \((-50, 200, 0)\). The corner point will help solve the identifying problem when two trajectories are symmetric. The \( D_i^l \) is calculated by the function below:

\[ D_i^l = \sqrt{(X_i^l - X_{\text{cor}})^2 + (Y_i^l - Y_{\text{cor}})^2 + (Z_i^l - Z_{\text{cor}})^2} \]

(5)

The reference point and corner point play the role as multilateration.

3. Angular position from the reference point \( \Theta \). It shows the variation (turning status) of trajectory with respect to the reference point. \( \Theta \) is defined as:

\[ \Theta_i^l = \arctan \left( \frac{Y_i^l - Y_{\text{cor}}}{X_i^l - X_{\text{cor}}} \right) \]

(6)

To sum up, the re-sampled dataset includes original features: position \((X, Y, Z)\), heading \( \Psi \) and additional features: distance from the reference point \( R \), distance from the corner point \( D \), angular position from the reference point \( \Theta \). To avoid the discontinuity at \( \pm \pi \), the sine and cosine values of \( \Theta \) and \( \Psi \) is adopted.

Next, to make every feature on the same scale, each feature is normalized in \([0, 1]\). The general formula is given as:

\[ x^* = \frac{x - \min(x)}{\max(x) - \min(x)} \]

(7)

where \( x \) is the original feature and \( x^* \) is the normalized feature. Replacing \( x \) with our features, finally, the trajectory is organized as follows:
\[ T_i = \begin{bmatrix} p_i^r & r_i^r & d_i^r \cos(\Theta) \sin(\Theta) & \sin(\Psi) \cos(\Psi) \end{bmatrix} \] (8)

\[ T = \begin{bmatrix} T_1 \\ \vdots \\ T_n \end{bmatrix} \] (9)

where \( T_i = [x_i^r, y_i^r, z_i^r] \). Then, each trajectory is re-sampled with 450 dimensions. Matrix \( T \) is \( n \times 450 \).

**Principal Component Analysis**: As shown in Eq. (8), trajectories are related to various factors. Nevertheless, among these factors, some is more related, while the other is less related. Redundant elements will decrease computational efficiency, even lead to larger errors. To solve this problem, Principal Component Analysis (PCA) is introduced. PCA is a powerful tool used to reduce the dimension of dataset without losing too much information. The main idea of PCA is to derive an orthogonal linear transformation to project each of the vector variables into principal components for the maximum amount of variance that can be presented in lower dimensions [14].

PCA performs a linear transform on the \( n \times m \) (in this case \( m = 450 \)) matrix \( T \):

\[ Y = E \cdot T \] (10)

Where \( E \) is a rotation matrix, \( Y \) is the new principal component matrix. The variance of \( Y \) is:

\[ \text{var}(Y) = E^T \cdot C \cdot \] (11)

Where \( C \) is the covariance matrix of \( T \), which can be written as:

\[ C = \frac{1}{n-1} T \cdot T^T \] (12)

The eigenvalues of \( C \) can be calculated as \( \{\lambda_i\}_{i=1}^{m} \), which correspond to the variances in \( Y \) as \( \{\lambda_i\}_{i=1}^{m} \), with \( \lambda_1 > \lambda_2 > ... > \lambda_m \).

To map a dataset \( X \subset \mathbb{R}^m \) to a dataset \( Y \subset \mathbb{R}^q \) with \( q \in \{1, m\} \), a rotation matrix \( E = [v_1, ..., v_q] \) can be used. The dimension can be reduced by choose the number of \( q \). It is required that the projection should better covers 95% of the variances, i.e., the cumulative percentage or variance explained \( G(q) \) is greater than 95%:

\[ \sum_{i=1}^{q} \lambda_i \geq 95\% \] (13)

