

Application of Machine Learning for ATM Performance Assessment – Identification of Sources of En-Route Flight Inefficiency

Rodrigo Marcos, Ricardo Herranz, Rita Rodríguez Vázquez, Pedro García-Albertos, Oliva G. Cantú Ros
Nommon Solutions and Technologies
Madrid, Spain
rodrigo.marcos@nommon.es

Abstract— The Single European Sky Performance Scheme aims to drive performance improvements in European aviation by setting binding targets for performance indicators. This process is not trivial due to the high number of stakeholders and the complex interdependencies between indicators and influence factors. This paper proposes a novel approach, based on machine learning techniques, to identify and evaluate the sources of flight inefficiency. A Random Forest regressor is trained to predict flight efficiency as a function of different flight properties derived from flight plans and ideal routes, such as heading, altitude and airspace crossed. The predictor enables the identification and evaluation of the relative importance of the factors that determine flight efficiency. We conclude by discussing the limitations and room for improvement of the proposed approach, as well as the future developments required to produce reliable performance assessments by means of machine learning techniques.

Keywords- *Machine learning; performance modelling; flight efficiency; Random Forest*

I. INTRODUCTION

Performance orientation is one of the key pillars of the Single European Sky (SES). By setting down EU-wide and local performance targets, as well as ensuring monitoring and corrective actions, the SES Performance Scheme aims to drive performance improvements in European aviation. Air Traffic Management (ATM) performance management implies target setting, measurement of indicators, and intervention to ensure that goals are met. Setting down targets is a complex process due to the high number of stakeholders involved. In addition, a targeted indicator may limit the maximum achievable value for other indicators due to interdependencies. Most indicators are correlated with more than one indicator, which makes trade-off evaluation a challenging task. Similarly, in the SES ATM Research programme (SESAR), there is a need to evaluate the impact of technological solutions on performance and to provide common assumptions to correctly evaluate impact mechanisms, influence factors and interdependencies between Key Performance Areas (KPIs) [1].

In the context of SESAR, several initiatives have tackled the problem of target setting and technology evaluation. Project 19 [2] is responsible for developing the SESAR Performance Framework and a common strategy to validate SESAR solutions, considering influence factors and interdependencies

between KPIs. Some Exploratory Research (ER) projects like ACCESS [3], SATURN [4], COMPAIR [5] or VISTA [6] have developed macroscopic models of the network to evaluate the impact of different policies and scenarios. Another approach explored in ER projects has been the use of microsimulation to evaluate the impact of different Concepts of Operations (ConOps) – on a set of performance indicators (e.g., the APACHE project [7]) – or to propose more representative performance indicators – (e.g., the AURORA project [8]).

Machine learning provides an alternative approach to performance analysis, by allowing the prediction of complex patterns and relationships among big and often heterogeneous data [9]. In the context of ATM performance, machine learning can assess interdependencies between KPIs [10], analyse safety anomalies [11] or analyse delay events [12]. In this paper, we explore how to use machine learning techniques to identify cause-effect relationships between ATM performance drivers and indicators. We do so through a case study focused on environmental flight efficiency.

The main environmental flight efficiency Key Performance Indicators (KPIs) for the first and second SES Reference Periods (RP) (2012-2015, and 2015-2018) are the average Horizontal en-route Flight Efficiency (HFE) of the last filed flight plan trajectory (KEP) and of the actual trajectory (KEA). However, it is often argued that Air Navigation Service Providers (ANSPs) have limited ability to influence these KPIs, which also depend on the coordination with other ANSPs and military authorities, airspace user preferences, and external events such as strikes or long-lasting airspace closure [13]. Traditionally, the identification of these influence factors has often relied on expert judgement [14].

The goal of this paper is to identify sources and drivers of en-route flight efficiency by means of machine learning techniques. A flight efficiency predictor is used to explore the dependencies between the interface component of the HFE in a single Area Control Centre (ACC) and several influence factors, such as the heading of the flight or the airspaces crossed by the flight trajectory. The predictor is trained with historical traffic data and used to identify the main drivers of the observed inefficiencies.

The rest of the paper is organized as follows: Section II describes the selected case study, the data sources used, and the approach and methodology followed for flight efficiency prediction and influence factor assessment; Section III describes the exploratory data analysis carried out to identify potential influence factors; Section IV evaluates the prediction power of the flight efficiency predictor; Section V presents the assessment of flight efficiency influence factors; Section VI concludes by discussing the limitations of the proposed method and outlining future research directions to produce reliable performance assessment by means of machine learning techniques.

