

# TREE model: a tool to explore delay reduction scenarios in the ECAC area

B. Campanelli, P. Fleurquin,  
V. M. Eguíluz, J. J. Ramasco  
Instituto de Física Interdisciplinar y  
Sistemas Complejos IFISC (CSIC-UIB)  
Campus UIB, 07122 Palma de Mallorca, Spain  
Email: bruno@ifisc.uib-csic.es

C. Ciruelos, A. Arranz  
Transport and ICT Directorate  
Ingeniería de Sistemas para la Defensa  
de España, S.A. ISDEFE,  
Beatriz de Bobadilla 3, 28040 Madrid, Spain  
Email: aarranz@isdefe.es

**Abstract**—Air transportation systems display a rich phenomenology connected with several key topics in Complexity Science, such as complex networks, cascading failures and percolation. One example is flight delays that have usually the origin in primary events localized in limited areas of the network, but can later multiply and magnify as the daily operations go along. Given the large costs that delays convey, it is important to characterize their propagation and to model with predictive power the potential components or areas of the network affected. In this paper, we discuss the validation of an agent-based model, developed within the framework of the SESAR WP-E TREE project and aiming at simulating the propagation of delay in the ECAC airport network. Simulation outcomes are systematically confronted with empirical flight performance, the results show a good level of agreement with accuracies and precisions. Furthermore, we use the model to assess the effect on delay reduction in the network of two delay reduction scenarios: dropping passenger connections if the delay induced goes beyond a given threshold  $\tau$ , or decreasing the service time of aircraft in the airports if they delay is larger than a given value  $\Delta$ . Our results show how optimal values of  $\tau$  and  $\Delta$  can be found within the simulation framework.

**Keywords**—Reactionary delays; Complexity Science; Disruption Management; Network Performance

## I. INTRODUCTION AND BACKGROUND

Yearly direct costs of flight ATFM delays (not counting indirect costs such as reputation damages suffered by the airlines or business opportunities missed by the passengers) amount in Europe to over one billion euros [1]. Understanding the patterns in which delay propagates through the airport network is therefore a problem of high economical relevance, besides also being interesting from a more theoretical point of view as a real-world example of complex system displaying a rich dynamics. Since airlines and airports operate as a highly interconnected network, they can easily be subject to cascading failure effects: a local disruption originated in one part of the system can spread and multiply, affecting other parts which may be geographically far away and/or not be connected in an obvious way with the source of the perturbation.

The state of the art in this subject is predominantly focused on the US system, even in cases when the investigation is performed by European researchers. This might be caused

by the fact that data are more readily available for the US [2]. Optimization of airline schedules for robustness against delay spreading has received much attention. For instance, [3] described a model to produce schedules minimizing the crew cost and maximizing the number of crew members who are available for swapping during operations in case the necessity arises. On the other hand, [4]–[6] focused respectively on maintenance routing constraints, redistribution of existing slack and multi-objective optimization. [7] identified the reduction of primary delays in the earlier part of the schedules as a key ingredient to limit the propagation of delays, and [8] carried the analysis further, considering the relationship between the system's resilience and the schedules of both aircraft and crew. In [9], several of the authors of the present paper introduced a model to reproduce the propagation of delay in the US airport network. The model was later employed to study the robustness of the network to perturbations at the flight and airport levels [10]. [11] presented a mathematical model of delay propagation taking into account the explicit distinction between controllable and random factors influencing delay propagation, with the goal of estimating the needed adjustments to slack and flight time allowance to minimize the impact of the random components, namely the variable airport turnaround time between flights and the variable flight time. The approach presented in [12] models each airport as a stochastic queuing system, and uses a delay propagation algorithm to update schedules and demand levels at the network level in response to local delays. A statistical departure delay prediction model for a single airport is developed in [13], classifying delay-inducing factors as seasonal trends (such as seasonal demand change), daily propagation patterns and random residuals, i.e. those not accountable in the first two categories, such as unexpected mechanical failures. On the European side, [14] found that approximately 50% of delays for low-cost airlines were reactionary, while the percentage dropped to 40% for regular airlines. This is consistent with the estimates published by Central Office for Delay Analysis (CODA) of Eurocontrol [15]. [16] used a mesoscopic model to capture the emergence of network properties such as performance degradation, behavior predictability, amplified impact of exter-

nal events and geographical stability, obtaining straightforward performance results associated to specific flight prioritization rules. Resilience to delay propagations depends strongly on the strategies employed by the airlines; [17] offers a good description of the most important factors to be considered, while [18] argued that an approach based on passenger-centric metrics is desirable. Last year, we also introduced an agent-based (TREE) model extending and adapting the model developed for the US [9] to the different way of carrying out network operations of the ECAC area [19]. These modeling approaches can help to tackle scenarios such as [20], which uses a dynamic programming approach to solve the problem of finding the right amount of time a late passenger should be waited.

In this paper, we show the systematic validation of the TREE model by comparing model predictions and empirical flight performance data. The validation is implemented at three different scales: predictions on delayed flights, congested airports and major network congestion. In all three levels, the TREE model predictions adjust to empirical events with high precision and accuracy. Since the performance data come from different sources, it is also possible to study the impact of uncertainty of the input in the final model performance. After the validation of the model, its potential for policy assessment is shown by considering two delay reduction scenarios and how they affect the system level congestion. The first scenario considers a threshold to the maximum time that an airline waits for a passenger before dropping the connection. The model allows to explore the possible threshold values systematically and to search for an optimal value. Similarly in the second scenario the service time of an aircraft in the airport is reduced (with a cost) to recover delay. This second mechanism does not have such a strong impact in the network, although an optimal time before triggering this process may also exist.

## II. MODEL DESCRIPTION

### A. Usage of empirical data within the model

The model follows the state of each aircraft and airport as the aircraft attempt to perform the scheduled flights in their daily rotations. Delay can be propagated between flights through several mechanisms: aircraft rotations, passenger and crew connections, and limited airport capacities - the maximum number of aircraft movements which can take place during each hour. The model is data-driven in the sense that as many details of the simulated system as possible are reconstructed from empirical data, accounting for airport capacities, passenger connectivity patterns, flight schedules and primary delays. The output of each simulation run is primarily the amount of reactionary delay each flight has suffered, as well as the structure of the simulated delay propagation trees, i.e. from/to which flight the delays were propagated. Since stochastic processes are involved, the results need to be averaged over multiple runs when compared with empirical data, as described in section III.

