Chaos

# Modeling and optimizing the delay propagation in Chinese aviation networks

Cite as: Chaos **29**, 081101 (2019); https://doi.org/10.1063/1.5111995
Submitted: 01 June 2019 . Accepted: 19 July 2019 . Published Online: 12 August 2019

Shuo Qin, Jianhong Mou, Saran Chen, and Xin Lu

Chaos ARTICLE scitation.org/journal/cha

# Modeling and optimizing the delay propagation in Chinese aviation networks

Cite as: Chaos **29**, 081101 (2019); doi: 10.1063/1.5111995
Submitted: 1 June 2019 · Accepted: 19 July 2019 ·
Published Online: 12 August 2019

View Online Export Citation CrossMark

Shuo Qin,[1,a)] Jianhong Mou,[1,a)] Saran Chen,[1,2] and Xin Lu[1,3,4,b)]

## AFFILIATIONS

[1]College of Systems Engineering, National University of Defense Technology, 410073 Changsha, China
[2]School of Mathematics and Big Data, Foshan University, 528000 Foshan, China
[3]School of Business, Central South University, Changsha 410083, China
[4]Department of Public Health Sciences, Karolinska Institute, Stockholm 17177, Sweden

[a)]S. Qin and J. Mou contributed equally to this work.
[b)]**Electronic mail:** xin.lu@flowminder.org

## ABSTRACT

We define metrics to quantify the level of overall delay and propose an agent-based data-driven model with four factors, including aircraft rotation, flight connectivity, scheduling process, and disturbance, to build a simulator for reproducing the delay propagation in aviation networks. We then measure the impact on the propagation by the delay at each airport and analyze the relevance to its temporal characteristics. When delay occurs, airline schedule planning may become infeasible, and rescheduling of flights is usually required to maintain the function of the system, so we then develop an improved genetic algorithm (GA) to reschedule flights and to relax the root delay. Results indicate that priority-based strategy rather than First-Come-First-Serve can achieve minimum overall delay when congestion occurs, and aircraft rotation is the most important internal factor contributing to delay propagation. Furthermore, the reschedule generated by the improved GA can decrease delay propagation more significantly compared to the agent-based model.



**Air transportation has become the main means of travel connecting different countries and different races. However, flight delays have become increasingly serious in recent years. In aviation systems, aircraft delay propagates if the slack time between consecutive flights is insufficient, and the propagation is further aggravated along strong ties connecting airports with heavy traffic. If the arrival of the former flight gets delayed and cannot be absorbed by the buffer time at the airport, the departure of the second flight may be delayed; and if the departure delay of the next flight cannot be absorbed through increasing the in-air speed, its arrival may also get delayed. This cascadelike effect enables the propagation of flight delays and may cause severe consequences. We propose an agent-based data-driven model for reproducing the delay propagation in aviation networks, and we investigate the behavior of the model under different factors. At the end of this study, we discuss two rescheduling strategies to guide the design of controlling strategies for effectively reducing flight delay propagation. This study is crucial for understanding delay propagation in airport networks and considering the temporal interactions among flights from all companies, it provides potential strategies for developing more robust schedules such that the overall delay of flights can be minimized.**

## I. INTRODUCTION

Air transportation has become the main means of travel connecting different countries and different races.[1–5] Further, it has overtaken rail transportation to become the more popular method used by travelers, and it has made many global pursuits much more accessible, such as resource allocation,[6] the forecasting of epidemics,[7] the optimization of transportation systems,[8] and disaster response.[9]

Air transportation systems have been traditionally expressed as graphs with vertices and edges representing airports and flights, respectively. These graphs are called aviation networks or airport networks and have been studied at different spatial[10–13] or temporal resolution.[14–16] While the study of aviation networks has been extended to help understand complex phenomena such as economic growth[17] or the spreading pattern of globally transmitted diseases,[18] many researchers also started focusing on transportation system itself for complexity dynamics such as delay propagation.[19–24]