II. DATA AND METHODOLOGY

A. Case Study

The identification of flight efficiency influence factors is applied to the Bordeaux ACC (LFBBCTA) during Aeronautical Information Regulation and Control cycle (AIRAC) 1702 (02/02/2017-01/03/2017). This ACC was selected due to its high volume of traffic (more than 50,000 flights during AIRAC 1702) and position (it lays in the path of the South-West air-traffic axis, which is one of the main air traffic flows in Europe).

This study is focused on the influence that nearby ACCs have on the en-route flight efficiency achieved in a certain ACC, which may limit the achievable performance improvement in flight efficiency. Therefore, the factors considered are only those related to airspace and trajectory. Other factors such as weather are out of the scope of the study. Moreover, only en-route segments (above flight level 245) of non-military instrumental flights with origin and destination within the European Civil Aviation Conference (ECAC) area are considered to ensure the consistency of the results.

In this study we consider the HFE indicator for the actual trajectory [15] rather than the actual fuel burnt currently used in the SESAR 2017 PF [1] to measured flight efficiency. The main reasons for this are that HFE has been used as the main flight efficiency indicator for RP1 and RP2, and it offers a consistent methodology to be measured and disaggregated from trajectory data. The application of the proposed method to other flight efficiency indicators, including those suggested by recent SESAR projects [16], is left for future research.

B. Data Sources

1) Demand Data Repository (DDR)

The DDR is a restricted-access flight database maintained by EUROCONTROL, which records data for almost all flights flying within the European ECAC area. This database has been fully operational since 2013. The DDR information used in this case study includes:

- Flight description: ID, airline, aircraft, origin, destination, date, departure time, arrival time, most penalising regulation and associated delay.
- Trajectory description: coordinates, timing, altitude and length of the flight.

- Intersections between airspace and flights: airspace intersected, entry time and exit time.
- Airspace information: airspace definition (coordinates, altitudes) and airport coordinates.

This information is available for both the actual flown trajectory and the last filed flight plan.

2) Correlated Position Records (CPR)

CPR data contain spatio-temporal data of airborne flights. This information is shared with EUROCONTROL by most of the ANSPs in the Network Manager (NM) area of operations. This dataset contains information on flight position records:

- flight identification,
- timestamp,
- position (latitude and longitude), and
- altitude.

The granularity of these records is higher than that of DDR, as shown in Figure 1. Moreover, trajectories stored in DDR are based on CPR.

For this study, we were granted access to a subsample of these data corresponding to the 20th of February 2017.

C. Approach and Methodology

1) Dataset Preparation

For the whole AIRAC cycle 1702, a dataset was created with flight information from DDR data. Additionally, for the 20th of February 2017, two datasets of flight information were created using DDR and CPR data.

The flights crossing the ACC under study during the 20th of February 2017 (DDR and CPR data) or the AIRAC cycle of study (1702, DDR data). Each of the resulting datasets, consisting of around 1,700 flights for the 20th of February 2017 and 51,000 flights for the whole AIRAC cycle, was randomly split into three disjoint subsets:

- Training dataset: A subset containing the majority (70%) of the flights.
- Validation dataset: A subsample (15%) of the flights.
- Testing dataset: A subsample (15%) of the flights.

Extreme values of the HFE indicator are removed from the datasets, which consisted in a small proportion of the flights

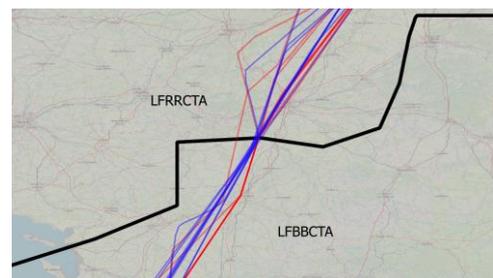


Figure 1. Comparison of actual trajectories from Barajas to Charles de Gaulle in DDR (red) and CPR (blue)

usually corresponding to flights entering the ACC for a few seconds and/or containing data inconsistencies.

2) HFE Indicator

The HFE indicator considers the segments of a flight crossing an airspace and compares the flown and the achieved distance. For every segment p of a flight, let L_{jp} be the horizontal length flown and H_{jp} the achieved distance inside airspace j . The HFE in that airspace is defined as:

$$HFE_j = \sum L_{jp} / \sum H_{jp} - 1.$$

The achieved distance is defined as the mean of the distance increased from the origin and the distance reduced to the destination. The formula to compute the achieved distance for a segment p is:

$$H_p = 1/2 (d(O_p D) - d(E_p D) + d(O E_p) - d(O O_p)),$$

where O_j and E_j are the origin and end point of segment p , respectively, O and D are the origin and destination of the flight, and d is the great circle distance between two points.