The capacities of the airports are obtained from Eurocontrol's DDR2 (Demand Data Repository). Capacity of other

airspace structures than airports such as sectors has not been considered. In the case of strong external perturbations, such as bad weather or strikes, airports typically operate at a lower capacity, and this must be taken into account in the modelling in order to obtain good results. In this paper, however, we are focusing on the validation of the model in nominal conditions (those in which no major external disruptions as extreme bad weather or strikes affect the system). Note that the airports may have more than one capacity value depending on the operational configuration of the runway and the wind intensity and direction. However, here for simplicity sake, a single capacity value per airport is considered.

Passengers on one flight might connect to another at the destination airport, and the latter might need to wait for them if the airline determines it is economically convenient. In our model, this process is represented in a stochastic way. Each flight  $F_i$  has a set of *connection candidates*  $F_j$ , which satisfy the properties:

- $F_j$ 's origin airport coincides with  $F_i$ 's destination.
- $F_j$ 's scheduled departure time lies in a time window starting at  $F_i$ 's scheduled arrival time plus a buffer time  $T_P$ , and ending at  $T'_p$ .  $T_P$  represents the minimum time needed for passengers to transfer from one flight to the other. In the model, it is the same for all airports and flights, and set at 45 minutes, while  $T'_p$  is set at three hours.

The actual connections are selected randomly between the candidates at the beginning of each simulation run. The probability of a flight being selected depend on the airlines, origins and destinations of both  $F_i$  and  $F_j$ , and the time period (the month) in which the day to be simulated is located, and were determined by analyzing market sector data obtained from Sabre [21]. Connections are only allowed between flights operated by companies in the same alliance, and passengers of point-to-point airlines (such as Ryanair or Easyjet) do not have connections. We introduce an effective parameter  $\alpha \in [-1, 1]$ , so that there are no connections for  $\alpha = -1$ , all passenger connect for  $\alpha = 1$ , and connection probabilities are the same as in the market sector data for  $\alpha = 0$ . For every other value of alpha, the probabilities are obtained by linearly interpolating between the extreme cases above. The market sector data of Sabre contains information on the monthly flows of passengers in each route. It is indicative of the possible connections but it does not necessarily reflects the daily flows. The parameter  $\alpha$  is thus needed to modulate the importance of the connections. Although its value should be close to zero ( $\alpha = 0$  can be used in absence of further information), its optimal may vary from one day to another.

As a final flight coupling mechanism, crew members may also connect from one flight to another. Typically, the effect of crew connectivity in the model is minor since the crew teams are maintained together along the aircraft rotation. However, in some cases the teams may split and some crew members may connect to develop their tasks on other flights. The absence of a crew member is major issue, if we are

talking about pilots or co-pilots it may ground the aircraft until the arrival of the connection. This is implemented in the model in an effective way by adding a certain probability of connecting flights within the same airline. The probability of a two-flights connection is approximated by the fraction of connecting passengers of the airline in the particular airport multiplied by a tuning parameter  $\gamma$ . In hub-and-spoke airlines, this method ensures that the crew connections occur in the main hubs of the company. While in the case of point-to-point airlines, the connections are set in the logistic hubs that are the only airports for which we assign a non zero crew connection probability. The simulations presented here have been performed with  $\gamma = 0.01$ . Other values of  $\gamma$  have been also tested but the results are not very sensitive to this parameter as long as it is maintained within a reasonable range. Passenger connections play usually a much more significant role than crew connectivity.

### B. Simulation strategy

A detailed description of all the elements of the model was provided in [19]. We summarize next the most relevant mechanisms involved in the simulations presented in this work. The model uses an event-based approach, where events consist of determining whether a flight is ready to depart or delayed ( $\varepsilon_F$ ), or whether the traffic demand at one or more airports exceeds the nominal capacity ( $\varepsilon_A$ ). Initially, the (priority) queue of events is populated as follows: one event  $\varepsilon_F(i)$  is placed at the scheduled departure time of each flight  $F_i$ , and one  $\varepsilon_A$  is placed at the beginning of each hour.

When executed, events of type  $\varepsilon_F(i)$  will cause flight  $F_i$  to depart if its actual departure time  $\tau$  (sum of scheduled departure time and total delay) coincides with the current time in the simulation. If the actual departure time has not yet been reached, another  $\varepsilon_F(i)$  is scheduled at  $\tau$ , in case  $i$  acquires further delay before being able to depart. Then, delay is propagated to the connections according to the following rules:

- If  $F_j$  is the next leg of the  $F_i$  rotation, then there must be  $T_T$  between the actual arrival of  $F_i$  and the departure of  $F_j$ .
- If  $F_j$  must wait for passengers coming from  $F_i$ ,  $T_P$  minutes must pass between the actual arrival of  $F_i$  and the actual departure of  $F_j$ .

Where  $T_T$  and  $T_P$  were defined in section II-A, and  $F_j$  acquires reactionary delay so that the above conditions are met. Once a flight has departed, it will arrive at destination at the scheduled arrival time plus its delay, i.e. the duration of all flights is fixed, and can no longer be influenced by any other event. This is a simplifying assumption to avoid the complexity of dealing with trajectories, sectors' congestion and weather effects on aircraft navigation. Typically, the flights within the ECAC area are not long so this approximation should hold but this is one of the limits of the model. If a flight has accumulated a delay of over four hours, or has been delayed past the end of the day, it is cancelled (and so are the next legs of the rotations, if any) instead, so it does not propagate delay any further.

Note that even though this condition may seem excessively restrictive, flights departing past the end of the day do not contribute to the evolution of the largest cluster (see section III), and cancellations during the day happen rarely in the simulations.

Events of type  $\varepsilon_A$  check the capacity demand at each airport, by counting the number of flights departing from or arriving at the airport taking the delays into account. At every airport with excess demand, the following actions take place:

- The hour is divided into a number of consecutive *slots*, whose number is equal to the airport's capacity.
- The affected flights choose a slot each, on a first-scheduled-first-served basis. Scheduled departure (arrival) is used for the departing (arriving) flights to determine the priority. Each flight chooses the free slot closer to its scheduled departure (arrival) time, and its delay is updated if needed.
- The flights that are not able to find a slot are "passed over" to the next hour, and the procedure is repeated until there are no more flights to re-schedule.

This method can be seen as a simplified version of the CASA (Computer Assisted Slot Allocation) algorithm. One exception to this rule is that flights which have already departed at the moment this procedure takes place are allowed to land regardless of the capacity constraints. Finally, there exists the possibility that both the origin and destination airports of the same flight are having demand problems at the same time period. In this case, the procedure might return two different pair of arrival/departure times for the same flight, one corresponding to the "preference" of each airport, and the earlier possible combination is chosen.