Flight delays are inevitable for many reasons, such as extreme weather conditions, unplanned maintenance issues, or air traffic control complications. In recent years, flight delays have become increasingly common, particularly in China, which has the second largest number of air transport passengers in the world after the US. However, according to the 2017 Statistical Bulletin of Civil Aviation Industry Development,[25] only 71.25% of 2.129 million flights are on time, and the direct economic impact caused by flight delays is about $50 billion. The situation may become more severe in the next decade with the rapid development of aviation networks and the large-scale growth of the number of future routes. When a delay happens, the coupling relationship between flights will lead to more frequent occurrence of delay propagation with consecutive flights. If the arrival of the former flight gets delayed and cannot be absorbed by the buffer time at the airport, the departure of the second flight may be delayed; and if the departure delay of the next flight cannot be absorbed through increasing the in-air speed, its arrival may also get delayed. This cascadelike effect enables the propagation of flight delays and may cause severe consequences. To figure out the mechanism underlying the propagating phenomenon, attempts have been made with a variety of approaches, such as Bayesian network,[26] survival model,[24] and epidemic spreading model.[18] These studies differ in the level of details included in their models, but in general they succeed in modeling the propagation for a few major airports. In addition, an alternative taking a systemic or network-wide perspective is proposed by Pablo Fleurquin.[27]

Furthermore, researchers have proposed different approaches to find improved strategies for recovering when disruptions occur, to minimize the overall delay in the system and prevent propagation. Abdelghany *et al.*[28] develop a decision support tool to automate crew recovery during irregular operations. Lettovskỳ *et al.*[29] develop a real-time recovery plan to restore a disrupted crew schedule. Yu and Qi[30] have studied the recovery models used by United Airlines in the context of disruption management. Stojković *et al.*[31] focus on how to modify an existing plan in order to recover from a set of minor disruptions. The objective of these approaches is to minimize the costs associated with extra resource utilization and passenger inconveniences. Lan *et al.*[32] consider how changes in aircraft routes can be used to reduce the potential for delays to propagate via connecting flights. The objective of their approach is to decrease the impact of delay on passengers' ability to make flight connections. Cohn *et al.*[33] decrease airline delay propagation by developing a flight retiming model which focuses on minimizing the propagation of root delays. Akturk *et al.*[34] undertake the first study in which the cruise speed is explicitly included as a decision variable in an airline recovery optimization model along with the environmental constraints and costs. Their model allows for an investigation of the trade-off between flight delays and the recovery cost.

However, though most of the above work succeeded in modeling and decreasing delay propagation in daily operations,[35–40] and most existing studies have found many factors which affect flight delays, it is not yet clear on how flight delays spread; it is difficult to accurately describe the mechanisms of delay propagation in the aviation network considering the airlines' dependence and all the internal and external factors in the system. To fill in this gap of knowledge, we try to create a model for reproducing delay propagation in aviation networks by considering four factors, including aircraft rotation, flight connectivity, scheduling process, and disturbance,[27] and then, we use this model as a benchmark to evaluate the influence of factors and to quantify the possibility of improvements by optimized rescheduling when delay happens. Unlike most previous studies, we have access to the most synthetic dynamic Chinese aviation network, with empirical data which includes all flight companies and contains the most up-to-date high spatial and temporal resolution. Moreover, as the mechanisms for the dynamics of aviation systems are extremely complicated and are not possible to quantify with tens of thousands of parameters, we build a black-box model, i.e., a data-driven agent-based simulator, to reproduce the system. Finally, compared to the current rescheduling strategy from real data, we find that it is possible to achieve much less overall delay with the new rescheduling strategy obtained by our algorithm.

## II. DATA SETS

The data were crawled from VariFlight,[41] which provides the actual and scheduled times of flights of Chinese Aviation Networks (CANs). The dataset covers all civil airlines in China and contains 33 995 scheduled and actual flights for 10 days in 2016 (from September 7 to September 13 and from December 3 to December 5) operated by domestic carriers connecting 211 different domestic airports. The information for each flight includes the actual and scheduled departure and arrival time, the origin and destination airport, historical punctuality, etc. The information about the aircraft includes the aircraft type, the age of aircraft, and the airline company. With these information, we then are able to generate the flight sequences of aircrafts. It is worth noting that the data collection period covers December 4, 2016, on which a heavy smog and fog occurred at the Chengdu Shuangliu Airport (CTU), and serious flight delays and congestion were recorded. The runway was closed for nearly 10 h, 49 flights were canceled, and over 20 000 passengers were left stranded by delayed or canceled flights.

