

## Research Article

# Data-Driven Modeling of Systemic Air Traffic Delay Propagation: An Epidemic Model Approach

Shanmei Li <sup>1</sup>, Dongfan Xie <sup>2</sup>, Xie Zhang,<sup>1</sup> Zhaoyue Zhang,<sup>3</sup> and Wei Bai<sup>4</sup>

<sup>1</sup>College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China

<sup>2</sup>School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

<sup>3</sup>School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China

<sup>4</sup>Air Traffic Control Department, North China Air Traffic Management Bureau, Beijing 100621, China

Correspondence should be addressed to Shanmei Li; amy820203@163.com

Received 10 April 2020; Revised 12 June 2020; Accepted 13 July 2020; Published 3 August 2020

Academic Editor: Chunjiao Dong

Copyright © 2020 Shanmei Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To better understand the mechanism of air traffic delay propagation at the system level, an efficient modeling approach based on the epidemic model for delay propagation in airport networks is developed. The normal release rate (NRR) and average flight delay (AFD) are considered to measure airport delay. Through fluctuation analysis of the average flight delay based on complex network theory, we find that the long-term dynamic of airport delay is dominated by the propagation factor (PF), which reveals that the long-term dynamic of airport delay should be studied from the perspective of propagation. An integrated airport-based Susceptible-Infected-Recovered-Susceptible (ASIRS) epidemic model for air traffic delay propagation is developed from the network-level perspective, to create a simulator for reproducing the delay propagation in airport networks. The evolution of airport delay propagation is obtained by analyzing the phase trajectory of the model. The simulator is run using the empirical data of China. The simulation results show that the model can reproduce the evolution of the delay propagation in the long term and its accuracy for predicting the number of delayed airports in the short term is much higher than the probabilistic prediction method. The model can thus help managers as a tool to effectively predict the temporal and spatial evolution of air traffic delay.

## 1. Introduction

Flight delays are one of the most important performance indicators of air transportation system. It has become increasingly more serious, which directly causes huge damage to the quality of civil aviation services, such as declines in operational safety, increases in operating costs, and more serious environmental pollution. Notably, commercial aviation players understand delays as the difference between scheduled and real times of departing or arriving flights [1]. According to the Federal Aviation Administration (FAA), a flight can be considered as delayed if the operation takes place 15 minutes after schedule [2].

The delay of an individual flight seems to be random at a glance. A flight delay may be transferred and amplified by consequent operations. Some delays that originate from upstream flights spread to downstream flights, which is

particularly evident when an aircraft flies multiple flight legs. This phenomenon is defined as delay propagation (DP) [3–6]. Actually, DP causes delays to obey certain statistical laws [7] when long-term delay records for a large number of flights are taken into account, which form airport delay propagation. Consequently, a congested airport may propagate delays to connecting airports through the delayed flights, which eventually has an impact on the performance of a significant part of the entire network [8]. The delay can be magnified when it is examined in multiairport networks [9].

There are many factors affecting airport delays. The factors can be divided into two categories: propagation factors (PF) (airport delays caused by connected delayed airports) and nonpropagation factors (NPF) (airport delays caused by original factors, such as extreme weather and equipment trouble) [10]. As the number of flights increases,

increasingly more airport delays are caused by the PF [11]. If there are many delayed flights in one airport, the connected airports may become delayed, which can affect further operations in a cascade-like effect. Due to the complexity of air transportation, the mechanisms of delay propagation at the airport level are not fully understood. Therefore, research on the mechanism of delay propagation is timely yet challenging.

Delay spreading has received lots of attention from the Air Traffic Management (ATM) community during the last decade. Some studies [12–14] have established flight delay propagation models based on Bayesian Networks and analyzed the internal factors influencing air traffic delay propagation. Pablo Fleurquin [15] introduced an agent-based model that reproduces the delay propagation patterns observed in U.S. performance data and identified passenger and crew connectivity as the most relevant internal factor contributing to delay spreading. Qiu [16] constructed a joint distribution of continuous flight delays by using the 2D copula function. Wong and Tsai [17] established a cox proportional hazard model for flight delay propagation based on survival analysis theory. Kalfe and Zou [3] promoted a delay propagation pattern based on an econometric method and analyzed factors by using the Heckman two-step method.

To the best of our knowledge, gaps still remain in understandings of the delay propagation in airport networks. The process of delay propagation needs to be analyzed from a broader and network-based perspective because flight scheduling for airlines and airport operations is increasingly synchronized from the perspective of network operation. The linkage in the airport networks is the direct operation by the airlines linking the airports [18]. It is an essential feature of the network structure of the air traffic system. Therefore, the propagation dynamics cannot be understood without referring to the underlying complex network structure. Some scholars have used complex network theory to characterize transportation [19–22]. The initial studies [23, 24] have identified a high heterogeneity in the traffic sustained by each edge. Most of the airport network exhibits a heavy-tailed degree distribution, which is often well approximated for a significant range of values of degree  $k$  by a power-law behavior ( $P(k) \sim k^{-\gamma}$ ), from which the name “scale-free network” originated [25–29]. Further studies have used the complex network characteristics to explain the propagation of air traffic congestion and flight delay [29–34]. From a macroperspective, the airport delay is usually caused by air traffic congestion. Thus, the application of the complex network theory to the air traffic problems is feasible. And most of them focused on the delay propagation between sequence flights from the perspective of single airport operation.

As aforementioned, due to the large number of airports and complex interactions, the features of delay propagation cannot be understood based on the information of an individual airport. Complex network theory and its associated metrics and tools present an opposite approach to study the air transport system beyond what is offered by classical techniques. To further understand the effects of delay

propagation, most of the existing achievements focused on the delay propagation between sequence flights from the perspective of single airport operation. In this paper, we propose a network-based approach to modeling airport delay propagation. There are several classical epidemic models in complex network theory, such as Susceptible-Infected (SI) model, Susceptible-Infected-Susceptible model (SIS), Susceptible-Infected-Recovered (SIR) model, and Susceptible-Infected-Recovered-Susceptible (SIRS) model. Because the propagation mechanisms of SIRS is the most similar to that of airport delay propagation (the details can be seen in Section 4.1), the SIRS model is utilized to understand the process of air traffic delay propagation in the context of an airport network and explain the spreading characteristics between different airports in this paper. The SIRS model has been normally used to simulate the process of how diseases [35], safety risk [36], or computer virus [37] spread.

To the best of the authors’ knowledge, this study is the first to apply the SIRS model to air traffic delay propagation. Firstly, the metrics for measuring airport delays are introduced. Then, the fluctuation of airport delay is studied from different time scales based on complex network theory in order to find out the propagation factors, PF or NPF. Then, an integrated airport-based SIRS (ASIRS) model is developed. At last, the effective and accuracy of this model is demonstrated using empirical data of China.

The outline of the remainder of this paper is as follows. The data sources and a measurement of airport delay are provided in Section 2. Section 3 is devoted to fluctuations of the average flight delay and the determination of airport delay status. The ASIRS model is established in Section 4. The data-driven description of the ASIRS model is presented in Section 5. Finally, Section 6 contains the conclusions.

## 2. Data Description and Airport Delay Measurement

The dataset analyzed in this paper is provided by the Civil Aviation Administration of China (CAAC) and consists of all flight information of China from June to December in 2015. There are totally 93630 records of flights. The number of carriers in the data is 295. The information from the historical flight data consists of flight ID, date of flight, real and scheduled departure (arrival) times (Beijing Time), origin, and destination. A short sample of the original database is shown in Table 1.

There are 205 domestic airports in the database. A sample of airports are shown in Table 2.

The identification of the airport delay state is the first problem to be solved. As we know, the delay state of an airport is the concentrated performance of individual flight delay. Thus, the airport delay is measured by the delays of arriving and departing flights.

In our study, the normal release rate (NRR) and average flight delay (AFD) are considered to measure airport delay:

- (1) The NRR of an airport is the ratio of the number of normal released flights to the total number of departure flights. According to the Federal Aviation

TABLE 1: A sample of the original database.