## III. VALIDATION METHODOLOGY

The validation of the model and the analysis of the results are based on the notion of cluster of congested airports. Since we are interested in modelling the European air transportation system at the daily level, we consider a different airport network for each day, so that two airports are connected if there is at least one direct flight between them in the day's schedule. We define an airport as congested if the average departure delay (calculated hour by hour) exceeds a given threshold, the proper value of which is to be determined so that it is coherent with the input dataset (see section IV). As a way to track the evolution of system-wide congestion, we use the size of the largest cluster of congested airports, i.e. the largest subset of the network satisfying the following properties:

- All the nodes in the subset are congested.
- For each pair of nodes in the subset, there is at least one sequence of adjacent edges connecting them that is entirely contained within the subset.

In principle there can be many congested clusters, but for the sake of brevity we will use the term "congested cluster" to refer to the largest one. The central idea is that the largest size reached by this cluster through one day could be a good proxy of whether the system displayed network-wide

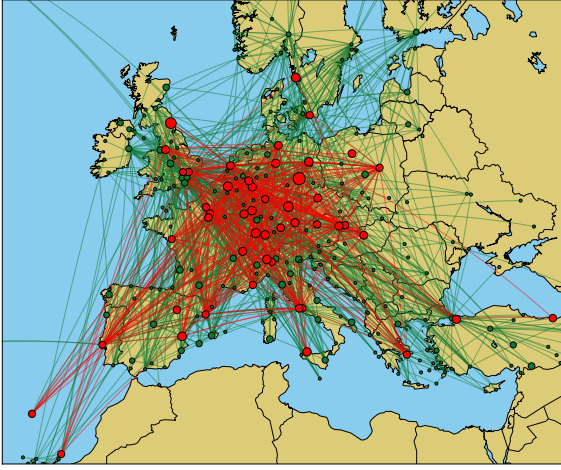


Fig. 1. European airport network from the 20 Jun 2013 Dataset. The airports in the largest congested cluster between 20:00 and 21:00 (CET) and their connections are highlighted in red. Node size is proportional to the average departure delay during the considered hour.

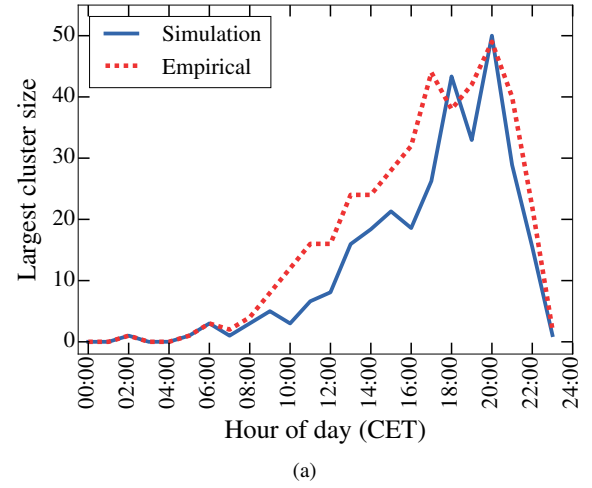
disruption problems, so the parameters of the model should be selected in such a way that correctly reproduces this value. Therefore, for each day it is possible to find a best-fit value of  $\alpha$  by minimizing the difference between empirical and simulated peak cluster size.

Once that is done, a more thorough analysis can be performed on the results of the simulations. From the qualitative point of view, we expect that if the model performs well, it should be able to reproduce the temporal evolution of the largest cluster's size and the total cumulative delay. Quantitatively, on the other hand, it is important being able to predict whether a particular flight will be delayed or an airport will be congested. Due to the stochastic nature of the model, delayed flights and congested airports may change from one model realization to the other. In an ideal context, the empirical situation would correspond to one of these model runs. We need to find, therefore, an alternative pragmatic way of assessing the quality of the model predictions.

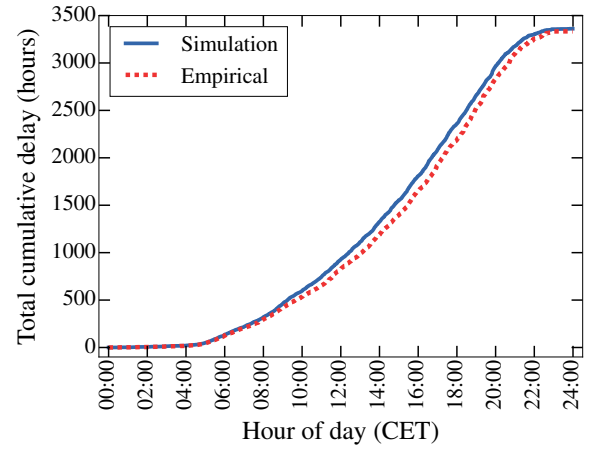
For the flights, this is done using delay probabilities across model runs by means of the following procedure:

- At the end of each simulation run, find the  $N_d$  flights with the highest total reactionary delay, where  $N_d$  is the number of reactionary delayed flights in the empirical data.
- The  $N_d$  flights which appear most often in the list of delayed flights across model runs are those most likely to incur into reactionary delays and they are compared with their empirical counterparts. The rest of the flights are considered to be non-delayed.

It is important to note that flights with primary delay in the input schedule are excluded from the above analysis, since they will always be delayed after every realization. The procedure for the airports is analogous, but the ranking is based on how many time the airports appear in the simulated largest cluster during the hour in which the cluster's size reaches its peak,



(a)



(b)

Fig. 2. Comparison between empirical data and simulation for the 20 Jun 2013 dataset: size of the largest congested cluster (a) and total cumulative delay (b) as a function of time.

and the number of airports predicted as delayed is equal to the number of airports in the empirical cluster at peak time. Once these steps are done, we test the predictions of the model using the standard definitions of accuracy and precision for binary classification problems, i.e.

$$\text{Accuracy} = \frac{TP + TN}{S}, \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FN}, \quad (2)$$

where  $TP$  is the number of true positives (correctly predicted as delayed/congested),  $TN$  the number of true negatives (correctly predicted as non-delayed/non-congested),  $FN$  the number of false negatives (incorrectly predicted as non-delayed/non-congested), and  $S$  the sample size. The accuracy alone cannot, in our particular case, account for the predictive power of the model, since in the datasets we used the number of positives typically is much smaller than the number of negatives. The use of the accuracy alone requires a more balanced configuration.