HFE can be separated into two components:

- Local extension: it is the difference between the flown horizontal length inside the airspace and the horizontal distance between the entry and exit point to/from the airspace.
- Interface contribution: it is the difference between the distance between the entry and exit point to/from the airspace and the achieved distance.

This separation in components is useful for classifying the sources of flight inefficiency. Whilst the local component can be basically attributed to the sector in question, the interface component may be influenced by other sectors.

In the present paper we focus on the prediction of the interface component of HFE, since the DDR data are not granular enough for predicting the local component. It is important to note that DDR trajectories are not radar tracks but a simplification of the trajectory, where only points that deviate significantly from the planned trajectory are stored.

3) Methodology for the Assessment of Flight Efficiency Influence Factors

The proposed approach entails three main steps:

- Selection of input features for the prediction algorithm.
- Random Forest regressor training and evaluation.
- Identification and assessment of influence factors.

a) Feature Selection

There are a vast number of factors that determine the efficiency of a flight. From a literature review and a working session with other SESAR ER performance projects (AURORA

and APACHE) the following factors were selected to be investigated:

- average heading;
- time of departure;
- great circle distance between origin and destination;
- length flown in the ACC and in contiguous ACCs;
- planned HFE (interface and local) in the ACC;
- reference flight level of the flight in the ACC;
- distance between ideal (if the great circle route is flown) and planned entry/exit point to/from the ACC and the contiguous ACCs;
- flights per active sector in the ACC and in the contiguous ACCs during the flight;¹
- day of the week, calendar day and number of flights crossing the ACC in the day.

These features can be obtained from actual data (e.g., number of flights per active sector), the last filed flight plan (e.g., reference flight level), and the ideal route, i.e., the great circle route (e.g., entry point of the great circle route to the ACC).

These features were visually analysed to determine their suitability for predicting the flight efficiency indicator (section III). These visualisations were supported by the visual work developed in the SESAR ER project INTUIT [17, Ch. 3.2]. Four statistical correlation indicators, namely Pearson's correlation, Spearman's correlation, distance correlation, and mutual information correlation factors, were computed between the HFE indicator in the ACC of study and the different factors.

b) Flight Efficiency Prediction

The flight efficiency predictor assigns each flight a value of the flight efficiency indicator according to a set of flight's characteristics (the features) by means of a Random Forest regressor. In this section, a description of the Random Forest algorithm and training process is provided.

A Random Forest is a machine learning algorithm formed by a set of Decision Trees [18]).

In Random Forests, first each tree is trained with a random subset drawn with repetition (bootstrapping) from the training data. Each split is obtained by considering only a random subset of the features. The output of the Random Forest is obtained by averaging the outputs of the trees.

The main advantage of the Random Forests regressors with respect to single Decision Trees is their lower variance, i.e., similar inputs will result in similar outputs. These may suffer from overfitting when applied to highly correlated data, and so do Neural Networks, whereas Random Forests avoid this by selecting only one or two features in each split. The predictor is implemented using the Python public library scikit-learn [19].

¹ In [22], the causes of flight inefficiency were studied. A relationship was found between route inefficiency and traffic levels.

Prior to the training, extreme values of the indicator were removed from the datasets, which consisted in a small proportion of the flights, usually flights entering the ACC for a few seconds and/or containing data inconsistencies.

The training process consists of assigning internal weights and hierarchy to each feature in the trees of the Random Forest, such that the output of the predictor fits the training data. This process is repeated for different combinations of the parameters of the algorithm (e.g., depth of Decision Trees).

Applying a Machine Learning algorithm on a small dataset might lead to overfitting. In order to avoid this, data augmentation was performed, following the methodology described in [20]. This process consists in creating copies of the training samples with random noise. Data augmentation reduces the prediction error between the validation and testing.

Validation consists in evaluating the performance of the trained predictor with further data. The validation is used to select the best parameters for the Random Forest.

The testing aims to obtain a final measure of the expected error of the prediction. Three scores are obtained: R_2 score, Pearson correlation and normalised root-mean-square error (NRMSE). A review of the achievable prediction power of random forests for different academic datasets can be found in [21]. It concluded that a NRMSE between 25% and 5% as acceptable, depending on the dataset. Note that this implies that random forests do not give the best accuracy but allow working with noisy datasets.

c) Assessment of Flight Efficiency Influence Factors

The assessment of the drivers of en-route flight efficiency and their influence on efficiency is done in a two-fold manner:

- Comparison of the prediction power and feature hierarchy for the random forests trained with DDR and CPR data for the 20th of February. This step serves to validate DDR data.
- Analysis of the most influencing features on flight efficiency for the Random Forest trained with DDR data of AIRAC 1702.

