Comparing the modeling of delay propagation in the US and European air traffic networks

Bruno Campanelli a, Pablo Fleurquin a, Andrés Arranz b, Izaro Etxebarria b, Carla Ciruelos b, Víctor M. Eguíluz a, José J. Ramasco a,*,

a Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Campus UIB, 07122 Palma de Mallorca, Spain
b Transport and ICT Directorate, Ingeniería de Sistemas para la Defensa de España, S.A., Beatriz de Bobadilla 3, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 31 July 2015
Received in revised form 8 March 2016
Accepted 18 March 2016
Available online 5 April 2016

Keywords:
Reactionary delays
Complexity science
Modeling delay propagation
Network performance

ABSTRACT

Complex Systems are those in which a very large number of elements interact, usually in a non-linear fashion, producing emergent behaviors that are typically difficult to predict. Air transportation systems fall in this category, with a large number of aircraft following a pre-scheduled program. It has been shown that it is possible to understand and forecast delays propagation in these systems. The objective of this analysis is to compare the modeling in the US and in the European air traffic networks, analyzing the propagation of delays due to failures in the schedule or to disturbances. We use two different agent based models recently developed to simulate the delays propagation and assess the effect of disruptions in the networks (US and ECAC areas). Our results show that a first-come first-served protocol managing the flights produces larger congestion when compared with an ATFM (Air Traffic Flow Management) slots priority system.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Among all the different means of transport, air transportation is the one that has experienced the fastest growth in the last century (Heppenheimer, 1995). According to the World Bank, in 2014 the number of domestic and international air passengers summed up 3.21 billions worldwide (World Bank, 2015), and it is expected to increase by 6.3% this year (ICAO, 2015). The rapid increase in demand comes at a high price, causing the transport network to become congested (Lan et al., 2006) (see also the evolution of the delays in Europe from the CODA digests of Eurocontrol since 1998 until the present (CODA). It is therefore of great importance to understand the interplay between the various components of the system. Delays are one of these components and have a great economic impact, a study for the US found that the costs imputable directly or indirectly to delays were around 40.7 billion dollars (US Congress, 2008). Delay related direct costs in Europe may look modest in comparison (1.25 billion euros) but still high (Cook and Tanner, 2011; Note). The intricacy and interaction between the elements that compose the air-traffic system qualifies it as a Complex System. Complexity is not used just to refer to complicated phenomena within Science; it emphasizes the notion of emergent behavior at the system level that surges from the interaction between its components. During the last decade, the scientific community has extensively studied these systems under the light of Network Science. In this context, air-traffic systems can be represented as networks whose vertices represent airports and its edges direct flights during a fixed period of time (Barrat et al., 2004; Li and Cai, 2004; Guimerà et al., 2005; Burghouwt and de Wit, 2005; Balcan et al., 2009; Gautreau et al., 2009). Several aspects of the air traffic network have been studied. The first works (Barrat et al., 2004; Guimerà et al., 2005) were focused on a topological description of the network structure. The results showed a high heterogeneity in the number of connections that bear each node (the so-called degree of a node) and the traffic sustained by each connection, finding a non-linear relation between the node degree and the fluxes of passengers in a given route (Barrat et al., 2004).

The Air Transportation Network can also be understood as the backbone where different dynamical processes take place. A story of notable success was the modeling and forecasting of disease spreading using air traffic data (Balcan et al., 2009).

Delay propagation dynamics can be also studied within this framework (Fleurquin et al., 2013, 2013b; 2014, 2014b; Campanelli

* Corresponding author.
E-mail address: jramasco@ifisc.uib-csic.es (J.J. Ramasco).
et al., 2014, 2015). Since airlines operate in an interconnected network, they are subject to propagation effects. A disruption in one flight or airport can quickly spread and multiply in cascade affecting other parts of the air transport network (Beatty et al., 1998; Allan et al., 2001; AhmadBeygi et al., 2008; Belobaba et al., 2009). The delay between flights may propagate due to several mechanisms: aircraft rotations, passengers and crew connections, or airport congestion. These factors are at the basis of the models developed to reproduce delay propagation.

