ARTICLE IN PRESS
Chinese Journal of Aeronautics, (2017), xxx(xx): xxx–xxx

Chinese Society of Aeronautics and Astronautics
& Beihang University
Chinese Journal of Aeronautics

cja@buaa.edu.cn
www.sciencedirect.com

Network analysis of Chinese air transport delay propagation

Massimiliano Zanin [a,b,*], **Seddik Belkoura** [a], **Zhu Yanbo** [c]

[a] *Innaxis Foundation & Research Institute, José Ortega y Gasset 20, 28006 Madrid, Spain*
[b] *Faculdade de Ciências e Tecnologia, Departamento de Engenharia Electrotécnica, Universidade Nova de Lisboa, Lisboa, Portugal*
[c] *Aviation Data Communication Corporation, 100083, Beijing, China*

Received 23 May 2016; revised 22 September 2016; accepted 14 December 2016

KEYWORDS

Air transport;
Complex networks;
Delays propagation;
Functional networks;
Granger causality

Abstract The Chinese air transport system has witnessed an important evolution in the last decade, with a strong increase in the number of flights operated and a consequent reduction of their punctuality. In this contribution, we propose modelling the process of delay propagation by using complex networks, in which nodes are associated to airports, and links between pairs of them are assigned when a delay propagation is detected. Delay time series are analysed through the well-known Granger Causality, which allows detecting if one time series is causing the dynamics observed in a second one. Results indicate that delays are mostly propagated from small and regional airports, and through flights operated by turbo-prop aircraft. These insights can be used to design strategies for delay propagation dampening, as for instance by including small airports into the system's Collaborative Decision Making.

1. Introduction

The Chinese air transportation system, one of the most important and fastest-growing in the world, is receiving an increasing attention from the research community due to its distinctive characteristics and challenges. On one hand, during the last decade, China has become the second largest market, with the Beijing Capital International Airport being the second busiest airport in the world. On the other hand, China is suffering from a limited availability of airspace resources, among others due to a strong military presence.[1] As a result, nowadays passengers are experiencing significant delays, with an average of 40% of flights being delayed every day. According to the flight statistics of Civil Aviation Administration of China (CAAC), the appearance of delays can be attributed to one of the four main categories: airlines related (≈40%), air traffic control (≈20%), weather related (≈20%), and military uses (≈10%). These four categories are responsible for more than 90% of the flight delays in China.

In this work, we propose the use of complex network theory to understand how delays appear and propagate through the Chinese airport network. Complex networks are powerful tools to help understanding the structures and dynamics of

* Corresponding author at: Innaxis Foundation & Research Institute, José Ortega y Gasset 20, 28006 Madrid, Spain.
E-mail address: mzanin@innaxis.org (M. Zanin).

Peer review under responsibility of Editorial Committee of CJA.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.cja.2017.01.012
1000-9361 © 2017 Chinese Society of Aeronautics and Astronautics. Production and hosting by Elsevier Ltd.

complex systems in general,[2,3] and of air transportation in particular. Airport networks and air route networks have been the focus of many studies around the world - see, for instance, Refs. [4,5] for some examples. Specifically to the Chinese system, examples include Refs. [6–10].

The structure induced by delay propagation between airports can be seen as a complex network, which lies on top of the network created by flights and which is the result of the dynamics of the system. This is similar to what has been considered in neuroscience: flight networks are equivalent to the way different brain regions are physically connected (i.e. the connectome), while a delay network represents how information (here, information about delays) is transmitted and processed at each node.[11] Once such network has been reconstructed, it can be used to obtain relevant information about the propagation process: for instance, which are the most important airports, both for delay generation and propagation; or how different airlines interact to produce the observed dynamics.