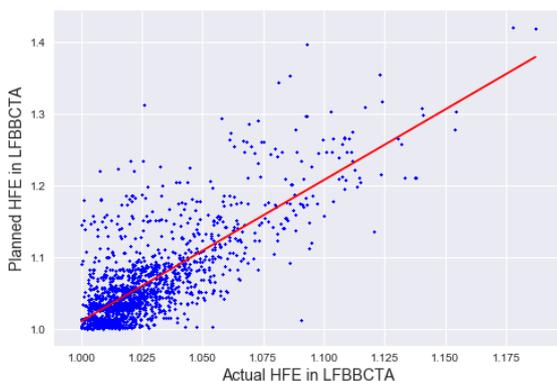


Figure 2. Comparison of planned and actual HFE in LFBBCTA

The analysis of the factors is done by presenting the relative importance of each feature, which is a measure of the influence on the indicator of horizontal flight efficiency.

III. DATA EXPLORATION

In this section, we present the results of the visual exploration of the candidate factors obtained from the literature review and expert consultation, which served to select the features used in the Random Forest regressor.

A. Features

a) Planned trajectory

The flight plan presents a first estimation of the route that the flight will follow. For the planned trajectory, the HFE indicator can be measured. From Figure 2, it can be observed that, despite the high correlation, the planned HFE is not always consistent with actual HFE and has in general higher values.

b) Congestion

A group of features was selected to try to explain the inefficiencies caused by congestion. In this group we include: the number of flights per active sector in the studied ACC and in contiguous ACCs, and the take-off time. In Figure 3, the hourly congestion and the flight efficiency distribution throughout the day of study are shown. It is clear that not only the HFE is higher during the peak hours (6-9 and 18-21) but also the dispersion of it. To consider the congestion at the time of flight, the number of flights per ATCO in an area is pondered over the duration of the flight.

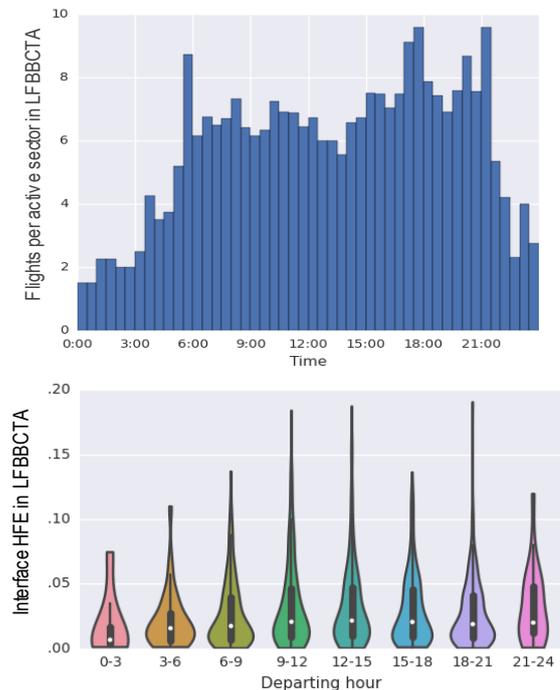


Figure 3. Top: Flights per active sector in LFBBCTA per hour of the day. Bottom: combination of boxplot and kernel density estimate of interface HFE in LFBBCTA vs take-off time. The contour represents the kernel density estimate, the white dot represents the mean, black dots are samples within the second and third quartiles, the vertical lines represent the range of the data en in the subsample.

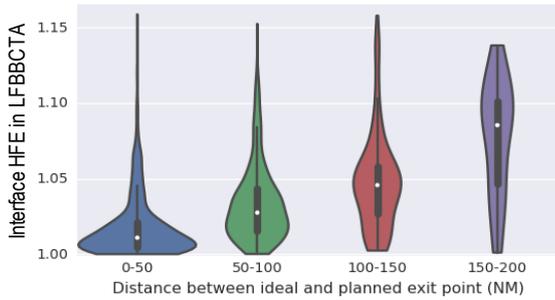


Figure 4. Interface HFE in LFBBCCTA vs distance between ideal and planned exit point.

c) Airspace structure

The airspace structure constrains the flight plans and, thus, the actual HFE. Two features were calculated: the distance between the ideal entry (or exit) point and the planned entry (or exit) point to the ACC. The cases when the ideal route did not cross the ACC are assumed as a maximum distance. The limited number of entry points forces to plan trajectories away from the ideal route, with the subsequent decrease in flight efficiency. As observed in Figure 4, these features have a clear negative influence on efficiency.

d) Flight range

The flight range, measured by the great circle distance (ideal distance) between departure and arrival, is also an influencing feature. This feature can be used to spot differences between long-, medium- and short-range flights. For instance, as observed in Figure 5, shorter flights tend to have slightly lower efficiency (higher HFE).

e) Airspace crossed

Another group of features is the one linked to the airspaces crossed. Two features are calculated for this: the ideal distance flown in the ACC and in contiguous ACCs; and the average heading of the route. The influence of the latter is shown in Figure 6. Note that the heading is computed from 0 to 180 degrees, thus not distinguishing the sense of the direction of the flight. In the figure one can appreciate how some directions have on average higher efficiency (red, green and cyan) than others (blue and yellow).