Flight ID	Scheduled departure airport	Scheduled departure time	Scheduled arrival airport	Scheduled arrival time	Real departure airport	Real departure time	Real arrival airport	Real arrival time
CES2265	ZLXY	201510071525	ZGHA	201510071700	ZLXY	201510071606	ZGHA	201510071737
CES2266	ZGHA	201510071740	ZLXY	201510071920	ZGHA	201510071848	ZLXY	201510072030
CES2266	ZLXY	201510072045	ZLIC	201510072150	ZLXY	201510072202	ZLIC	201510072256
HXA2689	ZUCK	201510071415	ZLLL	201510071555	ZUCK	201510071421	ZLLL	201510071549
GCR7567	ZGOW	201510071530	ZJHK	201510071650	ZGOW	201510071533	ZJHK	201510071648
JYH1032	ZBMZ	201510071625	ZYCC	201510071755	ZBMZ	201510071621	ZYCC	201510071750
CHH7702	ZGSZ	201510072140	ZBAA	201510080040	ZGSZ	201510072225	ZBAA	201510080106
UEA2755	ZSLG	201510071105	ZYTX	201510071250	ZSLG	201510071109	ZYTX	201510071241
CSN6633	ZHCC	201510071035	ZLXN	201510071225	ZHCC	201510071051	ZLXN	201510071242

TABLE 2: A sample of airports.

	1	2	3	4	5	6	7	8
Airport code	ZBAA	ZBMZ	ZGGG	ZGHA	ZGOW	ZGSZ	ZHCC	ZJHK
City	Beijing	Inner Mongolia	Guangzhou	Changsha	Shantou	Shenzhen	Zhengzhou	Haikou
Airport code	ZPPP	ZSHC	ZSLG	ZSPD	ZSSS	ZUCK	ZUUU	ZLXY
City	Kunming	Hangzhou	Lianyungang	Shanghai	Shanghai	Chongqing	Chengdu	Xi'an

Administration (FAA), a flight is considered as abnormal if the departure operation takes place more than 15 minutes after schedule:

$$\text{NRR} = \frac{N - n}{N}, \quad (1)$$

where  $N$  represents the number of departure flights and  $n$  represents the number of abnormal flights.

- (2) The AFD of an airport is the ratio of the total delay time to the total number of all the departure and arrival flights of the airport:

$$\text{AFD} = \frac{\sum_{i=1}^{N+M} d_i}{N + M}, \quad (2)$$

where  $M$  represents the number of arrival flights and  $d_i$  represents the delay of flight  $i$ .

The NRR and AFD are counted every 1 hour using the above database. Partial statistical results are shown in Tables 3 and 4.

### 3. Fluctuations of Airport Delay and Determination of Airport Delay Status

*3.1. Fluctuations of Airport Delay.* To explore the propagation laws of airport delay, the delay fluctuations should be studied from the perspective of airport networks. Fluctuations can be considered by investigating the coupling between the average flux and the fluctuations, which is actually the mean and standard deviation analysis, as developed in [38–42]. It is found that the standard deviation  $\sigma_i$  and the

average flux  $\langle f_i \rangle$  on individual nodes obey a unique scaling law as

$$\sigma_i \propto \langle f_i \rangle^\alpha,$$

$$\langle f_i \rangle = \sum_{t=1}^T f_i^{\Delta t}(t), \quad (3)$$

$$\sigma_i = \sqrt{\sum_{t=1}^T (f_i^{\Delta t}(t) - \langle f_i \rangle)^2},$$

where  $f_i^{\Delta t}(t)$  denotes the flux of node  $i$  in time interval  $[(t - \Delta t), t]$ .

As the strength of the external driving force increases, the value of  $\alpha$  gradually increases.

A method to separate the internal dynamics from the external fluctuations of complex systems is also promoted in [38–42].

The dynamical variable  $f_i(t)$  can be separated into two components:

$$f_i(t) = f_i^{\text{int}}(t) + f_i^{\text{ext}}(t), \quad (4)$$

where  $f_i^{\text{ext}}(t)$  is generated by external factors and  $f_i^{\text{int}}(t)$  is generated by internal factors. They can be described as follows:

$$f_i^{\text{ext}}(t) = A_i \sum_{i=1}^N f_i(t),$$

$$f_i^{\text{int}}(t) = f_i(t) - A_i \sum_{i=1}^N f_i(t), \quad (5)$$

$$A_i = \frac{\sum_{t=1}^T f_i(t)}{\sum_{i=1}^N \sum_{t=1}^T f_i(t)}.$$

TABLE 3: Samples of average flight delay (minute).

	0:00–1:00	1:00–2:00	2:00–3:00	3:00–4:00	4:00–5:00	5:00–6:00	6:00–7:00	7:00–8:00
ZGSZ	152.5	104	113	61.78	95	43.66	13.84	24.2187
ZJHK	0	0	0	0	0	0	28.37	14.5714
ZPPP	86.5	22	22	0	0	0	9.5	22.1395
ZBAA	40.36	31.27	30.75	27	52	89.91	31.5	30.7857
ZLXY	39	162	68	0	0	0	0	14.76

TABLE 4: Samples of airport normal release rate (%).

	0:00–1:00	1:00–2:00	2:00–3:00	3:00–4:00	4:00–5:00	5:00–6:00	6:00–7:00	7:00–8:00
ZGSZ	81.81	87.5	75	40	80	25	24	5.88
ZJHK	100	0	0	0	0	0	27.27	6.66
ZPPP	33.33	50	0	0	0	0	33.33	10.41
ZBAA	6.66	21.42	7.69	20	50	16.66	28.94	5.40
ZLXY	40	0	50	0	100	0	0	24.24

Furthermore, whether or not the fluctuations are mainly internally or externally imposed can be determined:

$$\eta_i = \frac{\sigma_i^{\text{ext}}}{\sigma_i^{\text{int}}}, \quad (6)$$

$$\sigma_i^{\text{ext}} = \sqrt{\langle f_i^{\text{ext}}(t)^2 \rangle - \langle f_i^{\text{ext}}(t) \rangle^2}, \quad (7)$$

$$\sigma_i^{\text{int}} = \sqrt{\langle f_i^{\text{int}}(t)^2 \rangle - \langle f_i^{\text{int}}(t) \rangle^2}. \quad (8)$$

If  $\eta_i \gg 1$ , the system dynamics are dominated by the network-wide factors, while for  $\eta_i \ll 1$ , local dynamics overshadow the network-imposed changes.

We aggregate the data and carry out the scaling law analysis at different time scales. In our study,  $\langle f_i^{\Delta t} \rangle$  represents the AFD in the time scale  $\Delta t$  of airport  $i$  and  $\sigma_i^{\Delta t}$  represents the standard deviation.

Figure 1 shows the relationship between  $\langle f_i^{\Delta t} \rangle$  and  $\sigma_i^{\Delta t}$ , with time scale  $\Delta t = 1$  h, 3 h, and 6 h. The scaling law between  $\sigma$  and  $\langle f \rangle$  can be clearly observed.

It can be seen that the value of the scaling exponent  $\alpha$  increases as  $\Delta t$  increases, suggesting that the system may have an inhomogeneous influence, as pointed out by Eisler and Kertesz [43]. The reason for this result is that the fluctuations of AFD are due mainly to network-wide factors such as the PF when  $\Delta t$  is much bigger; on the contrary, the fluctuations of AFD are due mainly to local factors such as the NPF when  $\Delta t$  is much smaller. When  $\Delta t$  is much bigger, airport delays may be caused by connected delayed airports. Some delays that originate from upstream flights spread to downstream flights, which is particularly evident when an aircraft flies multiple flight legs. When  $\Delta t$  is much smaller, airport delays may be caused by original factors, such as extreme weather and equipment trouble. Additionally, the ratio  $\eta_i$  for the 1 hour interval is calculated using the above method. The result reveals that the average  $\eta_i$  is 5.665353, which shows that the dynamic of airport delay of every 1 hour is dominated by PF. And the larger the time scale is, the bigger the value of  $\eta_i$  is. When the time scale is 3 hours, the value of  $\eta_i$  is 6.431234. When the time scale is 6 hours, the

value of  $\eta_i$  is 8.534778. Thus, we have to study how does the delay originating from an airport propagate to other airports in a large time scale.