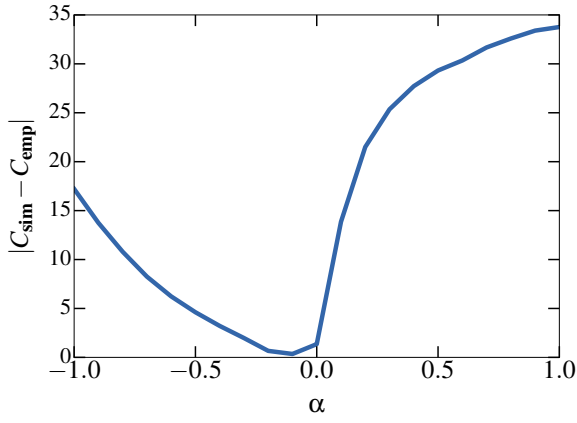


Fig. 3. Absolute difference between average largest cluster size in the simulations and empirical value vs. connectivity parameter  $\alpha$ .

As a benchmark on the quality of the model predictions, we will compare the results with those of a fully stochastic null model. For this model, the set of flights under consideration marked as "delayed" or "non delayed" are selected at random out of all the positives and negatives in the empirical data. This procedure mimics a completely uninformed guess on which flights may be delayed or not

#### IV. RESULTS

##### A. Using CODA data

The model is data-driven and so it needs the daily flight schedules and the primary delayed flights as inputs. The reactionary delays are then used for the validation. The first dataset that was had for validation corresponds to the ECAC network performance on June 20, 2013, and was obtained from the Central Office for Delay Analysis (CODA) of Eurocontrol. This data correspond to a day characterized by relatively high congestion in the absence (as far as we have been able to determine) of major, disruptive external events such as generalized bad weather conditions, strikes or technical problems. The dataset contains information on 19969 flights, of which only those internal to the ECAC area were considered. In a limited number of cases, aircraft rotations presenting "holes" (i.e. flights originated at a different airport than the one where the previous flight ended) were found and discarded. The schedule used as input for the simulations thus includes 15531 flights. Of these, 6549 have primary delays which serve as initial conditions, and are therefore excluded from the validation procedure, which is applied to the remaining 8982 flights. Fig. 2 shows the temporal evolution of size of the largest congested cluster and total cumulative delay for the real system and the simulated one. The value of  $\alpha$  used, determined according to the procedure described in II-A, is  $-0.02$ , and the goodness of fit vs  $\alpha$  is shown in Fig. 3. In this case we are using  $\Theta = 26.7$  minutes, i.e. the average delay per delayed flight over 2013 according to the 2013 CODA report [15]. In both cases, there is good qualitative agreement between empirical data and simulation. Specifically, we note that the position

TABLE I  
CONFUSION MATRICES FOR THE JUN 20, 2013 DATASET

Flights		Airports	
True Positives	False Positives	True Positives	False Positives
991	765	28	21
False Negatives	True Negatives	False Negatives	True Negatives
765	6461	21	352

TABLE II  
ACCURACY AND PRECISION FOR THE JUN 20, 2013 DATASET

	Flights		Airports	
	Our Model	Null Model	Our Model	Null Model
Accuracy	82.9%	68.5%	90.0%	79.5%
Precision	56.3%	19.5%	57.1%	12.3%

of the cluster peak is correctly identified, and the relative difference between empirical and simulated total delay is of the order of  $10^{-3}$ .

Tables I and II summarize the predictions of the model for this dataset at the single flight and single airport levels, with both accuracy and precision significantly higher than the null model. It is worth noting that, since there are many more non-delayed flights than *reactionarily*-delayed ones, a trivial model predicting no delay propagation at all would result in an accuracy of 80.4%, higher than the null model and close to the model being discussed, but its precision would be exactly 0%.

##### B. Extending the results, Flightradar24 data

A single day of data may be enough for a preliminary assessment of the capabilities of a model, but clearly a more extensive validation is needed. If the correct parameters have to be determined by fitting the data, there is no way of

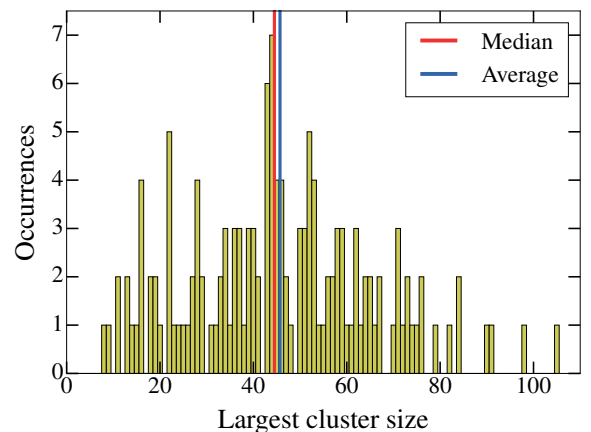


Fig. 4. Distribution of daily largest cluster sizes for the Flightradar24 dataset.

determining whether a particular schedule with a particular set of initial conditions will develop high congestion or not. The solution is to calibrate the model over a large set of days, finding a set of parameters which provide an overall best fit. This is the same strategy that has been employed on the US system in [9]. We could not access further CODA data on nominal conditions so in order to move ahead, we have been collecting data from Flightradar24 for several months: the analysis presented here covers a time period of 140 days, starting in March 3, 2015. Flightradar24 [22] is a well-known website, employing a crowd-sourcing strategy to collect and publish information emitted by aircraft via transponders. Due to the service's purposes and implementation, the data that can be collected from it are not as reliable as those coming from official sources, so additional cleaning is needed, and must be collected in real time.

The first issue to deal with is the fact that actual departures and arrivals recorded in this database are the actual take-off and landing times, so they have to be corrected by the taxi times, otherwise trying to determine the delays by taking the difference between actual and scheduled departure times would lead to an overestimation. The taxi times per flight cannot be accessed, so we solved the problem pragmatically by subtracting the average taxi-out (taxi-in) time at origin (destination) airport to the actual departure (arrival) time of each flight.

The problem with "broken" rotations in the CODA data mentioned in section IV-A is also present in Flightradar24 data, but in this case discarding the rotations with missing flights would be problematic, since they are more than in the CODA data. The number of "holes", however, is relatively small (no more than a few hundreds per day), so we fixed them by introducing "ghost" flights in their place in the rotations. The purpose of these is simply to propagate delay (if any) to the successive part of the rotation, so they are not taken into account in the analysis of the results.