These features explain additionally how the interface with contiguous airspace is optimised and the influence of the ideal

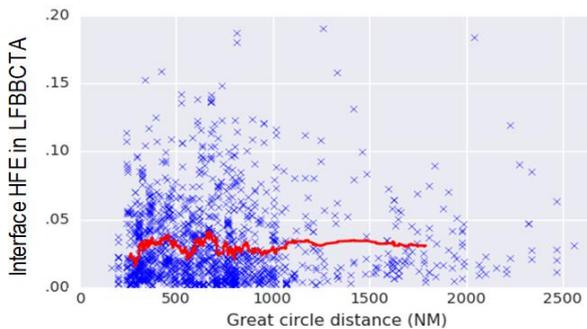


Figure 5. Interface HFE in LFBBCCTA vs great circle distance.

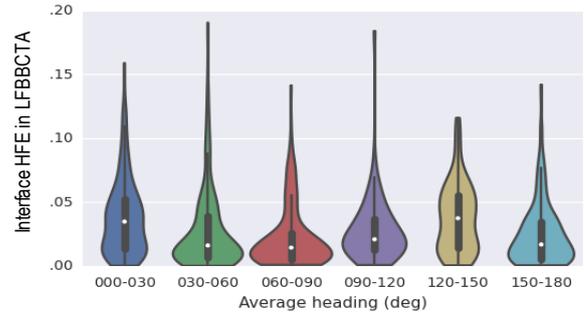


Figure 6. Interface HFE in LFBBCCTA vs average heading.

distance crossed in the ACC on flight efficiency. An example for the LECMCTA is shown in Figure 7. In the figure it is observed that flights whose ideal route does not cross (or only a few kilometres) LECMCTA have in general lower efficiency. Moreover, there are significant differences for flights crossing ideally lower, average and large distances in that ACC.

B. Feature Selection

Based on the previous analysis, the features selected to predict the interface component of HFE are:

- Planned HFE in LFBBCCTA. This indicator serves as a first estimation of the indicator.
- Flown distance in the ACC and in contiguous ACCs of the great circle route. This indicator can highlight low-efficient interfaces and the influence of the distance flown inside the ACC. Some interfaces with contiguous ACCs may be less optimised for certain flows (see Figure 6) and also the distance flown ideally in an ACC has proven to influence efficiency (Figure 7). The ACCs considered in this case study are:
 - LFBBCCTA (Bordeaux ACC).
 - LFRRCTA (Reims ACC).
 - LFFFCTA (Paris ACC).
 - LFMMCTA (Marseille ACC).
 - LECBCTA (Barcelona ACC).
 - LECMCTA (Madrid ACC).
- Global average heading of the route. This feature intends to discover any flight directions that are not

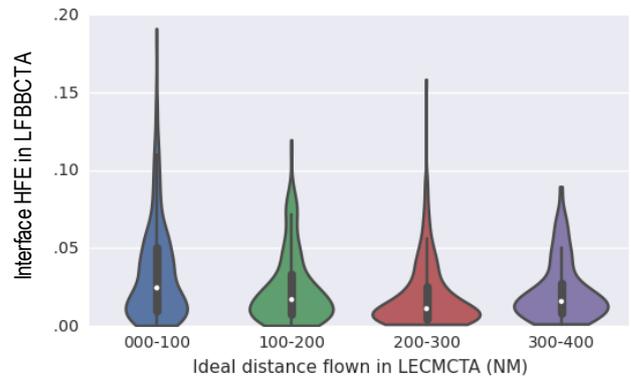


Figure 7. Interface HFE in LFBBCCTA vs ideal distance in LECMCTA.

favoured by the ACC. The feature is separated into two to make it cyclical: sine and cosine of the average heading.