**Clustering via DBSCAN**: As an unsupervised learning approach, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a commonly used density-based clustering algorithm [15]. The core concept of DBSCAN is to evaluate the density according to the number of points within the \( \epsilon \)-neighbourhood. DBSCAN classifies the points into three types: core point, density-reachable point and noise point. The algorithm expands to density-reachable areas from a selected core point, then obtaining a maximum area including the core point and density-reachable points. Being robust to the quality of datasets, DBSCAN can divide the dataset into several clusters and noises, where the a-priori selection of the number of clusters is not required. Besides, DBSCAN is able to find arbitrarily shaped clusters. The advantages of DBSCAN make it fits well with trajectory clustering scenarios.
There are two principle parameters in DBSCAN algorithm: the neighbourhood radius $\epsilon$ and the minimum number of points required to form a cluster $MinPts$. These two parameters should be well chosen: The value of $\epsilon$ will affect the size of clusters. The value of $MinPts$ will affect the noise identification and the significance of clusters. After the proposed processing approach, the dataset for machine learning model will have better quality and the performance will be increased.

D. MCNN-based learning model

The machine learning used in our short-term trajectory prediction is supervised learning method. Supervised learning finds a mapping function from the input to the output based on the training data. The prediction can be achieved by applying the mapping function to the new inputs. As one of the most classical machine learning algorithms, regression model is commonly used in 4D trajectory prediction problem [16], [12], [1], [9]. A regression model can be expressed as:

$$y = f(x, \beta),$$

(14)

where $y$ is dependent variable, $x$ is independent variable, $\beta$ represents parameters. More specifically, the Multiple Linear Regression (MLR) model is the most common form of regression analysis, frequently applied to prediction [12]. Given $n$ multiple independent variables $\{x_i | i \in [1,n]\}$ and corresponding dependent variable $y$, the model can be formalized as following:

$$y = \sum_{i=1}^{n} \beta x_i + \beta_0,$$

(15)

where $\{\beta_i | i \in [0,n]\}$ are parameters, which can be approximated by least squares approach.

In this paper, we use MCNN model to predict the Estimated Time of Arrival (ETA) based on preprocessed real 4D trajectory data. The advantage of the usage of Neural Network (NN) in each prediction cell is that they are able to learn the hidden and non-linear dependencies from the training data. The architecture of proposed NN model for each cell is composed of an input layer, a hidden layer and an output layer, shown in Fig. 3. Given input $\{x_i | i \in [1,n]\}$ and the hidden layer node number $m$, the network output can be calculated as:

$$y = \sum_{j=1}^{m} w_{ij} \left( \sum_{j=1}^{n} w_{ij} x_j + b_j \right) + c$$

(16)

Where $w_{ij}$ is the weight between the $j$-th input node and the $i$-th hidden node, $w_{ij}$ is the weight between the $i$-th hidden node and the output node, $b_j$ is the bias to the $i$-th hidden layer, $c$ is the bias to the output layer. $f$ is the activation function, in which Sigmoid function is commonly used. To find suitable weights such that the NN is in good performance, the cost function should be minimized. To increase the efficiency of updating the gradients, a prevailing cost function: cross-entropy cost function $J$ is used:

$$J = -\frac{1}{N} \sum_{t=1}^{N} \left[ t ln y + (1-t) ln (1-y) \right]$$

(17)

where $N$ is the number of training data, $t$ is the target output. The steep descent is used to update and obtain the optimized parameters, which can be computed by well-known back propagation algorithm.

The new input can be classified according to the initial point of each trajectory. In view of arrival flights in TMA, initial points of trajectories in each cluster belong to a certain
range in 3D Cartesian coordinate system. This character of dataset can be used to realize an effective classification on each new input trajectory.