Understanding how delays propagate in the airport network starting from primary events is thus of high economic relevance. In the last years, we have introduced two agent-based models to study and forecast delay propagation in the US and European networks (Fleurquin et al., 2013; Campanelli et al., 2014, 2015). The main difference between them is the method to prioritize the flight management in the airports. While in the US model a first-come first-served (FCFS) protocol is used, in Europe an ATFM (Air Traffic Flow Management) slot system is simulated (SM). This applies to the tactical phase of the flights and compresses processes such as slot reallocation and swapping. The purpose of this work is to compare the performance of the networks with each of these management systems. For this and since the models are data-driven, two days with large network congestion not caused by external disturbances have been selected: June 20, 2013, in Europe, and July 13, 2012, in the US. Both models are run in the same conditions and the results compared. Comparisons between the US and European networks have been carried out in the last years (Reynolds-Feighan, 2010; Vilaplana, 2010; Eurocontrol and FAA, 2013). Eurocontrol and the Federal Aviation Administration (FAA) published a joint report on the similarities and differences in ATM performance between the two areas (Eurocontrol and FAA, 2013). These studies take an empirical data analysis perspective, while this work is focused on the comparison of the ATM systems simulating both management systems in standardized conditions.

2. Materials and methods

2.1. Metrics

The results in this paper are analyzed in terms of two kinds of performance metrics. While the first one is the straightforward total cumulative delay in the system as a function of time, the second is less conventional, and it is intended to assess the level of network delays. We have previously used it in other works concerning network-wide congestion and delay statistics (Fleurquin et al., 2013, 2013b; 2014, 2014b Campanelli et al., 2014, 2015). First, we build a daily (unweighted) airport network using direct flights as edges. Then, for each hour of the day, we extract the sub-network containing the airports where the average hourly departure delay is above a given threshold, the value of which should be calculated over a long time period (e.g. a year or several months), so that the properties of the network can be analyzed in a stable way. As in all the metrics based on an arbitrary threshold selection, the exact results may depend on the particular value used but the general system trends must be consistent. We refer to the airports where the threshold is exceeded as congested airports, and use the size of the largest connected cluster found in the congested sub-network as an indicator of the presence of network-wide problems. It should be noted that this quantity contains information about correlation, not causation: two airports connected in the same congested cluster may be affected by each other, but the source of the delay is not identified.

2.2. FCFS (US) model

We will describe next the elements of the models starting by the simplest: the FCFS model adapted to the US air traffic. A detailed description of the modeling framework is provided at (Fleurquin et al., 2013). This model, as the SM one, needs as inputs the daily schedules of the flights that are typically extracted from flight performance data. A more detailed description of the datasets is provided in Section 2.4, but essentially the inputs needed are the scheduled arrival and departure times, aircraft and airline identification codes, flights’ origins and destinations, the airport capacities and the flights primary delays. Cancellations and diverted flights are not used in the model. With the aircraft code and the spatio-temporal information of the flights obtained from the data, we can reconstruct the aircraft rotations and consequently approximate the airline schedules throughout the day.

The model takes as basic units the aircraft and follows them as they complete the daily schedule. The minimum time resolution in the airport operations is 1 min. In the absence of disruption (primary delays) of any kind, daily operations would be carried out exactly as specified in the schedule. The flight operations are generated following three microscopic subprocesses that rule the agents’ reaction to each other and the system: aircraft rotation, flight connectivity and airport congestion. The rotation is the itinerary of each aircraft throughout the day, i.e., it goes from airport A to B and then to C following the scheduled arrival and departure times. An aircraft rotation is completed when all the previous legs have been fulfilled sequentially. A flight is not considered finished as far as the aircraft is in the gate-to-gate phase (offblock), which comprehends the taxi-in, taxi-out and airborne time. As a model simplifying assumption, it is not possible to absorb delay offblock.