In the line of what has already been studied for the European network,[12,13] we here propose to reconstruct delay propagation networks for the Chinese air transportation system by applying the well-known Granger Causality metric[14,15] on airport delay time series. A link between two airports will thus represent a direct delay propagation, i.e. a situation in which the delays observed at one airport are partly responsible for what we observe in a second airport. By characterising the resulting network using several established metrics, we make a first step towards the understanding of the dynamics of delay propagation in the Chinese air transport network, and towards the development of adequate strategies for improving the passengers' experience. Note that the approach here proposed is fundamentally different from standard delay studies. On the one hand, we here focus on the global structure created by delay propagation, and not on the local dynamics of individual flights and airports;[16,17] additionally, results are obtained by observing the real dynamics of the system, and not by inferring it from large-scale synthetic models.[18,19]

Beyond this introduction, this work is organised as follows. Section 2 presents the main techniques here applied, including a description of the data set (Section 2.1), the creation and preprocessing of delay time series (Section 2.2), the reconstruction of propagation networks through the Granger Causality (Section 2.3), and the analysis of the resulting networks (Section 2.4). Section 3 reports the results obtained, by considering three different networks: the global delay propagation one, the networks created by the three most important airlines, and the one created by aircraft types. Finally, Section 4 summarises our major conclusions and discusses future lines of research.

2. Methods

2.1. Data set description

The data set has been provided by Aviation Data Communication Corporation (ADCC) of CAAC, and comprises high-level information for all flights crossing the Chinese airspace during the 2015 summer peak (1st to 31st August 2015). For each flight, available information includes, among others, the scheduled departure and arrival airports, scheduled and real departure and arrival time, as well as the airline code and aircraft type. Only domestic flights, i.e. departing from and arriving at a Chinese airport, have here been considered; additionally, flights involving non-ICAO airports, marked as ZZZZ, have further been discarded. Cancelled flights were also not included in the data set, which only covers executed operations. This yielded a final data set comprising 152 airports and 45.151 flights - see Fig. 1 for a graphical representation.

The delay of each flight has been calculated as the difference between the real and scheduled arrival time, and therefore represents the real delay as perceived by passengers. It has to be noted that this is different from the official announced delay, as the latter is not reported if the flight arrived within 10/15 min from the expected arrival time.

2.2. Time series pre-processing

Once the landing delay of each flight has been assessed, it is necessary to reconstruct the average dynamics of each airport, i.e. the average landing delay within one-hour time windows. An example of the resulting time series is represented in Fig. 2(a), for the Beijing Capital International Airport. It can be appreciated that the time series is characterised by peaks appearing with a daily frequency - note the peaks at 24 h and multiples in the cross-correlation depicted in Fig. 2(b). This is mainly due to the non-stationarity of delays, whose expected value is higher at the end of the afternoon, and lower during the first hours of the day - see Fig. 2(c).

In order to reduce the non-stationarity of the time series, which may result in a biased evaluation of the Granger Causality metric, we here apply a Z-Score detrend procedure. In detail, the detrended delay time series for one airport is calculated as:

$$D'(d,h) = \frac{D(d,h) - \langle D(\cdot,h)\rangle}{\sigma(D(\cdot,h))} \tag{1}$$

$D'(d,h)$ being the detrended delay for day d and time h, $D(d,h)$ the original average delay, and $\langle D(\cdot,h)\rangle$ and $\sigma(D(\cdot,h))$ the average and standard deviation of the delays recorded at time h for all available days. A graphical representation of this transformation is depicted in Fig. 3.

The resulting time series have some interesting characteristics. First of all, their average value is zero by construction - note that we subtract the average delay $\langle D(\cdot,h)\rangle$ from each element. Second, positive and negative delays represent how unusual (in terms of the number of standard deviations) is the observed delay with respect to the delay expected at the same hour. Also note that, while negative delay values are usually discarded in delay propagation studies, here they encode important information, as they represent situations in which the behaviour of the system is better than expected; the objective thus includes understanding if such positive situations can be propagated between different airports. Finally, and as a direct consequence of the previous point, the resulting time series are detrended: if, at a given time, a high delay is observed, but this is aligned with what expected in that time window, D' will be close to zero.

2.3. Granger causality

The Granger Causality (GC) test[14,15] has extensively been used to assess the information exchanged between different elements of a system, and more specifically, to test for causality relationships that one might intuitively assume. GC is an extre-