**3.2. Determination of Airport Delay Status.** NRR and AFD are used to determine if the airport is in the delay state. The specific criteria is as follows:

If  $\{AFD_i(t) > \xi_1 \& NRR_i(t) < \xi_2\}$ , airport  $i$  is in the delay state in the time interval  $t$ .

To explore the characteristics of airport delay propagation, the value of the time interval  $t$  is 1 hour, as the dynamic of airport delay of every 1 hour is dominated by PF. When the delay propagates in the airport network, the airport delay is usually severe. The airport delay deduced by NPF is much small in the time scale of one hour. Thus, in order to eliminate the influence of NPF, the values of threshold  $\xi_1$  should be relatively large and the values of threshold  $\xi_2$  should be relatively small. Here,  $\xi_1 = 15$  min and  $\xi_2 = 0.7$ .

## 4. Epidemic Model of Airport Delay Propagation

**4.1. Airport-Based Susceptible-Infected-Recovered-Susceptible Model.** From the discrimination of delay propagation, we find that the epidemic model in a complex network is a valuable research tool for the exploration of fundamental laws and trends of delay propagation in airport networks. There are three kinds of individuals in the SIRS model: susceptible ones ( $S$ ), infected ones ( $I$ ), and recovered ones ( $R$ ). The susceptible ones are currently in a healthy state, and when they contact the source of the infection, they will become infected ones with an infection rate  $\alpha$ . The infected ones are unhealthy ones, and they can infect susceptible ones. The infected ones will be cured with a cure rate  $\beta$  and become recovered ones, and the recovered ones become healthy ones with an immunity ability. The immunity will disappear under some certain situations, and the recovered ones will become susceptible ones with an immunity-loss rate  $\gamma$ . The infectious mechanism is described in Figure 2.

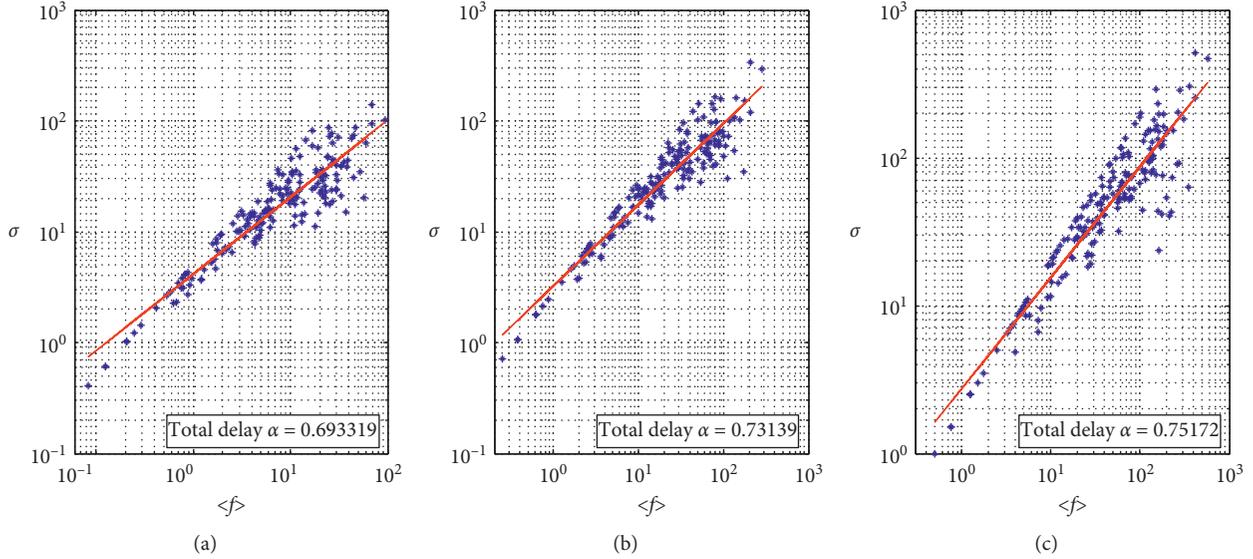


FIGURE 1: Scaling relations between fluctuations  $\sigma$  and AFD  $\langle f \rangle$  at different time scales. (a)  $\Delta t = 1$  h. (b)  $\Delta t = 3$  h. (c)  $\Delta t = 6$  h. The estimated values of scaling exponent  $\alpha$  are also indicated.

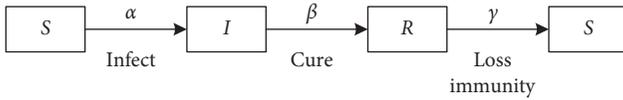


FIGURE 2: Infectious mechanism of SIRS.

In an airport network, the original airport delay may be due to capacity reduction, airport equipment trouble, and extreme weather. In the process of delay propagation for resource-shared flights, delays are propagated from an upstream flight at the departure airport to the arrival airport. As shown in Figure 3, the airports with “delay root” represent the susceptible ones, the airports with “delay propagation” represent the infected ones, and the airports with “delay termination” represent the recovered ones. The propagation of airport delay has traditionally been described as graphs with vertices representing airports and edges representing connectivity. When the delay is serious in one airport, the delay of its connected airports may be increased due to the delay spreading. Furthermore, the delay of spread airports may be absorbed in the subsequent operations, and they would not be influenced again by the same initial airport delay. However, they may be affected by another original airport delay later. Because of the complexity of airport networks [32], the evolution of delay within them possesses the characteristics of propagation in complex networks.

From the above analysis, we find that the infectious mechanism of SIRS is similar to the propagation characteristics of airport delay discussed before. Suppose there are three kinds of airports in the network at time  $t$ : non-delayed airports ( $S$ ) which are easily infected, delayed airports ( $I$ ), recovered airports ( $R$ ) which used to be delayed but are back to normal. The recovered airports only have immunity to the current delay spread and may become susceptible ones later.

As the stochastic process is applied to all flights, the airports are affected by probability.

The dynamics of ASIRS model can be written as

$$\begin{cases} \frac{dS(t)}{dt} = -\alpha S(t)I(t) + \gamma R(t), \\ \frac{dI(t)}{dt} = \alpha S(t)I(t) - \beta I(t), \\ \frac{dR(t)}{dt} = \beta I(t) - \gamma R(t), \\ S(t) + I(t) + R(t) = 1, \end{cases} \quad (9)$$

where  $S(t)$ ,  $I(t)$ , and  $R(t)$  represent the fraction of susceptible airports, infected airports, and recovered airports, respectively, at time  $t$ ;  $\alpha$  is the infection rate;  $\beta$  is the cure rate; and  $\gamma$  is the immunity-loss rate.

Assume that the proportion of infected airports, susceptible airports, and recovered airports at the initial moment  $t_0$  is  $I_0$  ( $I_0 > 0$ ),  $S_0$  ( $S_0 > 0$ ), and  $R_0$  ( $R_0 = 0$ ), respectively:

$$\begin{cases} \frac{dI(t)}{dt} = \alpha S(t)I(t) - \beta I(t), & I(0) = I_0, \\ \frac{dS(t)}{dt} = -\alpha S(t)I(t), & S(0) = S_0. \end{cases} \quad (10)$$

The phase trajectory of the ASIRS model is analyzed. The  $S$ - $I$  plane is called the phase plane, and the domain of the phase trajectory is  $(S, I) \in D$ :

$$D = \{(S, I) \mid S(t) \geq 0, I(t) \geq 0, I(t) + R(t) \leq 1\}. \quad (11)$$

Let  $\delta = \beta/\alpha$ . The following equation can be obtained:

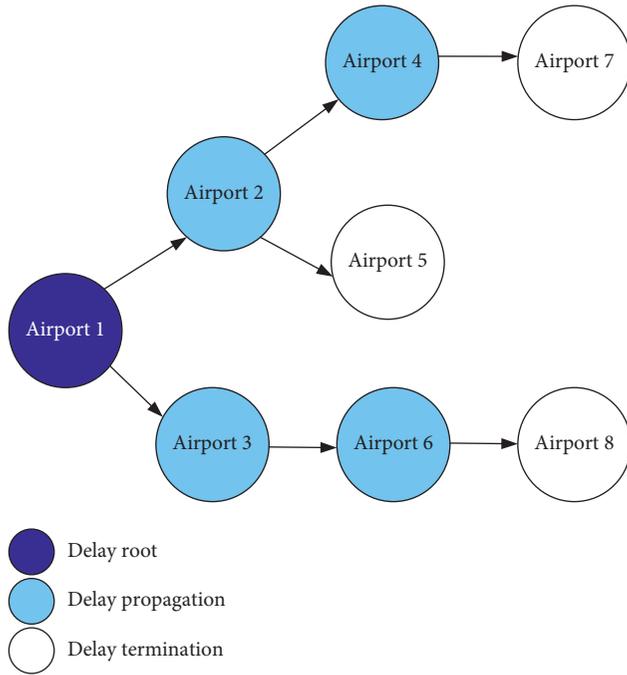


FIGURE 3: Delay propagation tree.

$$I(t) = I_0 + S_0 + \delta \ln \frac{S(t)}{S_0} - S(t). \quad (12)$$

The phase trajectory diagram is shown in Figure 4.

When  $t \rightarrow \infty$ , the limit values of  $I(t)$ ,  $S(t)$ , and  $R(t)$  are  $I_\infty$ ,  $S_\infty$ , and  $R_\infty$ , respectively:

- (1) No matter how the initial values of  $S_0$  and  $I_0$  change, the airport delay situation will eventually disappear,  $I_\infty = 0$ .
- (2) In equation (10), let  $I(t) = 0$ . The value of  $S_\infty$  can be calculated, which is the root value of equation (11) in the range of  $(0, \delta)$ .  $S_\infty$  is the abscissa of the intersection point between the phase trajectory and transverse axis in the range of  $(0, \delta)$ :

$$I_0 + S_0 + \delta \ln \frac{S(t)}{S_0} - S_\infty = 0. \quad (13)$$

- (3) If  $S_0 < \delta$ ,  $I(t)$  increases first; if  $S(t) = \delta$ ,  $I(t)$  reaches its maximum and then decreases to zero. At the same time,  $S(t)$  is monotonically reduced to  $S_\infty$ :

$$I_m = I_0 + S_0 - \delta \left( \ln \frac{1}{\delta} S_0 + 1 \right), \quad (14)$$

where  $I_m$  is the maximum of  $I$ .

According to the above analysis, the following conclusions can be drawn:

- (1) If  $\alpha S(t) < \beta$ ,  $I(t)$  increases and the airport delay will spread to more airports

- (2) If  $\alpha S(t) > \beta$ ,  $I(t)$  decreases, the delay situation of the airport network will be alleviated, and the airport delay will not spread to others
- (3) If  $\alpha S(t) = \beta$ ,  $I(t)$  reaches the maximum and the delay situation of the airport network is the most serious

**4.2. Parameter Analysis.** As mentioned above, the ASIRS model contains three parameters:  $\alpha$ ,  $\beta$ , and  $\gamma$ . To investigate the change of  $I(t)$ ,  $S(t)$ , and  $R(t)$  over time, suppose  $\alpha = 1$ ,  $\beta = 0.1$ , and  $\gamma = 0.1$ . The change of  $I(t)$ ,  $S(t)$ , and  $R(t)$  is shown in Figure 5.

**4.2.1. Analysis of Parameter  $\alpha$ .** First, the influence of  $\alpha$  on airport delay propagation is discussed. Figure 6 shows the changes of  $I(t)$ ,  $S(t)$ , and  $R(t)$  under different values of  $\alpha$  with the assumptions that  $\beta = 0.1$  and  $\gamma = 0.1$ .

As seen, the higher  $\alpha$  is, the earlier  $I(t)$  and  $R(t)$  reach their peak value and the higher the peak value is. The higher  $\alpha$  is, the earlier  $S(t)$  reaches its minimum and the lower the minimum is. Thus, it can be concluded that, as the speed at which airport delay propagates increases, more airports will be infected more quickly, and more airports will be recovered.

**4.2.2. Analysis of Parameter  $\beta$ .** Figure 7 shows the changes of  $I(t)$ ,  $S(t)$ , and  $R(t)$  under different values of  $\beta$  with the assumption that  $\alpha = 1$  and  $\gamma = 0.1$ .

When the value of  $\beta$  becomes smaller, the peak of  $I(t)$  appears later and the trough of  $S(t)$  appears later also. The trends of  $R(t)$  with  $\beta = 0.5/1$  and  $\beta = 0.1/0.2$  are not the same. Thus,  $R(t)$  is much more complex than  $I(t)$  and  $S(t)$ .

**4.2.3. Analysis of Parameter  $\gamma$ .** Figure 8 shows the changes of  $I(t)$ ,  $S(t)$ , and  $R(t)$  under different values of  $\gamma$  with the assumption that  $\alpha = 1$  and  $\beta = 0.1$ .

The higher  $\gamma$  is, the earlier  $I(t)$  reaches its peak value and the higher the peak value is. The lower  $\gamma$  is, the later  $S(t)$  reaches its minimum, the lower the minimum is, and the higher  $R(t)$  is. Parameter  $\gamma$  has little effect on  $I(t)$  and  $S(t)$  in the early stage and mainly affects the later stage.

## 5. Case Study

**5.1. Statistical Calculation of  $I(t)$ ,  $S(t)$ , and  $R(t)$ .** According to Section 3.2, we calculate  $I(t)$ ,  $S(t)$ , and  $R(t)$  based on the following criteria:

- (1) If  $\{NRR_i(t) > 15 \text{ min} \ \& \ ADF_i(t) < 0.7\}$ , airport  $i$  is infectious at time  $t$
- (2) If  $\{NRR_i(t) \leq 15 \text{ min} \ | \ ADF_i(t) \geq 0.7\} \ \& \ \{NRR_i(t-1) > 15 \text{ min} \ \& \ ADF_i(t) < 0.7\}$ , the airport is recovered at time  $t$  and infectious at time  $t-1$
- (3) Apart from the above situations, the airport is susceptible and can be easily infected by infectious airport

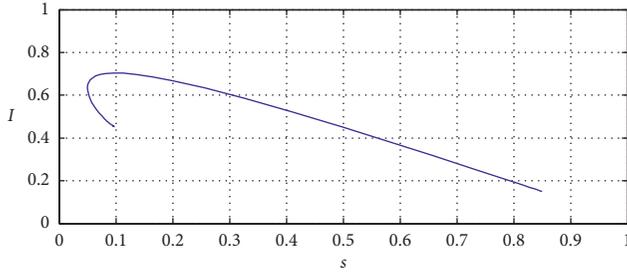
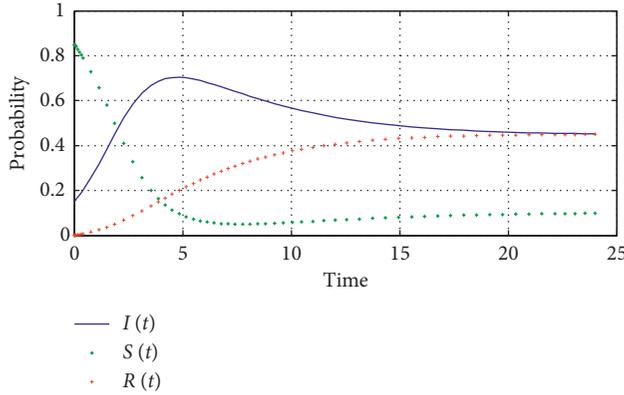


FIGURE 4: Phase trajectory diagram.

FIGURE 5: Change of  $I(t)$ ,  $S(t)$ , and  $R(t)$ .