- Distance between the ideal entry (or exit) point and flight plan entry (or exit) point to the ACC. This is an indicator of the closeness between ideal route and flight plan obeying airspace constraints. Routes that are highly deviated in the flight plan with respect to the ideal lead to less efficient interface component.
- Reference Flight Level. Flights that plan a lower flight level may suffer less the effects of congestion or some routes may be restricted at some flight levels.
- Number of aircraft in LFBBCA per active sector in the ACC averaged between the departure and arrival time of the actual flight.
- Distance from origin to destination, which explains the influence of the route length in the efficiency.
- Take-off time. This may highlight differences in efficiency depending on the time of the day. It could be improved by using the planned entry time to the ACC. As the global average heading, it is separated into two to make it cyclical.

IV. COMPARISON OF DATA SOURCES AND ANALYSIS OF INFLUENCE FACTORS

In this section, the results of the prediction power of the Random Forest regressor trained for the day of study are presented and compared for the two data sources provided.

A. Comparison of Prediction Power with different Data Sources

The results of the training, validation and testing score are summarised in Table I. The results for both datasets are similar in accuracy, with slightly lower performance with CPR data. This could be due to the higher detail of unexpected events (such as storm/convective cloud avoidance or direct routings) captured by CPR data. Such events cannot be predicted by the proposed predictor as no inputs are considered for that end.

To validate the results, further accuracy and score metrics are calculated with the testing dataset of DDR data:

- Testing Pearson correlation factor: 0.878.
- Testing mean square error: 7.42E-05.
- Testing NRMSE: 6.91 %.

The achieved accuracy with DDR data is considered to be adequate and thus it can be used to infer influence factors of the interface component of HFE.

TABLE I. TABLE I RANDOM FOREST FOR THE PREDICTION OF THE INTERFACE EFFICIENCY INDICATOR.

Data	Random forest regressor R^2 score	
	DDR	CPR
Training	0.973	0.980
Validation	0.804	0.836
Testing	0.770	0.757

B. Comparison of Influence Factors with Different Data Sources

Table II shows the relative importance of the features used for the Random Forests trained with DDR and CPR data. The relative importance [21, Ch. 10] is obtained as the rate of misclassification when one feature is excluded in the out-of-bag dataset (a subsample of the training data used for error estimation), and then normalised so that the sum is one.

From the table it is observed that the order and measure of importance is coherent for both datasets, thus validating DDR data to infer influence factors for a larger dataset. The most influencing feature is the planned interface HFE, followed by the distance from the planned exit and entry points to the ideal exit and entry points. This suggests that the interface component of the horizontal flight efficiency is highly dependent on the structure of the airspace, the available routes and the interfaces with nearby airspaces. An example is the interface between LFBBCA and LECMCTA in the Pyrenees, which has only a few handover options.

The ideal distance flown in the LFFFCTA is the fourth most influencing feature. This is due to the fact that there is a restriction to cross that airspace, reserved for Paris departures and arrivals, as shown in Figure 8. The next influencing features are the ideal distance in LFBBCA, the ideal distance between origin and destination and the average heading, which explain the type and direction of the flight and how it crosses the ACC. The planned local component of HFE shows a similar importance.

The ideal distances in the rest of contiguous ACCs have similar importance. These features are clearly linked to the direction of the flight, which shows a strong influence on HFE.

The last features in order of importance are the reference flight level, the number of flights per sector, and the take-off time.

TABLE II. TABLE II COMPARISON OF FEATURE IMPORTANCE WITH DIFFERENT DATA SOURCES

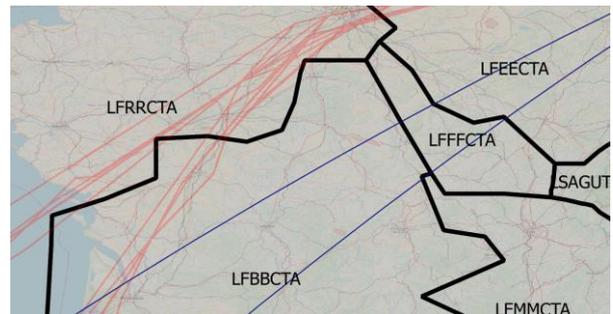


Figure 8. Comparison of ideal (blue) and flown (red) route of flights from Lisboa and Madrid to Frankfurt, crossing ideally LFFFCTA.