Once an aircraft is at the gate, in the turn-around phase, it has to comply with a minimum service time (TS) for ground operations. For the sake of simplicity, the value of TS has been fixed at 30 min. The next model ingredient represents flight connectivity due to crew and passenger connections. It is implemented as a stochastic mechanism due to the lack of information on passenger and crew connections along the day. The fraction of passengers connecting in each airport is estimated from Market Sector Data (DB18 Ticket and T100 Domestic Market repositories of the Bureau of Transportation Statistics (BTS)). Each flight (of the same airline) has a probability of connection proportional, with a factor α, to the connectivity levels of each airport. With this in mind, a connection is randomly chosen by considering flights of the same airline within a time window of the scheduled arrival time of 3 h. A flight is able to depart if and only if all its connections have already arrived. Note that the connections are at flight level, they may represent the connections of several passengers or of a single one in first class. The important issue is that once assigned, the flights must wait for their connections. The calibration of α to reproduce the global level of delay in the network provides a way to estimate how many of these flight connections are present in the system. As a simplification, the minimum connecting time for passengers is set to zero. More involved versions of the models (Fleurquin et al., 2013b; Campanelli et al., 2015) have the minimum connecting time into account but this is a feature that can vary from airport to airport and impact differently both models so adding it can render harder the comparison of the models’ results.

Airports’ capacity is measured as the scheduled airport arrival rate for each hour (SAAR) of the day multiplied by a factor β. When a perturbation occurs, the demand at the airport may vary and the actual arrival rate can exceed the schedule rate. Whenever this happen the next incoming aircraft will have to wait in order to be served. A queuing protocol based on first-come, first-served (common operating procedure in the US) is implemented in each
airport. This process may produce congestion at the airport and propagate delays to flights of the different airlines.

In the US network, the FAA has implemented the Ground Delay Program (GDP) to try to reduce capacity problems in the airports. In (Fleurquin et al., 2013b), the model was tested simulating a mechanism similar to GDP. The application of this measure was restricted to a set of airports affected by weather perturbations. In the present work, the following approach is used: if the load of an airport is larger than four flights and goes beyond 1.5 times capacity a GDP measure is declared. This implies that aircraft still on the ground and directed to this airport should wait to depart until the moment in which their arrival would coincide with the expected solution of the capacity problems. The order of arrival of the flights is maintained.

### 2.3. SM (European) model

The details of this model have been published in (Campanelli et al., 2014, 2015; 2015b). The basic structure is similar to the FCFS model, with an agent-based approach taking the aircraft as the minimal units. The aircraft are followed along their daily rotations, there is also a minimum service time (T_s set at 30 min) and maximum capacity values for the airports. As in the FCFS model, flights are not allowed to recover delay offblock. As described in the previous works (Campanelli et al., 2014, 2015), the flight connectivity of this model is more elaborate than the effective approach taken in the US. It involved the use of Market Sector Data (Sabre, 2015) to estimate the passengers’ connectivity levels between airlines of the same alliance in the different airports. Still, for the sake of a fair comparison, we have simplified this point and the same rule as in the US has been implemented with a tuning parameter \( \alpha \) to control the connectivity level.

The main difference between both models refers to the way flights are managed in the airports. While in the FCFS model a first-arrived first served protocol was implemented, the SM model runs with a system of priority based on ATFM slots (first-planned, first-served). The model uses the daily schedules as an input including the flight arrival and departure times in the program so these slots refer only to flow operation elements. When a flight \( F \) loses its ATFM slots, the model tries to assign it a new suitable pair of departure (at origin) and arrival (at destination) slots, first through rescheduling and then through slot swapping. If these processes fail, the flight and all the successive legs in the same aircraft’s rotation are canceled.