Finally, since no information about the cause of delays is present in the data, a way of identifying flights with primary delay is needed. First, we define for each flight that is not the first in a rotation the quantity

$$\rho_{i>0} = \frac{d_D^i - d_A^{i-1}}{d_D^i + d_A^{i-1}}, \quad (3)$$

where the index  $i$  denotes the flight's position within its rotation,  $d_D^i$  is its departure delay and  $d_A^{i-1}$  is the arrival delay of the previous leg. The idea is that a negative value of  $\rho_i$  means the aircraft has used the turnaround time to recovered delay between flights, while if  $\rho_i$  is positive it has gained additional delay. Then, for each day, we sort all the delayed flights by their  $\rho$  in descending order, with the ones which are the first ones in their rotations on top, and the delays of the first  $X N_{\text{Flights}}$  flights (with  $X$  to be determined later) are considered to be primary.

We motivate this choice by noting that these large, anomalous increases in delay are better candidates to be primary since in the case of very large reactionary delays the airlines

TABLE III  
PREDICTION OF CONGESTED DAYS IN THE FLIGHTRADAR24 DATASET:  
CONFUSION MATRIX (A) AND ACCURACY AND PRECISION (B)

<i>True Positives</i>	<i>False Positives</i>			
61	11	<i>Our Model Null Model</i>		
<i>False Negatives</i>	<i>True Negatives</i>	Accuracy	85.7%	50.0%
9	59	Precision	82.7%	50.0%
(a)		(b)		

would react by, for example, cutting the connection. The question is then how to determine  $X$ . In the CODA dataset, 78% of the flights have primary delay, but a flight could have a small amount of primary delay and a large amount of reactionary, and for the Flightradar24 database we have no way of separating the total delay by cause, so a value close  $X = 0.78$  would lead to introducing an excessive amount of initial delay into the system. We chose the value  $X = 0.6$  after taking the following considerations into account:

- In the CODA database, 60% of the delayed flights have primary delay larger than 150% of their reactionary delay, i.e. their delay is predominantly primary.
- By applying this procedure to each day of the Flightradar24 data, the ratio between primary delay and total delay is typically close to 0.58, which is the value reported by CODA for the year 2013.

Since this procedure is likely to introduce errors in the determination of delays, we cannot expect to reproduce with high precision metrics depending on the exact values of the delays, such as the total cumulative delay or the distribution of delays. However, we are mostly interested in determining the probabilities of flights being delayed and airports and whole days of operations displaying high levels of congestion, and this dataset is good enough for this purpose. Note, however, that there was a price to pay in the accuracy and precision of the method for not knowing the primary delays. In the case of the accuracy, the drop from the CODA case to the average value in Flightradar is of approximately 17%.

Like with the CODA dataset, we proceed by comparing the simulations' outcomes with empirical data. In this case, however, after determining the best value of  $\alpha$  for each day, we selected the value  $\alpha = 0.011$  (the median value of the distribution of fitted  $\alpha$ s), simulated again each day in the dataset with this value, and then calculated the confusion matrices. In this case we set  $\Theta = 19.23$ , the average delay per delayed flight over all the Flightradar24 data, *after* departure and arrival times have been corrected as described above. The results for each day are shown in Fig. 5; accuracy and precision are comparable with what was obtained with the CODA dataset. One exception to this are a small number of days where the precision drops to a very low value. All of them exhibit a rather small amount of congestion (and are correctly identified by the model as non-problematic days -

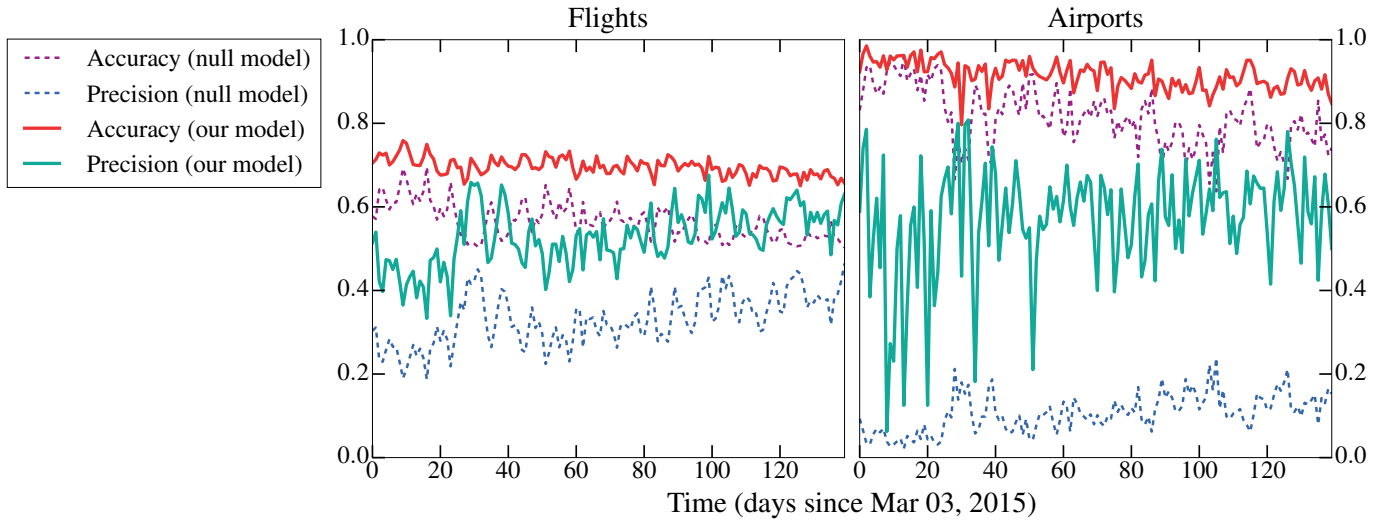


Fig. 5. Accuracy and precision for the Flightradar dataset.

see below). The problem can come from the uncertainty in the predictions associated to a day with very few flights with reactionary delays.

We then defined  $\Xi = 44.5$ , the median value of the largest daily cluster size over all days in the dataset (see Fig. 4 for the distribution of cluster sizes). If we consider a day as having high congestion when its largest cluster size  $\xi$  is larger than  $\Xi$ , the question arises whether the model is able to predict the truth value of  $\xi > \Xi$ , i.e. whether a schedule will develop significant system-wide congestion in response to the input delay. The results of this exercise are shown in table III. As can be seen, the model is able to predict highly congested days with high accuracy and precision.