Feature	Relative importance	
	DDR	CPR
Data Source	DDR	CPR
Planned interface HFE in LFBBCA	42.08%	43.55%
Distance between planned and ideal exit point from LFBBCA	17.57%	17.06%
Distance between planned and ideal entry point to LFBBCA	6.02%	7.56%
Ideal distance in LFFFCTA	4.42%	4.14%
Ideal distance in LFBBCA	3.94%	3.52%
Ideal distance	3.79%	4.20%
Planned local HFE in LFBBCA	3.17%	3.18%
Average heading	2.99%	2.62%
Ideal distance in LECMCTA	2.52%	2.18%
Ideal distance in LFRRCTA	2.42%	1.50%
Ideal distance in LFMMCTA	1.65%	1.63%
Ideal distance in LECBCTA	1.37%	1.48%
Reference FL in LFBBCA	1.37%	1.52%
Flights per sector in LECBCTA	1.12%	0.80%
Flights per sector in LECMCTA	1.00%	0.75%
Flights per sector in LFMMCTA	0.85%	0.82%
Take-off time - cosine	0.83%	0.84%
Take-off time - sine	0.81%	0.81%
Flights per sector in LFFFCTA	0.74%	0.67%
Flights per sector in LFBBCA	0.73%	0.59%
Flights per sector in LFRRCTA	0.62%	0.55%

V. ASSESSMENT OF FLIGHT EFFICIENCY INFLUENCE FACTORS

In this section, we provide a further analysis of the influence factors of the interface HFE. A Random Forest regressor is trained with a dataset corresponding to the flights during AIRAC 1702 (02/02/2017-01/03/2017). The predictor achieved the following score and error metrics:

- Testing Pearson correlation factor: 0.905.
- Testing mean square error: 8.72E-05.
- Testing NRMSE: 4.30 %.

The accuracy obtained is even higher than for the day of study, due to the larger number of samples in the training dataset.

Table III shows several indicators of the magnitude of the influence of each feature on the output.

The most influencing feature remains the planned interface HFE, with even higher importance than for one specific day. The distance from the planned and ideal exit or entry point remain as the next two features, confirming the importance of route structure and interfaces with nearby airspaces for flight-efficiency.

The ideal distance flown in the LFBBCA is in this case the fourth most important feature, just before the ideal distance in LFBBCA, average heading and ideal distance between origin and destination, highlighting the influence of the direction of the flight.

TABLE III. TABLE III FEATURE IMPORTANCE IN AIRAC 1702.

Feature	Relative importance
Planned interface HFE in LFBBCA	51.16%
Distance between planned and ideal exit point from LFBBCA	14.22%
Distance between planned and ideal entry point to LFBBCA	6.41%
Ideal distance in LFBBCA	4.86%
Ideal distance in LFFFCTA	4.36%
Average heading	3.42%
Ideal distance	3.17%
Ideal distance in LFRRCTA	2.34%
Planned local HFE in LFBBCA	1.89%
Ideal distance in LECMCTA	1.87%
Ideal distance in LFMMCTA	1.84%
Ideal distance in LECBCTA	1.75%
Reference FL in LFBBCA	0.69%
Take-off time - cosine	0.43%
Flights per ATCO in LECMCTA	0.22%
Take-off time - sine	0.20%
Flights per ATCO in LFBBCA	0.17%
Flights per ATCO in LFFFCTA	0.16%
Flights per ATCO in LFMMCTA	0.16%
Flights per ATCO in LFRRCTA	0.16%
Flights per ATCO in LECBCTA	0.15%
Number of the day	0.11%
Flights crossing LFBBCA	0.10%
Weekday - cosine	0.07%
Weekday - sine	0.07%

The ideal distances in the rest of contiguous ACCs have similar values of relative importance, together with the planned local HFE and reference flight level, as in the day of study.

The features linked to congestion (number of flights per sector, and take-off time) present an even lower influence than in the day of study.

Finally, the features concerning the day show the least influence, which indicates that the results are applicable during the whole AIRAC.

VI. CONCLUSIONS

This paper proposes a first approach for the use of machine learning techniques for performance assessment, in particular to assess flight efficiency at ACC level. Performance is predicted as a function of the flight properties by means of Random Forest regression. The regressor was used to evaluate and rank the relative importance of several influence factors for the interface HFE. The results suggest that the route structure has the most important influence, followed by the direction of the flight. Congestion and daily variability show a low influence.

This approach demonstrates the potential of machine learning techniques for analysing ATM performance, further research in these lines can tackle the problem of assessing new metrics and KPIs. However, this approach does not aim to challenge current indicators, but to provide a data-driven technique to analyse the influence factors of a given indicator. The resulting efficiency predictor is not meant to be used as a prediction tool but rather as an assessment tool of trends and correlations. Although the prediction power achieved is high for

the interface HFE, the predictor showed some error due to: (i) lack of some indicators, such as weather conditions; (ii) events such as convective clouds avoidance or ATC shortcuts that cannot be deterministically predicted. The consideration of such indicators would improve the prediction accuracy and allow to study the effect of weather and events on performance.

The proposed approach presents several novelties:

The conclusions are purely data-driven, which can enhance and update traditional influence diagrams with additional information. In this regard, the influence factors and hierarchy are coherent with the outputs collected from expert consultation, although some differences arose: for example, the daily variability and congestion showed much lower influence than expected from expert consultation, which may be explained because February is a month of low congestion and higher is needed to affect flight efficiency.