Going step by step, if \( F \) lost its slots then a new “proposed departure time” is searched for. This time is given by the earliest time at which, having waited for all the other flights to which it is connected and dealt with its own primary delay, \( F \) can depart counting with the possible capacity constraints at the airports of origin and destination. This implies that a delayed flight will not depart immediately unless there are free ATFM slots in both airports, otherwise the free combination closest in time will be searched. If there is no eligible pair of slots in a certain time window (\( \Delta T = 6 \) h), the re-scheduling procedure fails and the ATFM slot swapping mechanism is triggered.

Through slot swapping, the model tries to avoid the cancellation of \( F \) at the expense of delaying another flight \( G \) operated by the same airline \( A \). \( G \) is selected among \( A \)’s flights departing from or arriving at the same airport of \( F \) and with origin/destination deemed as “less important” than \( F \). As a proxy of the importance of an airport, we use the total daily movements (departures and arrivals). Once flight \( G \) is selected, it loses its slots in favor of \( F \) and after that it must be reallocated. Finally, if all this fails \( F \) is canceled.

### 2.4. Model assumptions

We now list the common assumptions used in the two models:

- The duration of each flight is fixed, i.e. en-route delay recovery is not allowed.
- Turnaround times and crew/passenger transfer times are fixed throughout the whole system, regardless of the specific airline, airport and aircraft type involved in particular connections and rotations.
- Intercontinental flights are not taken into account.
- The flight connection probabilities are only a function of the airport where the connection takes place, and do not depend on other factors such as the airline, the arrival/departure times of the flights involved or the difference of such times, and the flights’ lengths.

These assumptions were chosen to simplify the models and considering the amount information available in both networks to us (e.g., the US data does not contain information about intercontinental flights). It is possible and relatively straightforward to relax them in order to improve the predictive performance of both models, and/or assess their impact on the results obtained from the simulations. We have done this in previous studies focused on a single network (Fleurquin et al., 2013, 2014b; Campanelli et al., 2014, 2015, 2015b).

### 2.5. Input data

The first input for both models is the airport capacities. In the case of Europe, we have information on the nominal capacities of most of the airports. However, we do not have the same information for the US. A pragmatic approach has been taken in both areas and the capacity of each airport will be fixed at \( \beta = 1.3 \) times the operations scheduled hourly.

The next set of input data refers to the daily programs of the airlines. These are approximated by the schedules extracted from the flight performance data. The source of data for the US is the Bureau of Transportation Statistics (BTS) of the American Government that makes available at its Web page (http://www.rita.dot.gov) daily performance data disaggregated at the level of flights for over two decades. Out of all this data, we selected July 13, 2012, as one day with an important congestion level in the network without a clear external cause such as generalized bad weather or labor conflicts. Note that this does not exclude local problems in some airports. The daily network can be seen in Fig. 1 along with the location of the airports in part of the US. This day the BTS performance data registered 17,894 domestic flights, with 8,735 showing some delay. The average delay per flight was 24.9 min, while the average delay of delayed flights was higher at 50.9 min. The total delay in the domestic US network was 445,077 min Fig. 1 also shows the (cumulative) delay distribution and the delay evolution of the largest congested cluster per hour.

The data selected for Europe comes from Eurocontrol Central Office for Delays Analysis (CODA), which is the office in charge of monitoring flight performance at the continental scale. It refers to June 20, 2013, some days before than the US data. This day a major congestion also developed throughout the European network without an external cause such as extreme bad weather or strikes. As before, this does not exclude local congestion or the presence of ATFM regulations in some airports. A map with the airports and the main statistical properties are included in Fig. 2. This day the data registers 15,565 intra-European flights, with 8,326 of them delayed. The total delay is 202,095 min, while the average delay per flight is 13 min and the average delay of the delayed flights over 24 min.
Comparing Figs. 1 and 2, one realizes that the congestion in the US was stronger, with larger congested clusters and also values of delay per flight. This variety is interesting to test the models in different contexts.