## III. MODEL FOR DELAY PROPAGATION IN AVIATION NETWORKS

Agent-based models are a kind of microscale model which is used to simulate the simultaneous operations and interactions of multiple agents in an attempt to recreate and predict the appearance of complex phenomena. The data-driven approach focuses on building a system based on a large number of observed datasets. The strength of this approach is that it depends on the empirical data rather than on a set of rules or parameters hypothesized by the users. In this section, we propose an agent-based data-driven model consisting of four factors, including aircraft rotation, flight connectivity, scheduling process, and disturbance, to build a systematic model for reproducing the delay propagation in aviation networks. As shown in Fig. 1, first, the model is configured with initial delay (the delay in the first mission of each aircraft), which is extracted from the empirical data. Second, after initialization, the model considers the delay caused by both aircraft rotation and flight connectivity as well as the disturbance which may occur for flights waiting for departure at the

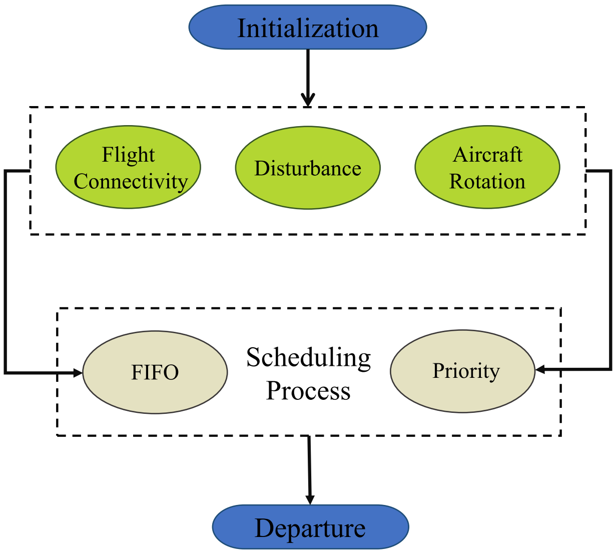


**FIG. 1.** The agent-based data-driven model.

airport. Third, the model applies and compares two strategies for scheduling aircraft takeoff sequences, on the premise of meeting the airport's takeoff and landing capacity.

## A. Aircraft rotation

Throughout a day, each aircraft follows the connections given in the schedule, the so-called aircraft rotations.[27] To make the most efficient use of their aircraft, the airline companies typically schedule several flights for each aircraft. Consequently, one flight may be delayed if the preceding flight cannot land on time, and this kind of delay can cascade and propagate further delays. However, the propagation of delays can be diminished gradually via two time-saving processes: accelerating the flight preparation work at airports and increasing speed during the fly.

It is common knowledge that flights can arrive early but cannot depart early. The propagation of delay caused by aircraft rotation is shown in Fig. 2, in which $S^{apt}{}_i = \alpha \times O_i$ and $S^{air}{}_i = \beta \times F_i$ represent the saved time generated by the two previously mentioned processes: faster flight preparation ($S^{apt}$) and increased flight speed ($S^{air}$). In the above notation, $\alpha = (O_i - O_i^*)/O_i$, in which $O_i$ and $O_i^*$ represent the scheduled and actual stopover time of flight $i$ at the airport, respectively; and $\beta = (F_i - F_i^*)/F_i$, in which $F_i$ and $F_i^*$ represent the scheduled and actual flight time, respectively. Obviously, during the two process of flight preparation/stopover and flying, when $\alpha$ and $\beta$ are negative, it means that there is no time saved by these two process, the actual stopover time and flight time have exceeded the scheduled time; and when their values are positive, it means that the processes are accelerated than planned. If its former flight $j$ is delayed for $D_j$ and cannot recover via the buffer time at the airport ($D_j > S^{apt}{}_i$), flight $i$ will depart late for $D_j - S^{apt}{}_i$. Then, if $i$ can absorb the departure delay along the way, it will arrive early for $S^{air}{}_i - (D_j - S^{apt}{}_i)$; otherwise, it will delay for $D_j - S^{apt}{}_i - S^{air}{}_i$. exceed

According to the Normal Statistical Methods for Civil Aviation Flights, flights more than a 15 min late are considered delayed. In addition, faster preparation time at airports is only applicable to aircrafts carrying out multiple missions. We analyze the distribution of faster flight preparation work and increased flight speed from the empirical data (Fig. 3). $\alpha < 0$ indicates that the flight was delayed at the airport [Fig. 3(a)], which may attributable to slow boarding by the passengers, mechanical failure, etc. However, it is worth noting that there are other common factors which may lead to delay, such as airport congestion, regardless of the on-time status of the preceding flight. From Fig. 3(b), we can see that both delayed and on-time flights tend to speed up during the fly, on average, each flight saves about 25% traveling time by speeding up. Only a small proportion (1.3%) of flights are slower than scheduled, i.e., with $\beta < 0$.