## V. ASSESSMENT OF DELAY REDUCTION SCENARIOS

### A. Passenger Connections Moderated by Delay

As a first scenario testing, we focus our attention on the effect of cutting passenger connections after a certain amount of delay has been accumulated. In real operations, airline managers will decide to wait for a connecting passenger only if it is deemed economically convenient. This consideration can include factors such as the ticket class and the next connections of the passenger (long range connections may take preference). However, if the delay that waiting induces in subsequent flights there can be a moment in which the most advantageous solution is to drop the connection.

Here, we represent this situation by introducing a threshold in the delay  $\tau$  allowed beyond which the connection is disregarded and the propagation of delay cannot continue. This policy is applied in the simulation to all the airlines with passenger connections. By tuning  $\tau$ , one can pass from a situation in which virtually no passenger is waited to another in which only very long delays would justify the connection drop.

The simulations are run with  $\alpha = -0.02$ , as discussed in the previous section, and several values of  $\tau$ . The results

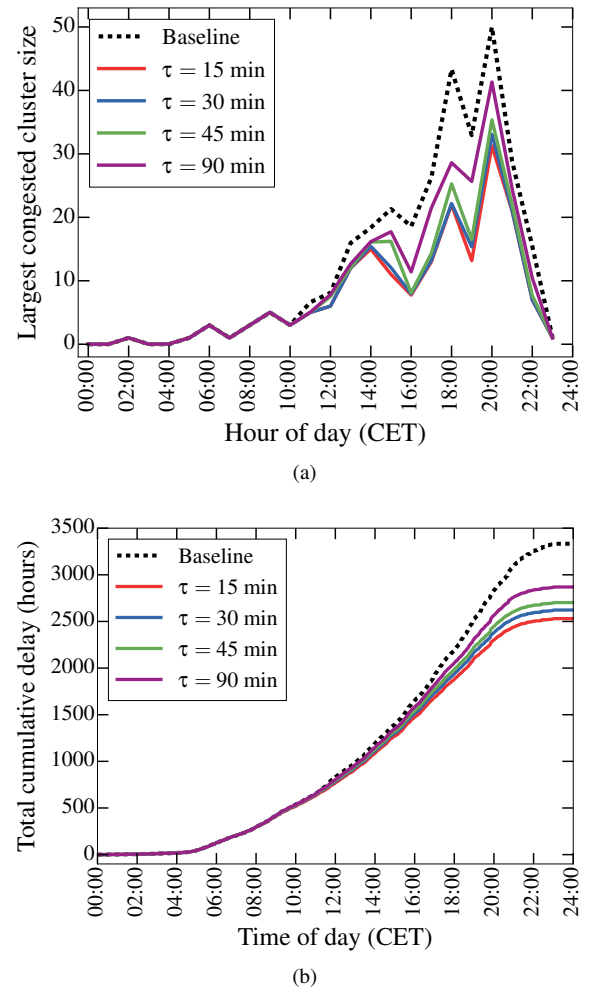


Fig. 6. Impact of cutting passenger connections after a certain amount  $\tau$  of delay on the congested cluster (a) and total delay (b).

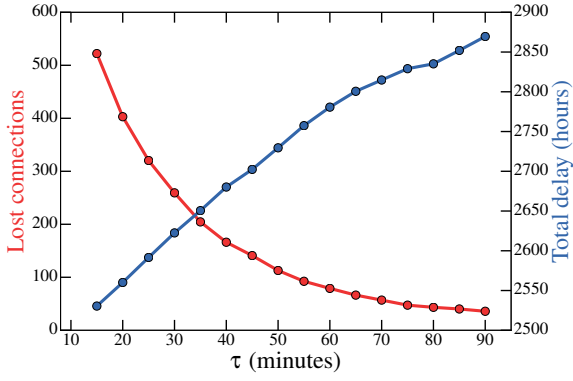


Fig. 7. Total lost connections and delay as a function of the maximum waiting time  $\tau$ .

for the full network are displayed in Figure 6 both for the total minutes of delay accumulated in the network until a certain hour and for the average size of the congested cluster. Lowering  $\tau$  the connections play a less significant role in delay propagation and the accumulated delay decreases. More interestingly, the effect on the congested cluster is non linear and relatively strong as can be observed if  $\tau$  goes from one hour and a half to 45 mins. Further reductions are not so impacting.

This new policy has, however, a catch: every connection dropped between flights means one or several passengers left on the ground and imply cost for the airlines in terms of rerouting passengers and fines. There is, therefore, a balance between the cost incurred due to the passengers not connecting and the savings obtained from the prevented delays. The TREE model does not include at this stage economic considerations, but it is possible to calculate the number of connections dropped and the improvement in the delay as shown in Figure 7. The lower  $\tau$  is, the more connections are lost. Assigning economic values to the delay and the connections, it is possible to find an optimal balance allowing for a maximum saving in delays while impacting as little as possible to the connections.

### B. Aircraft Service Time Depending on Delay

Another scenario that can be tested is recovering delay by accelerating ground operations, which in our model essentially means an on-demand reduction of the parameters  $T_D$ ,  $T_C$  and  $T_P$ , accounting for faster aircraft handling (un-boarding, boarding, fueling and baggage handling) and transfer of crew members and passengers - possibly moving them, with an extra cost, directly from one aircraft to the other without passing through the terminal. This action is also in the hands of the airport manager, instead of only on those of the airlines as in the previous case. Although, the acceleration in the ground operations can imply an extra cost also for the airports.

In this exercise, we introduce a new parameter  $\Delta$ , the threshold delay after which faster operations are triggered. This could be seen as complementary to the previous scenario, in the sense that this time the system attempts to reduce delay