**FIGURE 3 — Neural network architecture used in this paper**

**E. Nested cross validation**

In order to well select the parameters of prediction model, and to achieve an unbiased performance of the prediction model, this paper utilizes nested cross validation method. It consists of the outer loop and the inner loop. In the outer loop, there is a $k_1$-fold cross validation that splits the data into $k_1 - 1$ folds of training sets and one fold of test set. Then in the inner loop, there is another $k_2$-fold cross validation, which will further split the training set into $k_2 - 1$ fold of training sets and one fold of validation set. Taking $k_1 = 5$, $k_2 = 5$, the concept of the whole process is demonstrated by Fig. 4. The proportion of training sets, validation sets and test sets is 64%/16%/20%. The purpose is that the inner loop is for parameters selection, such as learning rate, number of hidden nodes, and the outer loop is to validate the robustness of our prediction model.
Here, we use Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to evaluate our trajectory prediction model performance:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|$$  \hspace{1cm} (18)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}.$$  \hspace{1cm} (19)

where $\hat{y}_i$ is the $i$-th predicted value and $y_i$ is the $i$-th observed value of ETA. A smaller value of MAE or RMSE represents a better accuracy of prediction.

Given that each outer iteration produces a $\text{MAE}_i$ and a $\text{RMSE}_i$, $i \in [1, k_1]$, the average MAE and RMSE can be computed as follows:

$$\text{MAE} = \frac{1}{k_1} \sum_{i=1}^{k_1} \text{MAE}_i,$$  \hspace{1cm} (20)

$$\text{RMSE} = \sqrt{\frac{1}{k_1} \sum_{i=1}^{k_1} \left(\text{RMSE}_i\right)^2}.$$  \hspace{1cm} (21)
IV. Simulation and result

A. Dataset

The dataset that we used in the experiments contains 8677 arrival flights of QFU 36 extracted from the available dataset described in section III-B.

B. Results and discussion

In this study, the cumulative percentage of variance is calculated and presented in Fig. 5. We can see that when the principal component reaches over 32, the variance explained will be more than 95%. Let \( q = 32 \), then the dimension of each trajectory was reduced to 32 from 450. To sum up, dimensionality augmentation enriches the features that principle components can choose. PCA reduce the dimension of the dataset, which makes the following clustering step more efficient and accurate in the projected principal component space.

**FIGURE 5 — The cumulative percentage of variance in PCA**

For DBSCAN step, experience shows that setting the parameters as \( \varepsilon = 1.8 \) and \( \text{MinPts} = 200 \) is an optimum choice for this dataset. The distance metric used is Euclidean distance. taking a randomly generated training fold & validation fold for demonstration proposes. The resulting clusters is presented in Fig. 6a (trajectories in 2D) and Fig. 6b (trajectories in 3D), the noises is presented in Fig. 7.

According to Fig. 6 and Fig. 7, the trajectories are divided into 5 clusters. Clustered trajectories account for 93.47% of total trajectories. Noises represent 6.53%. Fig. 7 shows that the noise is mainly composed of holding patterns and trajectories with large vectoring, which will have an interference for prediction stage. Therefore, the noise should be removed from the dataset. In addition, there is no significant reduction on numbers of trajectory in the dataset.

The clustered partitions for each iteration is illustrated in Fig. 8, in which each trajectory is presented with its first 3 principle components. As we can see, 5 similar partitions were clustered for each iteration. The minimum proportion of clustered trajectories represent 93.22% and the corresponding noises account for 6.78% of all trajectories. The percentage is reasonable, which will not only eliminate the bad effect by noise, but also will keep most of the information.
FIGURE 6 — Cluster result of QFU 36 arrival trajectories example

FIGURE 7 — Noises result of QFU 36 arrival trajectories example
To compare the performance of MCNN learning with the simple machine learning model, the Multiple Linear Regression (MLR) was proposed with the same clustering preprocessing step, and 5-fold cross validation is applied. The average proportion of test sets in each clusters and the ETA prediction errors of the proposed NN model and MLR were summarized in Tab. I. According to the Tab. I, with the same preprocessing procedure, the proposed NN model performs significantly better than MLR model in view of MAE and RMSE, not only in total, but also for each cluster.

To illustrate the importance of the proposed clustering preprocessing step mentioned in section III-C, the prediction errors of NN model and MLR model both without preprocessing are presented in Tab. II. We can see from the Tab. II and Tab. I that in view of the same machine learning method, the model with clustering preprocessing step has less prediction errors than the one without clustering preprocessing step, which proves that the clustering preprocessing is effective in improving the prediction accuracy. Besides, the NN model prevails against the MLR model.