## B. Flight connectivity

Flight connectivity involves the transfer of passengers and crews between flights with different aircrafts. Without loss of generalizability, in this study, we assume that transfer time of passengers and crews of former flights requires at least $t_{min} = 1$ h, and that the temporal gap between the departure time of the subsequent flight and the landing time of the former flight shall not exceed $t_{max} = 3$ h. In addition, transfers occur only between airports without direct flights. The probability of connecting is set as the historical punctuality of

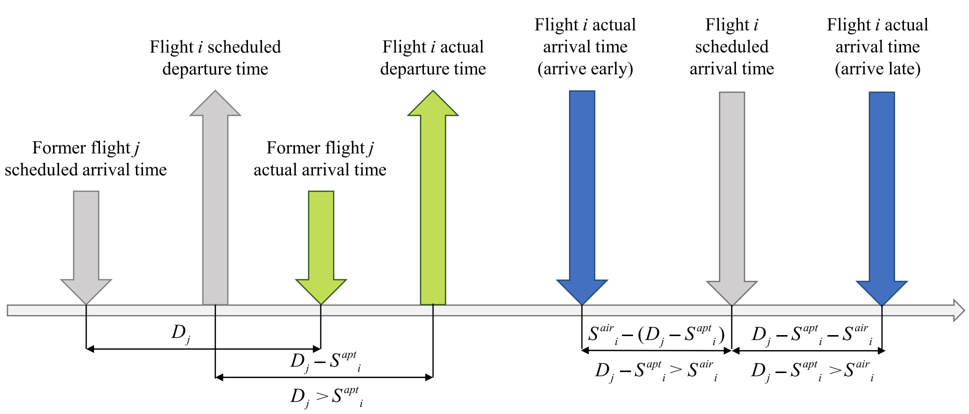


**FIG. 2.** The subprocess of aircraft rotation.

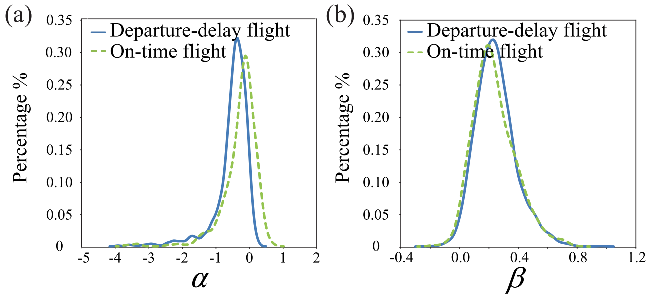


**FIG. 3.** Two time-saving processes: (a) accelerating the flight preparation work at airports and (b) increasing speed during the fly in CAN during the data collection period.

each flight. A more complicated factor is that passengers on a second flight may be a combination of individuals transferring from different incoming flights, and the actual departure time depends on the boarding time of the last passenger to board. As illustrated in Fig. 4, passengers and crews of flight $i$ come from flights $a$, $b$, and $c$, with scheduled stopover time $O_a > O_b > O_c$. When flight $a$ is on time but flights $b$ and $c$ are delayed, the final delay of flight $i$ equals $D_b - \alpha \times O_b$, while $D_b - \alpha \times O_b > D_c - \alpha \times O_c$.

## C. Scheduling process

### 1. Air traffic control

Airports have limited capacity for departing and landing flights in a given time period. In addition, a temporal gap between departure flights or arrival flights must be guaranteed to ensure safety. Thus, some flights may be delayed when the number of flights in the queue surpasses the capacity of the airport during a given time period. The takeoff capacity of an airport is the number of scheduled departure flights (SDF) taken by the airport during a takeoff time, and the landing capacity refers to the number of scheduled arrival flights (SAF) of an airport during a landing time. If the number of actual departure flights (ADF) or actual arrival flights (AAF) is greater than the capacity threshold, flights in the airport will need to wait for takeoff or landing. Airport dispatchers must take into account the capacity constraints of the departing and arriving flights in scheduling, and flight takeoff can only be arranged when both of them fit the threshold. After investigating the departuring and landing frequency during busy hours from the empirical data, we assume in the model that the duration of takeoffs is about 10 min for a single-runway airport and 5 min for two runways.