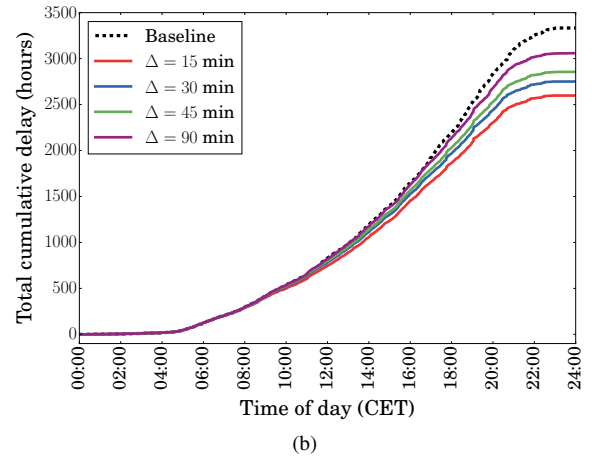
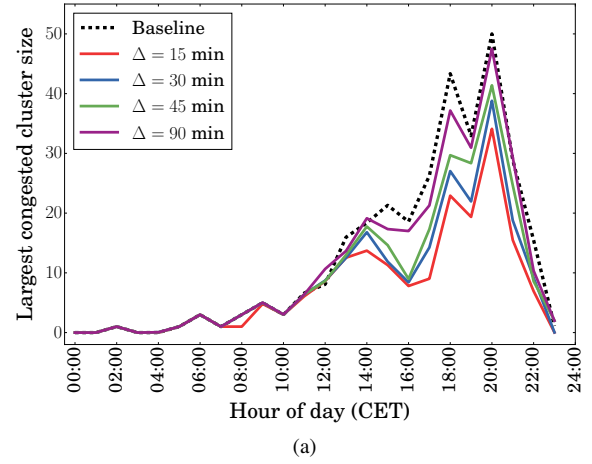


Fig. 8. Impact of a 60% reduction of ground operations time for aircraft arriving at destination with a delay greater than  $\Delta$  minutes on the congested cluster (a) and total delay (b).

by operating on the flights acting as a source of delay, rather than on those who are going to be delayed.

The reductions considered are to 80% and 60% of the minimum standard times in the model. Note that the baseline minimum times - 30 minutes for the aircraft and 45 for crew and passengers - were chosen after an expert consultation and are already relatively tight. Here it is worth stressing that these values are applied to all the aircraft and all the passengers: this is an oversimplification because different type of aircraft need longer or shorter service times and the minimum connection time of passengers depends on the distance between the particular gates and the presence or not of security controls in between. The results of the simulations at network scale with different values of  $\Delta$  are displayed in Figure 9 for 80% reduction and in Figure 8 for 60%. The effect with an 80% reduction is noticeable in the total accumulated delay and in the congested cluster size but not strong. When a 60% reduction is implemented instead, the effect on the network is clearer.

Finally, Figure 10 shows the evolution of the total delay as a function of  $\Delta$  for both 60% and 80% time reductions. In both cases, the relation appears to be non-linear. Since reducing

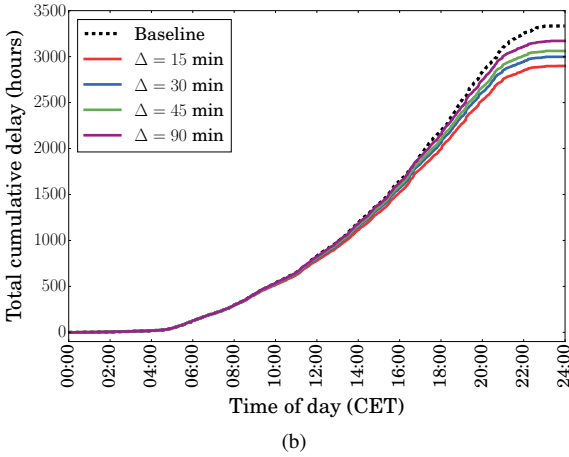
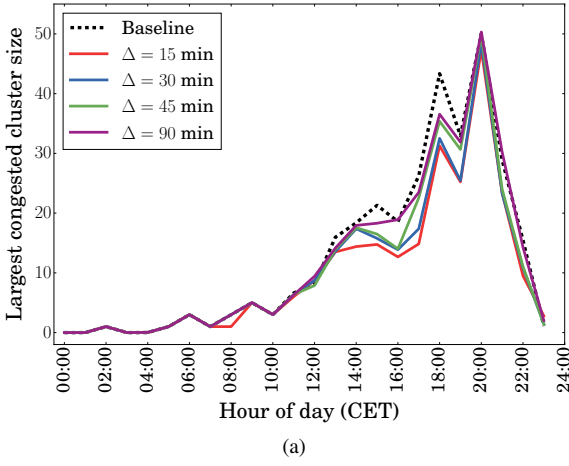


Fig. 9. Impact of a 80% reduction of ground operations time for aircraft arriving at destination with a delay greater than  $\Delta$  minutes on the congested cluster (a) and total delay (b).

service times translates into larger costs, it can be reasonably expected that, by calculating the savings introduced through this kind of policy is possible to find a value of  $\Delta$  for balancing costs and savings

## VI. CONCLUSION

We introduce a model for the propagation of delay in the ECAC area and validate it against empirical data obtained from different sources. The observed performance is satisfactory in all the levels: predictions on the probability of delay of single flights, congestion in airports and predictions on whether a schedule will develop large problems or not. The model performance can conceivably be further improved with the use of higher-quality input datasets. The model is easy to adapt to take into account different policies applied by the network managers or other stakeholders. The relative simplicity of the model implies that it is also efficient, which suggests that this kind of approach can be used to perform experiments on the system in order to assess the impact of, for instance, airline delay management strategies or severe weather perturbations. These questions are being investigated at the time of writing, and will be the subject of future work.

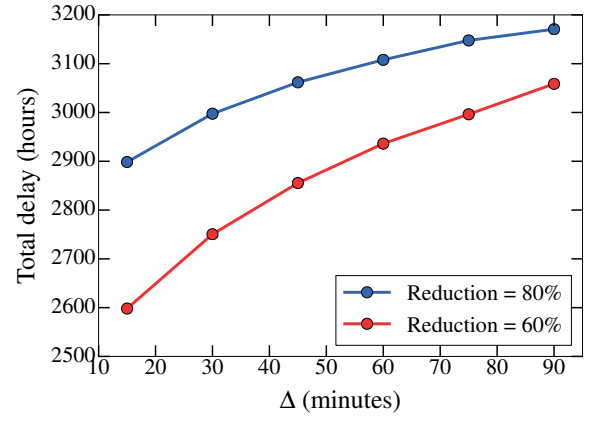


Fig. 10. Total cumulative delay at the end of the day at different thresholds  $\Delta$  for triggering a reduction of ground operations time.