Some data cleaning and processing has been required in the two datasets before passing the daily schedules to the models. The information available refers to flights of commercial airlines including flights directed to or starting in airports outside the considered areas (ECAC area for Europe and the US domestic airports). These flights are discarded in the modeling exercise. A very small number of aircraft rotation presented also inconsistencies such as missing legs. These rotation have been disregarded since the impact in the datasets was minimal. For further details on the data cleaning process see (Fleurquin et al., 2013, 2014; Campanelli et al., 2015, 2015b).

Both models, representing different approaches to air traffic management, are run on the two datasets. The last input missing is the initial conditions for the propagation, that is, the primary delays. These are delays caused by hard to predict circumstances such as localized bad weather, technical issues in the aircraft, airport or air control. In the European data, we have information about the delay causes and it is possible to separate primary from reactionary delays. The primary delays are used as inputs. However, this did not apply to the American case where we needed a pragmatic criterion: the first delays that appear at the beginning of the aircraft’s rotations are taken as primary. This criterion was used in previous works (Fleurquin et al., 2013, 2013b; 2014, 2014b) producing good results.

3. Results

The first action to take is to calibrate $\alpha$ in the two models. Even though the objective is to compare the models outputs and not so much with the empirical data, it is helpful to run the simulations with a realistic value of the connectivity parameters. In the case of the US model, a reasonable cluster is found for a value of $\alpha = 0.29$. This value is thus fixed for the simulations with US data of both models. The SM model, on the other hand, gives a good fit to the empirical delay for $\alpha = 0.57$, which is also established for all the simulations of both models in Europe. Note that $\alpha$ is large but the probability of connection between flights is calculated multiplying $\alpha$ by the fraction of connecting passengers in each airport obtained.
from the Market Sector Data. This fraction can be small and therefore the product with $a$ may still be below the unit even for values of $a$ larger than one.

Once the values of $a$ are fixed, the two models are run on the two datasets. The results for the simulations with the US data can be seen in Fig. 3. The empirical total delay is between the predictions of the two models. The FCFS model produces more delay (especially at the end of the day) than the SM one based on slots. This result is confirmed when the congested cluster size is analyzed as a function of time during the day in Fig. 3B. Both models produce similar results in the early hours up to almost 20:00 EST. The activity and also the congestion after this time start to propagate toward the West Coast. Interestingly, it also coincides with the divergence of both models. The FCFS model produces larger clusters than the SM one even beyond the error bars. It is important to note that in this case $a$ was not fit to the cluster size and so the agreement with the empirical values is not so good for any of the models.

On the contrary, in the European data case the value of $a$ was selected in such a way that the SM model was able to reproduce the height of the peak in the large congested cluster size. The simulations output can be observed in Fig. 4, where equivalent results to those in Fig. 3 are displayed for Europe. The SM model values agree well with the empirical data both in terms not only of the height of the peak in cluster sizes but also of total delay accumulated along

![Characterization of the European network on June 20, 2013. In A, a screenshot of the network structure with the airports as nodes in red and the links (direct flights) in green. The size of the airports correlates with the number of destinations. In B, complementary distribution function of the flight delays. In C, evolution of the size of the largest congested cluster along the day. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image)
time and in the shape of the curve of the congested cluster evolution. The FCFS model with a simpler queue system in the management of the flights produces more delay and larger clusters here too. The divergence between both models starts earlier than with the American data but the conclusions are the same. Simpler queue systems lead to larger congestion.