There may be a large difference between scheduled departure/landing capacity and actual departure/landing capacity in the case of large-scale delay. For example, on December 4, 2016, a serious delay caused by heavy smog and fog occurred at Chengdu Shuangliu Airport [Figs. 5(a) and 5(b)], which led to a large difference between its scheduled capacity and the actual capacity for taking off and landing, while Beijing Capital Airport [Figs. 5(c) and 5(d)], with a relatively small delay, had a scheduled landing and departure capacity similar to the actual landing and departure capacity.

### 2. Scheduling

When delay occurs, the airport needs to reschedule the delayed flights, so as to minimize the impact on other on-schedule flights, as well as to reduce the effect of cascading consequences. While in practice the rescheduling is a compromise of complex factors, in the model we consider two strategies: First Come First Serve (FCFS) and priority-based.

*a. FCFS.* FCFS operates on the premise of meeting the airport's departure and landing capacity; the flight departure order is queued according to the actual landing time of the flight.

*b. Priority-based.* With the priority-based approach, flights with higher priority will depart earlier. Factors affecting the priority of flights include the economic value of aircrafts, the delayed time of flights, and the number of flights with the same aircraft. While it is not

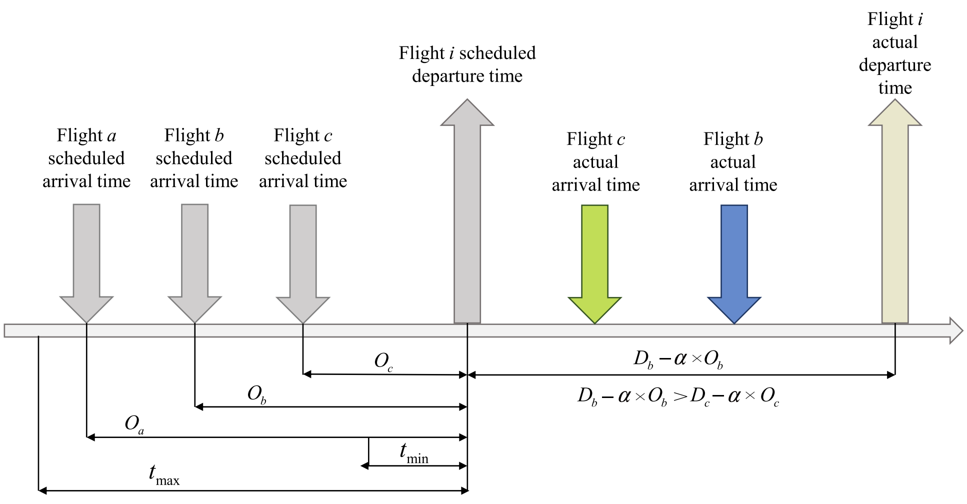


**FIG. 4.** The subprocess of flight connectivity.

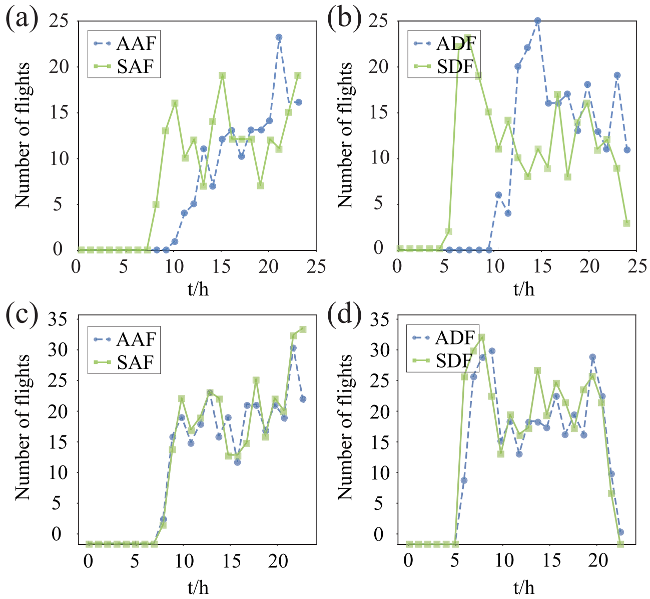


**FIG. 5.** The hourly AAF, SAF, ADF, and SDF at Chengdu Shuangliu Airport (CTU) and Beijing Capital Airport (PEK) on December 4, 2016. (a) AAF and SAF at CTU airport, (b) ADF and SDF at CTU airport, (c) AAF and SAF at PEK airport, (d) ADF and SDF at PEK airport.