**5.2. Determination of ASIRS Parameters.** The hourly states of all the airports in the network can be identified based on the above criteria using Tables 1 and 2. Taking the airport ZBNY in October 1, 2015, for example, Figure 9(a) shows the time varying state of airport ZBNY. It can be seen that the airport ZBNY is infected in  $\{6:00-10:00, 14:00-20:00\}$ , susceptible in  $\{0:00-6:00, 11:00-14:00, 21:00-24:00\}$ , and recovered in  $\{10:00-11:00, 20:00-21:00\}$ . We also investigate the air traffic flow of ZBNY, which is shown in Figure 9(b). Comparing Figures 9(a) and 9(b), we find that the larger the traffic flow is, the more the airport tends to be infected. It should be noted that, although the airport has a small number of flights, the flight delay is serious, which may be deduced by the delay propagation in the airport network.

In addition, we find that there are 7 infectious airports in 4:00–5:00: ZBAA, ZGSZ, ZUCK, ZGGG, ZSHC, ZSPD, and ZHCC. Six of these are the top 10 airports by throughput in China (the throughput of airports is provided by the CAAC), as shown in Table 5. There are many flights in the six airports. Large departure delays may influence the operations of the arrival airports, and the flights will arrive at or depart this airport with big delays. Thus, the phenomenon of delay propagation appears.

The simulation model is ASIRS established in Section 4.1, as shown in equation (7). The simulation method is that we use the real information of flights in China to calculate the values of  $I(t)$ ,  $S(t)$ , and  $R(t)$  and then calibrate the parameters of  $\alpha$ ,  $\beta$ , and  $\gamma$ . Thus, the ASIRS model for simulating the real flight data can be obtained.

Next, we calculate the values of  $I(t)$ ,  $S(t)$ , and  $R(t)$ . Partial statistical results are shown in Table 6.

The parameters in the ASIRS model are calibrated based on the statistical values of  $I(t)$ ,  $S(t)$ , and  $R(t)$  by using the numerical simulation method. For every day's traffic situation, we can always build an excellent ASIRS model. Taking October 1, 2015 (there are 16804 scheduled flights connecting 199 different commercial airports), for example, we find that the model fits the actual operation situation to the highest level when  $\alpha = 0.47$ ,  $\beta = 0.175$ , and  $\gamma = 0.45$ .

**5.3. Accuracy of the Long-Term Simulation of Delay Propagation.** The values of  $I(t)$ ,  $S(t)$ , and  $R(t)$  for October 1, 2015, in Figure 10 are the actual values, which are counted based on the database. The values of  $I(t)$ ,  $S(t)$ , and  $R(t)$  in Figure 11 are the simulations of the ASIRS model for that day.

Comparing Figure 10 with Figure 11, it can be seen that the long-term trends of the actual and predicted values of  $I(t)$ ,  $S(t)$ , and  $R(t)$  are similar, although there are some differences between local values. Thus, from a qualitative point of view, we can conclude that the ASIRS model is reasonable and can describe the real situation of airport delay propagation to a certain extent.

To further examine the application of the ASIRS model, we also study the characteristics of airport delay propagation derived from the degree distribution.

First, the ASIRS models for the peak period of 12:00–19:00, October 1, 2015, and the nonpeak period of 00:00–07:00, October 1, 2015, are constructed. The parameters' values of the two models are determined based on trial calculation using the actual data:

$$(1) \quad 12:00-19:00: \alpha = 0.4, \beta = 0.25, \gamma = 0.8$$

$$(2) \quad 00:00-07:00: \alpha = 0.38, \beta = 0.28, \gamma = 0.6$$

The  $\alpha$  and  $\gamma$  values of 12:00–19:00 are bigger than those of 00:00–07:00, and the  $\beta$  values of 12:00–19:00 are smaller than those of 00:00–07:00. According to the parameter analysis mentioned above, the conclusion can be obtained that airport delays propagate faster and wider in 12:00–19:00 than in 00:00–07:00.

The relative cure rate  $\rho$  is the ratio of the infection rate  $\alpha$  to the cure rate  $\beta$ :

$$\rho = \frac{\alpha}{\beta}, \quad (15)$$

where  $\rho$  reflects the recovery speed of delay propagation. The larger  $\rho$  is, the more the delay propagation tends to alleviate. The value of  $\rho$  for 12:00–19:00 is 0.625; the value of  $\rho$  for 00:00–07:00 is 0.737. The relative recovery rate of delayed propagation in 00:00–07:00 is higher than that in 12:00–19:00. This is consistent with the actual situation.

Next, the degree  $k$  and the degree distribution  $p(k)$  of the Chinese airport network in the above two periods are counted. The degree of airport A represents the number of airports with flights to or from A in the statistical time period. Thus, the same airport may have different degrees at

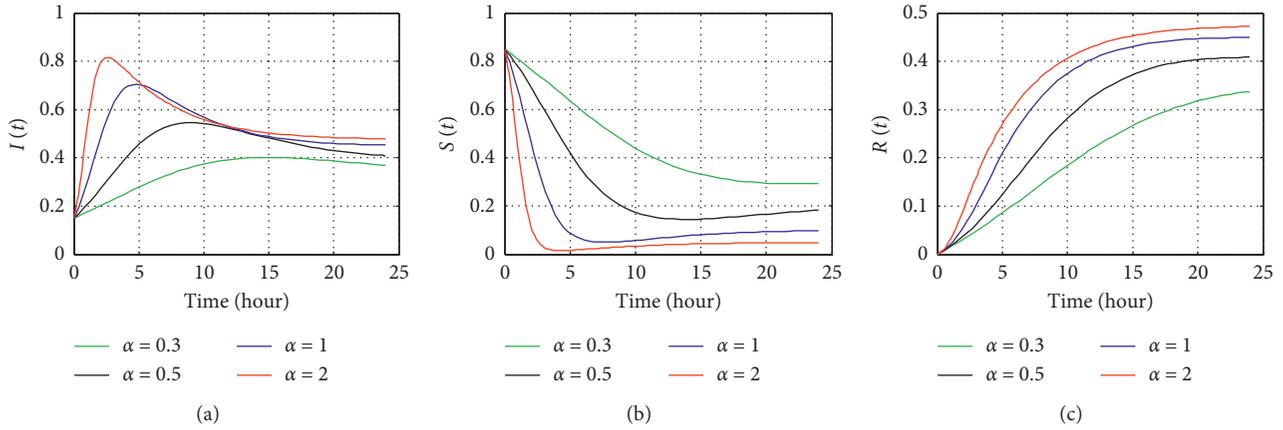


FIGURE 6:  $I(t), S(t)$ , and  $R(t)$  under different  $\alpha$ .

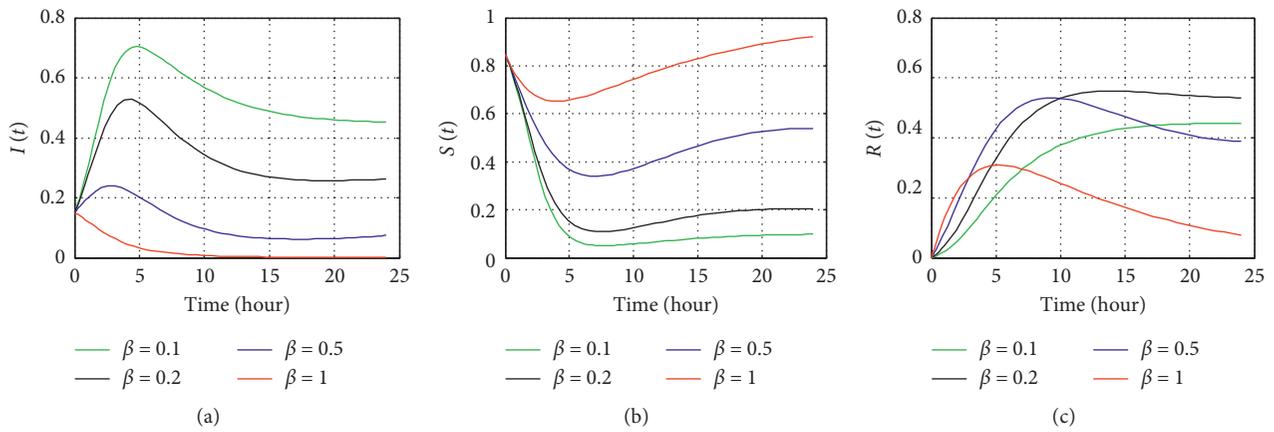


FIGURE 7:  $I(t), S(t)$ , and  $R(t)$  under different  $\beta$ .