Machine learning techniques are able to model non-linear dependencies between variables that typical correlations may fail to capture. For instance, Pearson correlations are meant for linear correlations and rank correlations consider the contribution as whether positive or negative. Another advantage is that the machine learning models consider the whole set of variables to compute the importance of each one.

The analysis can be tailored to any airspace and time-range. This is particularly useful to analyse certain low performance episodes, like peak congestion hours or seasonality. In this sense, the presented approach focused on a specific use case of interest. Nevertheless, it would be interesting to analyse further airspaces and time ranges to find commonalities and trends.

The approach allows the evaluation of the influence of nearby airspaces and interfaces, which is a first step towards isolating the source of inefficiencies and enhancing performance target setting.

This type of analysis could be used to identify focus areas for performance improvement.

To generalise the application of the proposed approach, further development would be needed. The predictor could be further developed to assess other KPAs and more sophisticated flight efficiency metrics, such as vertical efficiency indicators and fuel consumption, and the approach should be extended to other ACCs and over several seasons, in order to find commonalities (global influence factors) and specificities.

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 and the teams of the projects APACHE and

AURORA, 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

- [1] EUROCONTROL, "PJ19: Performance Framework (2017)," no. October, pp. 1–134, 2017.
- [2] "Content Integration - SESAR 2020 project PJ19 (CI)." [Online]. Available: <https://www.eurocontrol.int/articles/content-integration-sesar-2020-project-pj19-ci>. [Accessed: 19-Sep-2018].
- [3] "ACCESS - Application of agent-based Computational Economics to Strategic Slot allocation." [Online]. Available: <https://www.access-sesar.eu/#/>. [Accessed: 19-Sep-2018].
- [4] "SATURN - Strategic Allocation of Traffic Using Redistribution in the Network." [Online]. Available: <http://www.saturn-sesar.eu/>. [Accessed: 19-Sep-2018].
- [5] "COMPAIR - Competition for Air Traffic Management." [Online]. Available: <http://www.compair-project.eu/>. [Accessed: 19-Sep-2018].
- [6] "VISTA - Market forces trade-offs impacting European ATM performance." [Online]. Available: <http://vista-eu.com/>. [Accessed: 19-Sep-2018].
- [7] "APACHE - Assessment of Performance in current ATM operations and of new Concepts of operations for its Holistic Enhancement." [Online]. Available: <https://apache-sesar.barcelonatech-upc.eu/en>. [Accessed: 19-Sep-2018].
- [8] "AURORA - Advanced User-centric efficiency metrics for air traffic performance Analytics." [Online]. Available: <http://aurora-er.eu/>. [Accessed: 19-Sep-2018].
- [9] Performance Review Body, "Report on the preparation of the revision of the SES Performance Scheme addressing RP2 and beyond Prepared by," 2012.
- [10] A. Bujor and A. Ranieri, "Assessing Air Traffic Management Performance Interdependencies through Bayesian Networks: Preliminary Applications and Results," vol. 4, pp. 34–48, 2016.
- [11] L. Sherry and A. Nanduri, "Anomaly detection in aircraft data using Recurrent Neural Networks (RNN)," *Integr. Commun. Navig. Surveill.*, p. 5C2-1-5C2-8, 2016.
- [12] R. Marcos et al., "Visual Analytics and Machine Learning for Air Traffic Management Performance Modelling," in *SESAR Innovation Days 2016*, Delft, The Netherlands, 2016.
- [13] EUROCONTROL, "Support study for an ex-post evaluation of the SES performance and charging schemes," 2017.
- [14] L. Tabernier, R. Dowdall, S. Gillet, and P. Simonsson, "Use of Influence Diagram techniques to understand the influences and assess the environmental impact of ATM," 2010.
- [15] EUROCONTROL, "Performance Indicator – Horizontal Flight Efficiency," 2014.
- [16] PJ19, "Vision of the Future Performance Research in SESAR," D.4.0.40, 2018.
- [17] INTUIT Consortium, "D5.1 Performance Monitoring and Management Toolset," 2018.
- [18] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. 2013.
- [19] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *JMLR*, vol. 12, pp. 2825–2830, 2011.
- [20] S. C. Wong and M. D. McDonnell, "Understanding data augmentation for classification: when to warp?," 2016.
- [21] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [22] D. Howell, M. Bennett, J. Bonn, and D. Knorr, "Estimating the En Route Efficiency Benefits Pool," *5th USA/Europe Air Traffic Manag. ...*, pp. 1–10, 2003.