feasible to quantify all these factors, we adopt an approach developed by Milan[42] to measure priority with two calculations: static priority (SP) for on-time flights and dynamic priority (DP) for delayed flights.

The static priority of a normal flight depends on the intrinsic characteristics of the flight and the number of passengers and is calculated by[43]

$$(SPF)_i = n_i \cdot (c_i + \theta_i \lambda_i N_i) / s_i, \tag{1}$$

where $c_i$ is the direct economic loss per unit of time of delayed flight $i$, $\theta_i$ is the direct economic loss per unit of time of passengers, $\lambda_i$ is the passenger occupancy rate, $N_i$ is the number of seats, and $s_i$ is the scheduled service time for the flight at the destination airport.

Flights with short scheduled service time have little buffer time for absorbing delay, thus being more vulnerable to cause cascade-like propagation of delay. Consequently, the higher the passengers' "value" is, or the shorter service time of the flight at the destination airport is, the higher priority the flight has. Although this single standard can effectively reduce the economic losses, flights are not treated equally, especially for those with lower "value." Further, static priority also fails to take into account the heterogeneity of the length of each flight delay.

Dynamic priorities of delayed flights take into account the intrinsic nature of the flight and the flight delays, including

(1) *Flights' characteristics*: the economic loss per unit of time when flight delays occur on the ground or along the way, the number of seats, and the scheduled transfer time of flights as discussed above.
(2) *Passengers' characteristics*: the number of passengers, passengers' economic loss per unit of time when flights delay, and the proportion of transferring passengers and the time they require.
(3) *Length of flight delays*: the delay occured on the ground and/or during the fly.

Then, the dynamic priority of a flight can be calculated as

$$(DFP)_i = (h_{i1} + h_{i2})/s_i, \tag{2}$$

$$h_{i1} = n_i \cdot (c_i + \theta_i \lambda_i N_i) \cdot w_i, \tag{3}$$

$$h_{i2} = (\theta_i \lambda_i N_i q_i) \cdot w_i, \tag{4}$$

where $h_{i1}$ is the economic loss caused when the flight delays on the ground at the departure airport, $h_{i2}$ is the economic loss of passengers who need to transfer, $n_i$ is the number of flights carried out by the same aircraft of flight $i$, $w_i$ is the delay time, and $q_i$ is the ratio of transferring passengers. In this study, we assume that passengers will not choose to transfer between airports which have direct flights.

In the following, we assume that all flights are full, that is, $\lambda_i = 1$; the passenger transfer rate $q_i$ is proportional to the historical on-time rate of subsequent flights $p_{i+1}$, that is, $q_i \propto p_{i+1}$; and the direct economic loss per unit of time $c_i$ of delayed flights $i$ is proportional to the number of seats, that is, $c_i \propto N_i$ and $\alpha_i = 1$. Thus, we can rewrite Eqs. (1) and (2) as

$$(SPF)_i \propto 2n_i N_i / s_i, \tag{5}$$

$$(DFP)_i \propto (2n + p_{i+1}) \cdot N_i w_i / s_i. \tag{6}$$

In addition, we should take other factors into consideration, such as the boundary conditions, where there should be no delayed flights when all root flights are on-time. Further, the flight schedule is the result of the coordination of airlines and airports, and thus, the above priorities may not exactly match the order of flight schedules, so flights within a takeoff resource need to be classified and scheduled as the following:

(1) Flights are scheduled according to flight schedules when there exist only on-time flights in a unit of time.
(2) The priority of flight will be scheduled by the dynamic priority method when there exist only delayed flights in a unit of time, and the rest will delay for a unit of time and compete with flights within the subsequent unit of time.
(3) If there exist both on-time and delayed flights in a unit of time, the priority of delayed flights will be scheduled based on the dynamic priority method, while the priority of on-time flights will be scheduled according to the static priority method.