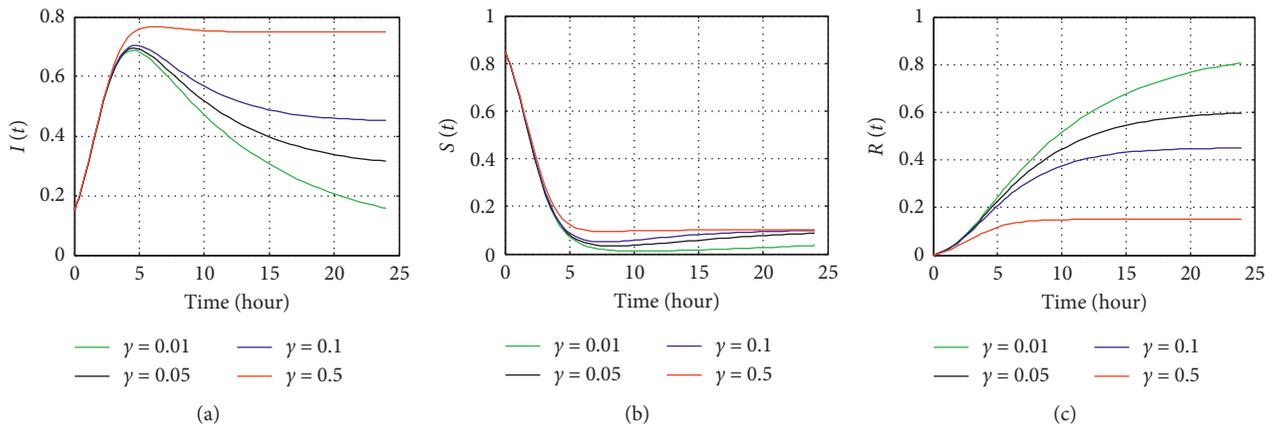


FIGURE 8:  $I(t), S(t)$ , and  $R(t)$  under different  $\gamma$ .

different time periods. Partial results of the statistics are shown in Table 7.

As seen in Table 7, the maximum airport degrees of the nonpeak period and the peak period are 92 and 275, respectively. Figure 12 shows the relationship between  $k$  and  $p(k)$ , with time periods 00:00–07:00 and 12:00–19:00,

respectively. A scaling law between  $k$  and  $p(k)$  can be observed clearly.

Figures 12(a) and 12(b) illustrate two segments that follow the power laws:

$$(1) \text{ 00:00–07:00 } p(k) \sim \begin{cases} k^{-1.1966}, & k \leq 53, \\ k^{-0.4054}, & k > 53. \end{cases}$$

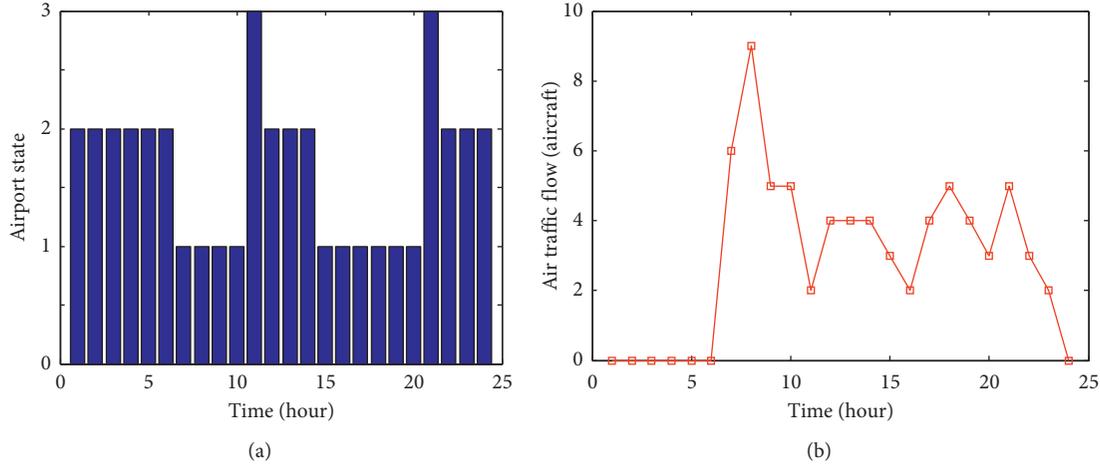


FIGURE 9: (a) Time varying airport states of airport ABNY. (b) Time varying air traffic flow volume of airport ABNY. In (a), “1” represents infected state, “2” represents susceptible state, and “3” represents recovered state.

TABLE 5: The top 10 airports.

Airport	Takeoffs and landings	Ranking
ZBAA	590199	1
ZSPD	449171	2
ZGGG	409679	3
ZGSZ	305461	4
ZPPP	300406	5
ZUUU	293643	6
ZLXY	267102	7
ZSSS	256603	8
ZUCK	255414	9
ZSHC	232079	10

TABLE 6: Samples of  $I(t), S(t)$ , and  $R(t)$  statistics of October 1, 2015.

	0:00-1:00	1:00-2:00	2:00-3:00	...	4:00-5:00	5:00-6:00
	00	00	00	...	00	00
$I(t)$	0.105528	0.080402	0.085427	...	0.381910	0.201005
$S(t)$	0.894472	0.879397	0.884422	...	0.497487	0.582915
$R(t)$	0	0.040201	0.030151	...	0.120603	0.216080

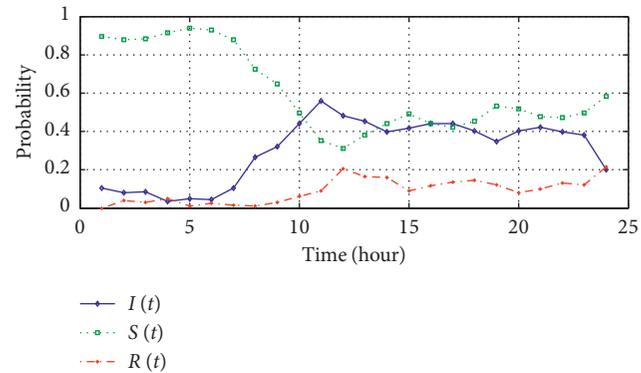


FIGURE 10: Actual value of  $I(t), S(t)$ , and  $R(t)$  of Chinese airport network.

$$(2) \text{ 12:00-19:00 } P(k) \sim \begin{cases} k^{-2.0460}, & k \leq 83, \\ k^{-0.4497}, & k > 83. \end{cases}$$

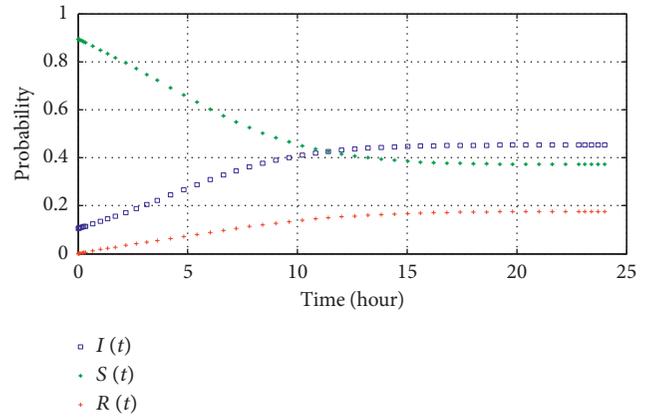


FIGURE 11: Simulated value of  $I(t), S(t)$ , and  $R(t)$  by the ASIRS model.

TABLE 7: Airport degree distribution.