We introduce a model for the propagation of reactionary delays in the ECAC area and validate it against empirical data obtained from different sources. The observed performance is satisfactory in all the levels: predictions on the probability of delay of single flights, congestion in airports and predictions on whether a schedule will develop large problems or not. The model performance can conceivably be further improved with the use of higher-quality input datasets. The model is as well easy to adapt to take into account different policies applied by the network managers or other stakeholders. In this sense, we have considered two examples of how the model can be used as a supporting tool in the policy assessment. Our focus was on the global impact of individual airline policies, but a more detailed analysis, e. g. at the single airline level, is definitely possible. Furthermore, even if we used the real primary delays, another interesting direction would be assessing the effectiveness of different policies against artificial configurations of primary delays, representing hypothetical crisis scenarios. The stochastic approach taken in modeling passenger connections naturally implies some level of uncertainty in the results, but also allows conclusions to be drawn from incomplete data, i.e. without knowing precisely the itinerary of each passenger.

In particular, we have analyzed two scenarios in which measures to reduce delay propagation in the network are tested. The first one consists of dropping passenger connections if the delay induced in the flights connecting get over a certain threshold  $\tau$ . This is a decision that should be in the hands of the airline managers. The simulations show that a balance between the number of dropped connections and the delay in the network can be searched. This is balance should mark the optimal threshold time for  $\tau$ . Depending on the particularities of each airline operations, this point can be in a different value of  $\tau$  but this is something easy to implement in the model.

The second case study contemplates a more complex scenario with a reduction on the service time of flights with delay over a certain value  $\Delta$ . The service time reduction involves airline and airport managers and can have a notable cost. We consider two reduction values 80% and 60%. Only for the

lowest values, one can see noticeable effects on the delay reduction in the network. As before an optimal value of  $\Delta$  may exist balancing cost and effect. These two are simple scenarios but they show already the potential of the model as policy assessment tool.

#### ACKNOWLEDGMENT

This work is co-financed by EUROCONTROL acting on behalf of the SESAR Joint Undertaking (the SJU) and the EUROPEAN UNION as part of TREE project under the Work Package E in the SESAR Programme. Opinions expressed in this work reflect the authors' views only and EUROCONTROL and/or the SJU shall not be considered liable for them or for any use that may be made of the information contained herein. B. Campanelli is funded by the Conselleria d'Educació, Cultura i Universitats of the Government of the Balearic Islands and the European Social Fund. J. J. Ramasco acknowledges funding from the Ramón y Cajal program of the Spanish Ministry of Economy (MINECO). The authors thank Flightradar24 for granting permission to use their data. Flightradar24 cannot be held liable for the accuracy of the information or for ensuring that the information is up to date at all times. It is thus not liable for any consequences that may arise as a result of possible inaccuracies in the information provided.

#### REFERENCES

- [1] Cook A. and Tanner G., European airline delay cost reference values, Performance Review Unit EUROCONTROL, 2011
- [2] Bourgois M. and Sfyroeras M. 2014. Open Data for Air Transport Research: Dream or Reality?. In Proceedings of The International Symposium on Open Collaboration (OpenSym '14). ACM, New York, NY, USA, pp. 17-24. DOI=10.1145/2641580.2641602 <http://doi.acm.org/10.1145/2641580.2641602>
- [3] Shebalov S. and Klabjan D., Robust airline crew pairing: Move-up crews, Transportation Science, 40:300-312, 2006.
- [4] Lan S., Clarke J. and Barnhart C. Planning for Robust Airline Operations: Optimizing Aircraft Routings and Flight Departure Times to Minimize Passenger Disruptions, Transportation Science, Vol. 40, No. 1, pp. 15-28, 2006.
- [5] Ahmadbeygi S., Cohn A. and Lapp M., Decreasing airline delay propagation by re-allocating scheduled slack, IEEE Transactions, Vol. 42, No. 7., pp. 478-489, 2010.
- [6] Burke E.K., De Causmaecker P. and De Maere G., A multi-objective approach for robust airline scheduling, Computers & Operations Research, Vol. 37, No. 5, pp. 822-832, 2010.
- [7] Beatty R., Hsu R., Berry L., and Rome J., Preliminary evaluation of flight delay propagation through an airline schedule. Air Traffic Control Quarterly 7, 259-270 1999.
- [8] Ahmadbeygi S., Cohn A., Guan Y., & Belobaba P., Analysis of the potential for delay propagation in passenger aviation flight networks, Journal of Air Transport Management 14, 221-236, 2008.
- [9] Fleurquin P., Ramasco J.J. and Eguiluz V.M., Systemic delay propagation in the US airport network, Scientific Reports, vol. 3, p. 1159, 2013.
- [10] Fleurquin P., Campanelli B., Eguiluz V.M. and Ramasco J.J. Trees of Reactionary Delay: Addressing the Dynamical Robustness of the US Air Transportation Network, Procs. SESAR Innovation Days 2014.
- [11] Wang P.T.R., Schaefer L.A., and Wojcik L.A., Flight connections and their impacts on delay propagation, Procs. of the IEEE Digital Avionic Systems Conference 1, 5.B.4-1-5.B.4-9, 2003.
- [12] Pyrgiotis N., Malone K.M. and Odoni A., Modeling delay propagation within an airport network, Transportation Research C 27, 60-75, 2013.
- [13] Tu Y., Ball M. O. and Jank W. S., Estimating flight departure delay distributions: A statistical approach with long-term trend and short-term pattern, J. Amer. Statist. Assoc., v103, pp. 112-125, 2008.
- [14] Jetzki M., The propagation of air transport delays in Europe, Thesis in the Department of Airport and Air Transportation Research, RWTH Aachen University, 2009.
- [15] CODA Digest - Delays to Air Transport in Europe - 2013
- [16] Network Congestion 2030 project, Final Report Volume II, Isdefe and Innaxis for Eurocontrol, 2010.
- [17] Palpant. M, Boudia M., and Others, ROADEF 2009 Challenge: Disruption Management for Commercial Aviation
- [18] A. Cook, G. Tanner, S. Cristóbal, M. Zanin, Schaefer Dirk (ed), Passenger-Oriented Enhanced Metrics, Proceedings of the SESAR Innovation Days (2013) EUROCONTROL. ISBN 978-2-87497-074-0
- [19] Campanelli B., Fleurquin P., Arranz A., Etxebarria I., Ciruelos C., Eguiluz V.M. and Ramasco J.J. Modelling Reactionary Delays in the European Air Transport Network, Procs. SESAR Innovation Days 2014.
- [20] Jacek Skorupski, Magdalena Wierzbńska, A method to evaluate the time of waiting for a late passenger, Journal of Air Transport Management, Volume 47, August 2015, Pages 79-89.
- [21] Market sector data have been purchased from Sabre: [www.sabre.com](http://www.sabre.com)
- [22] Flight performance data are publicly available from Flightradar24: [www.flightradar24.com](http://www.flightradar24.com)