## D. Disturbance

We set the delays caused by other factors in addition to the aforementioned as the disturbance of flights and assume that disturbance occurs at airports (i.e., on the ground). Taking into account the impact of number of flights at an airport on random delays for $a$ to $b$ minutes, we mimic the disturbance using a uniform distribution $(d_i/d_{max}) \cdot U(a, b)$, where $d_i$ and $d_{max}$ are the degree (the number of connections to other airports) and maximum degree of departure airport of flight $i$. Considering that the effect of disturbance only accounts for a small part in the model (random noise), we set $a = 0$,

$b = 60$. All of these parameters are set according to practice in the Chinese aviation system as well as to our simulation tests, and they will not affect the main conclusion of this study.

## IV. RESULTS AND DISCUSSION

### A. The effect of each factor

In order to analyze the behavior of the model under different days and different factors, we undertake experiments on a day with serious delays (December 4, 2016) and on a day with moderate delays (September 8, 2016). We then run the agent-based data-driven model not only to evaluate the effect of four individual factors on delay propagation separately but also to analyze the synthesized models with combinations of different factors.

We use the ratio of delayed flights $r_d$ and the average time of delay $d_{avg}$ to quantify the level of systemic delay. We obtain 48 pieces of networks with a sliding time window of $T = 30$ min for one day. We then calculate $r_d$ and $d_{avg}$ in each time window and use this sequence as the output of the model. To illustrate the distinction between initial delays and actual delays on CAN under different factors, we calculate the Euclidean distance as the similarity between a sequence of actual delays and a sequence of delays produced by these models,

$$E_r = \sum_{i \in S^r{}_1} (S^r{}_1(i) - S^r{}_2(i)),$$
$$E_d = \sum_{i \in S^d{}_1} (S^d{}_1(i) - S^d{}_2(i)), \tag{7}$$

where $E_r$ and $E_d$ denote the difference between the two sequences in terms of the proportion of delayed flights and the average delay time, respectively; $S^r{}_1(i)$, $S^r{}_2(i)$, and $S^d{}_1(i)$, $S^d{}_2(i)$ are the simulated and actual sequences in terms of the proportion of delayed flights and the average delay time, respectively. Under the expectation that strategies generating results closer to reality are more likely to be the mechanisms driving the evolution of CAN, we use a two-dimension comparison to rule out potential mechanisms for further simulation (Fig. 6).

As shown in Fig. 6, the model considering both disturbance and priority (PD model) can reproduce the delay propagation on days with moderate (September 8) and serious (December 4) delay very well, while the model considering FCFS and disturbance (FCFSD model) performs better on days with serious delay. Under normal circumstances when there exist fewer flights with serious delay, delays of a large number of flights may be absorbed at airports or during the fly, making the effect of aircraft rotation being nonsignificant or even less than the effect of disturbance. In addition, aircraft rotation is the most significant factor in delay propagation on days with serious delay [Fig. 6(a)], while the priority of flights plays an important role on days with moderate delay [Fig. 6(b)]. Moreover, for delay spreading occuring mainly through aircraft rotation, the FCFSD model sufficiently explains the delay propagation, and the PD model may overestimate the propagation on days with serious delay. Similarly, on days with moderate delay, the PD model performs well at calculating the delay propagation, and the FCFSD model, mainly influenced by aircraft rotation. Disturbance of flights may increase $r_d$ and $d_{avg}$ no matter whether it functions in conjunction with the priority-based

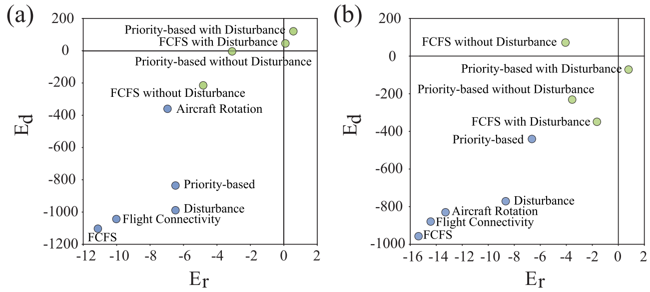


**FIG. 6.** The effects of four factors on different initial delay systems and the performance of models in CAN under different days (a) on December 4, 2016 and (b) on September 8, 2016. The synthesized models include the following combinations of factors: FCFS with disturbance, FCFS without disturbance, priority-based with disturbance, and priority-based without disturbance. The positive or negative values of numbers imply the overestimation and underestimation of actual delay, respectively.

model or the FCFS model on days with serious delay. Interestingly, as shown in Fig. 6(b), under normal circumstances, random disturbance under the FCFS mechanism will undoubtedly increase $r_d$ while decrease $d_{avg}$. Random disturbance may weaken the effect of the FCFS mechanism on days with moderate delay.