00:00-07:00			12:00-19:00		
$k$	$N$	$p(k)$	$k$	$N$	$p(k)$
1	33	0.340206	1	19	0.104972
2	16	0.164948	2	23	0.127072
3	8	0.082474	3	21	0.116022
4	7	0.072165	4	12	0.066298
...	...	...	...	...	...
53	1	0.010309	185	1	0.005525
55	1	0.010309	201	1	0.005525
60	1	0.010309	220	1	0.005525
92	1	0.010309	275	1	0.005525

The degree distributions are strikingly different from those of random graphs, small-world networks, and scale-free networks. First, the degree distributions of the Chinese airport network display two segments and follow the Double Pareto Law. The degree of the critical airports is approximately 53 for 00:00-07:00 and 83 for 12:00-19:00. The smaller the exponent is, the stronger the heterogeneous characteristics of the network are [44]. The exponents of 12:

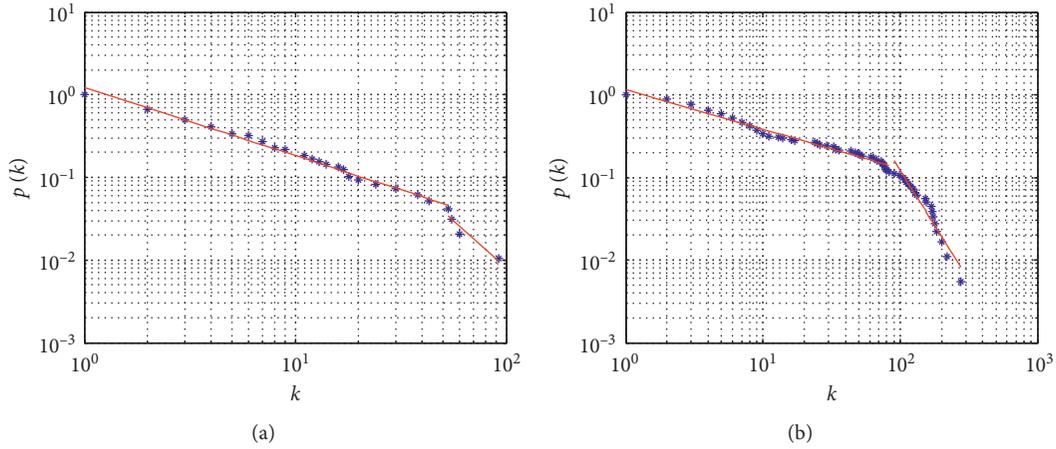


FIGURE 12: Scaling relations between  $k$  and  $p(k)$  at different time periods. (a) 00:00–07:00 and (b) 12:00–19:00.

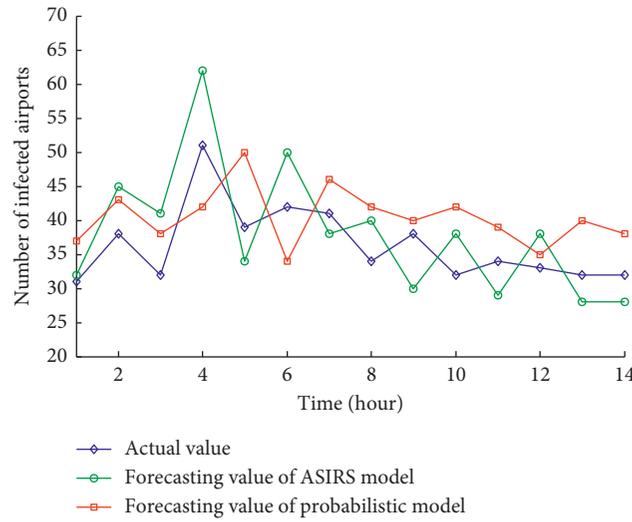


FIGURE 13: Comparison of actual and forecasting values.

00–19:00 are smaller than those of 00:00–07:00, which indicates that the heterogeneous characteristics of the Chinese airport network in 12:00–19:00 are stronger than those in 00:00–07:00. Thus, airport delays during 12:00–19:00 spread over the airport network more easily, quickly, and widely.

Obviously, the results from the ASIRS model and the degree distribution are consistent, which further indicates that the ASIRS model is well suited for characterizing airport delay propagation.

**5.4. Accuracy of the Short-Term Prediction of Delay Propagation.** As mentioned above, the ASIRS model can describe the characteristic of delay propagation in the long term, but the forecast accuracy is not very good (seen in Figures 10 and 11). How about the forecast accuracy of the ASIRS model in the short term?

To gauge the forecast accuracy, we introduce the probabilistic prediction method, which is common to delay forecast [45]. Intercepting the delay period (the period is

usually no more than 4 hours) on a typical day, we forecast the number of delayed airports by constructing ASIRS models for every 4 hours. The prediction results are shown in Figure 13.

We use the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) to compare the accuracy of the two methods.

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{A_t - F_t}{A_t} \right| \times 100\%, \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (A_t - F_t)^2},$$

where  $A_t$  is the actual value in the time interval  $t$ ,  $F_t$  is the forecast value in the time interval  $t$ , and  $T$  is the number of time intervals.

Table 8 shows a comparison of the two methods. It can be seen that the MAPE and RMSE of the SIRS model are smaller

TABLE 8: MAPE and RMSE of the two models compared with historical data.

Models	MAPE (%)	RMSE
ASIRS model	15.9	6.3
Probabilistic model	18.1	7.1

than those of the probabilistic model. Thus, the ASIRS model is more accurate than the model based on probability in forecasting the number of delayed airports in the short term.

## 6. Conclusion

Understanding the process and evolution of airport delay propagation is very important for both air traffic management and aviation planning. In this study, we investigated the mechanism of delay propagation among airports from a new perspective:

- (1) The delay fluctuations of airport networks are studied. To quantify the delay dynamics, we collected the airport delay at 199 Chinese airports and identified the existence of a certain scaling law, which indicates that the dynamic of airport delay is dominated by a propagation factor.
- (2) The ASIRS model for airport delay propagation is presented to reveal the macroscopic appearance of delay propagation. The modeling approach is data driven in the sense that it is based on real China performance data.
- (3) The long-term characteristics of delay propagation is described through building the ASIRS model. The accuracy of the short-term prediction of delay propagation is also examined.

It is remarkable that the airport delay is the result of the coupling of different factors, and there is no information on delay factors in the datasets. We cannot determine a delay is caused by which factors. Thus, we study the delay propagation from the overall delay data and simulate the overall delay without considering the specific factors.

Our ongoing work involves further calibration and validation of the ASIRS model. It is interesting to compare the epidemic model of airport delay propagation in different countries and investigate the practices of the countries. We will come up with insights for mitigating airport delay from such international comparisons.

## Data Availability

The data used to support the findings of this study have not been made available because the data also form part of an ongoing study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This study was supported by the National Natural Science Foundation of China (nos. 71801215, 71671014, and U1833103), Natural Science Foundation of Tianjin (nos. 18JQCQNJC04300), and Fundamental Research Funds for the Central Universities (nos. 3122016C009, 3122019129, and 3122018D026).