We further observe the distribution of ETA prediction errors with different prediction methods. In Fig. 9, X axis is the value of prediction error, Y axis is the frequency, which presents the percentage of trajectories on the associated error. With four different prediction methods, large part of trajectory predictions are all with less than 100 seconds error. Moreover, NN
method performs better than MLR method. MLR with preprocessing method can improve the accuracy of prediction. The method NN with preprocessing performs the best ETA prediction. In addition, Fig. 10 reveals the mean absolute error of ETA prediction with the fly time to destination (runway). With four different prediction methods, the results show the same trend, that is: when the time to destination is fewer, the absolute prediction error is smaller. The NN with preprocessing performs best. In conclusion, the proposed model in this paper is efficient and able to make an accurate 4D trajectory prediction.

**TABLE I — The performance on ETA prediction of NN and MLR with preprocessing step**

<table>
<thead>
<tr>
<th>Partition number</th>
<th>percentage</th>
<th>MAE for NN+P. (s)</th>
<th>RMSE for NN+P. (s)</th>
<th>MAE for MLR+P. (s)</th>
<th>RMSE for MLR+P. (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>13.85%</td>
<td>106.08</td>
<td>141.51</td>
<td>113.67</td>
<td>150.20</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>5.62%</td>
<td>82.91</td>
<td>108.08</td>
<td>92.99</td>
<td>118.59</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>58.39%</td>
<td>61.68</td>
<td>97.81</td>
<td>82.48</td>
<td>117.14</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>13.64%</td>
<td>46.00</td>
<td>69.37</td>
<td>51.09</td>
<td>75.12</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>8.51%</td>
<td>88.76</td>
<td>124.31</td>
<td>97.42</td>
<td>132.62</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>69.19</td>
<td>104.82</td>
<td>84.37</td>
<td>119.13</td>
</tr>
</tbody>
</table>

**TABLE II — The performance on ETA prediction of NN and MLR without preprocessing step**

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE (s)</th>
<th>RMSE (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR without P.</td>
<td>108.03</td>
<td>160.40</td>
</tr>
<tr>
<td>NN without P.</td>
<td>76.28</td>
<td>127.76</td>
</tr>
</tbody>
</table>

**FIGURE 9 — The distribution of ETA prediction errors with different methods**

- (a) NN without Preprocessing
- (b) MLR without Preprocessing
- (c) NN with Preprocessing
- (d) MLR with Preprocessing
V. Conclusion

In this paper, a novel trajectory prediction approach that combines clustering with machine learning is proposed, implemented and simulated for ETA prediction.

The proposed model contains clustering-based preprocessing step and MCNN-based machine learning prediction step. First, it clusters different traffic flows, then it trains the associated prediction model for different clusters. After that, it is performed on real traffic data in Beijing TMA with nested cross validation. The numerical experiments demonstrate that the proposed method, NN with preprocessing, performs best in terms of MAE and RMSE, compared with other methods, such as NN without preprocessing, MLR without preprocessing, MLR with preprocessing. It can make an accurate 4D trajectory prediction. In addition, the proposed method has a good robustness.

Future work could be conducted in different look-ahead times, on a comparison with results from model-based methods, as well as on studying prediction accuracy for other trajectory variables besides ETA. Moreover, more complex prediction model, such as deep learning approaches, would be very valuable.

VI. Acknowledgement

The authors would like to thank Serge Roux for his assistance with data collection, thank colleagues in the Optim group of ÉNAC and anonymous reviewers for their generous suggestions.

REFERENCES


Getting in touch with the EU

In person
All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email
Europe Direct is a service that answers your questions about the European Union. You can contact this service:
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696 or
– by email via: https://europa.eu/european-union/contact_en

Finding information about the EU

Online
Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications
You can download or order free and priced EU publications at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents
For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU
The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.