4. Conclusions

The US and the European networks have been compared in the past literature regarding flight performance. As happens in the examples selected in this paper, congestion in average delay per delayed flight, total delay or even the size of the congested airport clusters tend to be larger in the US than in the ECAC area. Some of the reasons for these differences lie in the diverse ways of managing flights on both sides of the Atlantic. Most of the US airports implement a first-come first-served protocol when handling flights, while in Europe a system based on ATFM slots is used. Comparing both systems in the same conditions is not a practical option in reality since changing the priority method would have important associated costs. This is precisely the kind of task that can be undertaken using modeling. In silico simulations allow the change of the prioritizing method and to assess the relevance of the impact. Here we have used two models developed to reproduce reactionary delay propagation in Europe and in the US. One model uses ATFM slots and the other first-come first-served priorities. The models have been simplified and run on the very same schedule samples: one from the US and another from Europe. Our results show that flight management based on FCFS produces larger delays with the two data samples. The ATFM slot framework may look more rigid but it seems to protect slightly better the system from the development of large congestion. The price to pay is, however, a more involved flight management.

Acknowledgments

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. Bruno Campanelli is funded by the Conselleria
de Educació, Cultura i Universitats of the Government of the Balearic
Islands and the European Social Fund. Jose J. Ramasco acknowl-
dges funding from the Ramón y Cajal program of the Spanish
Ministry of Economy (MINECO).

References

AhmadBeygi, S., Cohn, A., Guan, Y., Belobaba, P., 2008. Analysis of the potential for
at Newark International Airport. Proceedings of the 4th USA/Europe Air Traffic
Management R&D Seminar. Santa Fe, USA.
Multiscale mobility networks and the large scale spreading of infectious dis-
Propagation through an Airline Schedule. Proceedings of the 2nd USA/Europe
Air Traffic Management R&D Seminar. Orlando, USA.
Sons, Chichester, UK.
gov.
Burghouwt, G., de Witt, J., 2005. Temporal configurations of European airline net-
Flights. Proceedings of the 11th USA/Europe Air Traffic Management R&D
Seminar. Lisbon, Portugal.
Scenarios in the ECAC Area. Proceedings of the Sixth SESAR Innovation Days.
Bologna, Italy, 2015.
CODA. Central Office for Delay Analysis Digests on Delay to Air Transport in Europe.
eurocontrol.int/articles/coda-publications.
Cook, A., Tanner, G., 2011. European Airline Delay Cost Reference Values, Perfor-
mance Review Unit Eurocontrol.

Note: The costs calculated in the US Congress report include a wide range of
possible impacts like missing opportunities of the passengers and image
damage for airlines. The costs in the European network calculated in (Cook and
Tanner, 2011) refer only to ATFM costs. This is the reason for the apparent
discrepancy in the figures given for the two networks.

Eurocontrol, The Federal Aviation Administration (FAA) and the EU Commission,
2013. Comparison of Air Traffic Management-related Operational Performance:
US/Europe. Available online at: https://www.faa.gov/air_traffic/publications/
Fleurquin, P., Ramasco, J.J., Eguiluz, V.M., 2013b. In: Data-driven Modeling of Sys-
temic Delay Propagation under Severe Meteorological Conditions. Proceedings of
the 10th USA/Europe Air Traffic Management R&D Seminar. Chicago, USA.
tionary Delay: Addressing the Dynamical Robustness of the US Air Trans-
portation network: Anomalous centrality, community structure and cities’
John Wiley & Sons Inc, NJ.
ICAQ, 2015. International Civil Aviation Organization, Forecast of Scheduled Pas-
Lan, S., Clarke, J., Barnhart, C., 2006. Planning for robust airline operations: opti-
mizing aircraft routings and flight departure times to minimize passenger
046106.
US Congress Joint Economic Committee, 2008. Your flight has been delayed again:
flight delays cost passengers, airlines and the U. S. economy billions. Available
online at: http://www.jec.senate.gov (accessed 22. 05. 08).
Vilaplana, P., 2010. Hub & Spoke vs Punto a punto: análisis práctico de las tipologías
de redes aéreas. Master Thesis in the School of Civil Engineering of the Uni-
versitat Politècnica de Catalunya, director. Carles Casas. Spain, Barcelona.