## References

- [1] F. Wieland, "Limits to growth: results from the detailed policy assessment tool (air traffic congestion)," in *Proceedings of the 16th Digital Avionics Systems Conference (DASC16)*, Irvine, CA, USA, 1997.
- [2] Faa.gov (Internet), Washington DC: FAA; (up-dated 2018 January 3), [http://aspmhelp.faa.gov/index.php/APM:\\_Definitions\\_of\\_Variables](http://aspmhelp.faa.gov/index.php/APM:_Definitions_of_Variables).
- [3] N. Kafle and B. Zou, "Modeling flight delay propagation: a new analytical-econometric approach," *Transportation Research Part B: Methodological*, vol. 93, pp. 520–542, 2016.
- [4] S. Ahmadbeygi, A. Cohn, Y. Guan, and P. Belobaba, "Analysis of the potential for delay propagation in passenger airline networks," *Journal of Air Transport Management*, vol. 14, no. 5, pp. 221–236, 2008.
- [5] C. Dong, C. Shao, J. Li, and Z. Xiong, "An improved deep learning model for traffic crash prediction," *Journal of Advanced Transportation*, vol. 2018, Article ID 3869106, 13 pages, 2018.
- [6] P. T. R. Wang, L. A. Schaefer, and L. A. Wojcik, "Flight connections and their impacts on delay propagation," in *Proceedings of the 22nd Digital Avionics Systems Conference (DASC22)*, Indianapolis, Indiana, 2003.
- [7] P. Fleurquin, J. J. Ramasco, and V. M. Eguíluz, "Characterization of delay propagation in the US air-transportation network," *Transportation Journal*, vol. 53, no. 3, pp. 330–344, 2014.
- [8] B. Yu, Z. Guo, S. Asian, H. Wang, and G. Chen, "Flight delay prediction for commercial air transport: a deep learning approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 125, pp. 203–221, 2019.
- [9] A. Kondo, "Delay propagation and multiplier," in *Proceedings of the 51st Annual Transportation Research Forum*, pp. 11–13, Arlington, Virginia, March 2010.
- [10] Y. J. Wang, Y. K. Cao, C. P. Zhu et al., "Characterizing departure delays of flights in passenger aviation network of United States," 2017, <https://arxiv.org/abs/1701.05556>.
- [11] W.-B. Du, M.-Y. Zhang, Y. Zhang, X.-B. Cao, and J. Zhang, "Delay causality network in air transport systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 118, pp. 466–476, 2018.
- [12] J. S. Li and J. L. Ding, "Analysis of flight delay propagation using Bayesian networks," *Acta Aeronaut Astronaut Sin*, vol. 29, no. 6, pp. 1598–1604, 2008, in Chinese.
- [13] Y. J. Liu, W. D. Cao, and S. Ma, "Estimation of arrival flight delay and delay propagation in a busy hub-airport," in *Proceedings of the 4th IEEE Computer Society (ICNCA)*, Jinan, China, October 2008.
- [14] W. Wu, C.-L. Wu, T. Feng, H. Zhang, and S. Qiu, "Comparative analysis on propagation effects of flight delays: a case study of China airlines," *Journal of Advanced Transportation*, vol. 2018, Article ID 5236798, 10 pages, 2018.

- [15] P. Fleurquin, Ramasco, J. José, and V. M. Eguiluz, "Systemic delay propagation in the US airport network," *Scientific Reports*, vol. 3, no. 1, p. 1159, 2013.
- [16] S. P. Qiu, "Modeling and analysis of the correlation between flight delays and flight delay propagation based on copula," Dissertation, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 2015.
- [17] J.-T. Wong and S.-C. Tsai, "A survival model for flight delay propagation," *Journal of Air Transport Management*, vol. 23, no. 7, pp. 5–11, 2012.
- [18] X. Zheng, C.-M. Liu, and P. Wei, "Air transportation direct share analysis and forecast," *Journal of Advanced Transportation*, vol. 2020, Article ID 8924095, 12 pages, 2020.
- [19] C. J. Dong, A. J. Khattak, C. F. Shao, and K. Xie, "Exploring the factors contribute to the injury severities of vulnerable roadway user involved crashes," *International Journal of Injury Control and Safety Promotion*, vol. 26, no. 3, pp. 302–314, 2019.
- [20] W.-B. Du, Y. Gao, C. Liu, Z. Zheng, and Z. Wang, "Adequate is better: particle swarm optimization with limited-information," *Applied Mathematics and Computation*, vol. 268, pp. 832–838, 2015.
- [21] Y. Cao, C. Zhu, Y. Wang, and Q. Li, "A method of reducing flight delay by exploring internal mechanism of flight delays," *Journal of Advanced Transportation*, vol. 2019, Article ID 7069380, 8 pages, 2019.
- [22] C. Dong, C. Shao, H. Huang, X. Chen, and N. N. Sze, "Advances in traffic safety methodologies and technologies," *Journal of Advanced Transportation*, vol. 2018, Article ID 4129582, 2 pages, 2018.
- [23] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences*, vol. 101, no. 11, pp. 3747–3752, 2004.
- [24] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, "The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 22, pp. 7794–7799, 2005.
- [25] L. P. Chi, R. Wang, H. Su et al., "Structural properties of US flight network," *Chinese Physics Letters*, vol. 20, no. 8, p. 1393, 2003.
- [26] W. Li and X. Cai, "Statistical analysis of airport network of China," *Physical Review E*, vol. 69, no. 4, Article ID 046106, 2004.
- [27] G. Bagler, "Analysis of the airport network of India as a complex weighted network," *Physica A: Statistical Mechanics and Its Applications*, vol. 387, no. 12, pp. 2972–2980, 2008.
- [28] S. W. Cheng, Y. P. Zhang, S. Q. Hao, R. W. Liu, X. Luo, and Q. Luo, "Study of flight departure delay and causal factor using spatial analysis," *Journal of Advanced Transportation*, vol. 2019, Article ID 3525912, 11 pages, 2019.
- [29] M. Zanin, S. Belkoura, and Y. Zhu, "Network analysis of Chinese air transport delay propagation," *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 491–499, 2017.
- [30] W. B. Du, "Research on complex network and network dynamics for air transportation systems," Dissertation, Anhui: University of Science and Technology of China, Huainan, Chi, 2010.
- [31] S. M. Li, X. H. Xu, and L. H. Meng, "Fluctuations in airport arrival and departure traffic: a network analysis," *Chinese Physics B*, vol. 21, no. 8, Article ID 088901, 2012.
- [32] H. Wang, Z. Song, R. Wen, and Y. Zhao, "Study on evolution characteristics of air traffic situation complexity based on complex network theory," *Aerospace Science and Technology*, vol. 58, pp. 518–528, 2016.
- [33] A. Gautreau, A. Barrat, and M. Barthélemy, "Microdynamics in stationary complex networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 22, pp. 8847–8852, 2009.
- [34] X. P. Wu, H. Y. Yang, and S. C. Han, "Analysis of properties and delay propagation of air traffic based on complex network," *Acta Aeronaut Astronaut Sinica*, vol. 38, no. S1, pp. 1–6, 2017.
- [35] L. Zheng and L. Tang, "A node-based SIRS epidemic model with infective media on complex networks," *Complexity*, vol. 2019, Article ID 2849196, 14 pages, 2019.
- [36] T. Q. Chen, B. C. Ma, and J. N. Wang, "SIRS contagion model of food safety risk," *Journal of Food Safety*, vol. 38, no. 1, Article ID e12410, 2018.
- [37] Y. Muroya, Y. Enatsu, and H. Li, "Global stability of a delayed SIRS computer virus propagation model," *International Journal of Computer Mathematics*, vol. 91, no. 3, pp. 347–367, 2014.
- [38] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [39] A.-L. Barabasi, M. A. De Menezes, S. Balensiefer, and J. Brockman, "Hot spots and universality in network dynamics," *The European Physical Journal B—Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 169–175, 2004.
- [40] M. Argollo de Menezes and A. L. Barabasi, "Separating internal and external dynamics of complex systems," *Physical Review Letters*, vol. 93, Article ID 068701, 2004.
- [41] Z. Eisler, J. Kertész, S.-H. Yook, and A.-L. Barabási, "Multiscaling and non-universality in fluctuations of driven complex systems," *Europhysics Letters (EPL)*, vol. 69, no. 4, pp. 664–670, 2005.
- [42] R. N. Mantegna and H. E. Stanley, "Scaling behaviour in the dynamics of an economic index," *Nature*, vol. 376, no. 6535, pp. 46–49, 1995.
- [43] Z. Eisler and J. Kertesz, "Random walks on complex networks with inhomogeneous impact," *Physical Review E*, vol. 71, Article ID 057104, 2005.
- [44] J.-F. Zheng, Z.-Y. Gao, and X.-M. Zhao, "Properties of transportation dynamics on scale-free networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 373, pp. 837–844, 2007.
- [45] E. R. Mueller and G. B. Chatterji, "Analysis of aircraft arrival and departure delay characteristics," in *Proceedings of the AIAA Aircraft Technology, Integration and Operations (ATIO) Conference*, p. 5866, Los Angeles, CA, USA, October 2002.