The distribution of delay time of each flight is the best characteristic of delay in CAN. As shown in Fig. 7, a comparison of two days reflects that the delay on December 4 [Figs. 7(a) and 7(b)] exhibited a more pronounced fat-tailed distribution than that on September 8

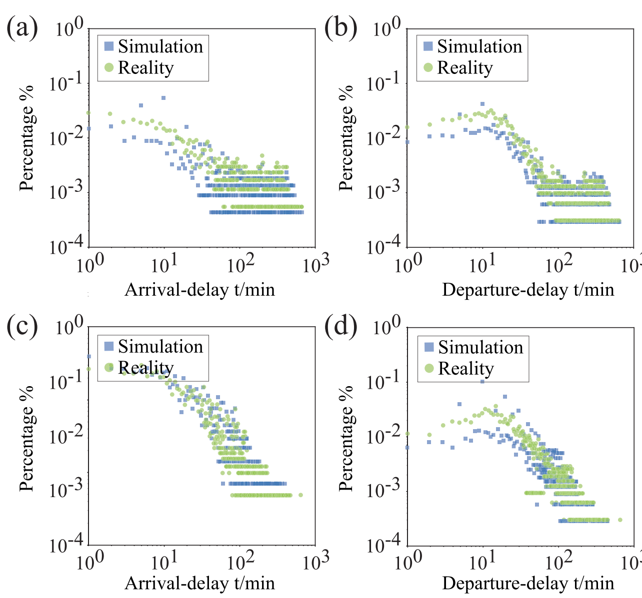


**FIG. 7.** Comparison of distribution of departure and landing delay time between the simulation and empirical data. The comparison of (a) arrival delay on December 4, (b) departure delay on December 4, (c) arrival-delay on September 8, and (d) departure delay on September 8.

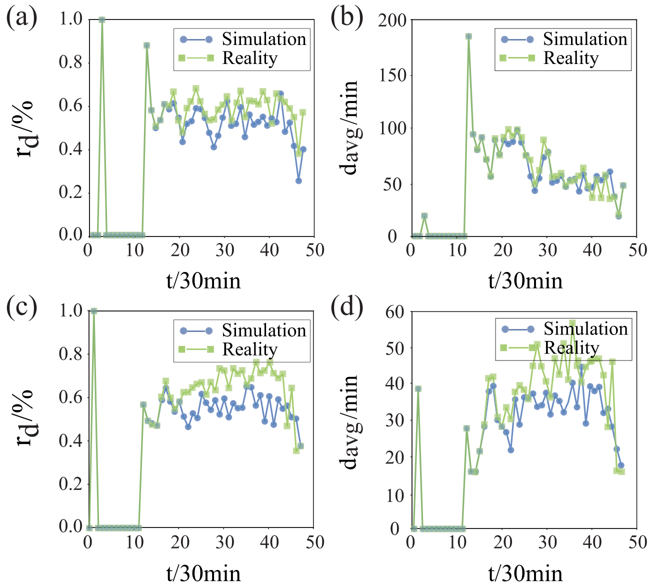


**FIG. 8.** Comparison of $r_d$ and $d_{avg}$ between the simulation and empirical data. The comparison of (a) $r_d$ on December 4, (b) $d_{avg}$ on December 4, (c) $r_d$ on September 8, and (d) $d_{avg}$ on September 8.

[Figs. 7(c) and 7(d)] on both takeoff and landing, indicating that more flights were delayed on December 4. In addition, the priority-based model has a better fit for the departure delay on December 4 than that on September 8. The main reason was that Chengdu Shuangliu Airport was closed on the morning of December 4, causing a long delay in the initial flights, and this dominating effect makes the impact of other factors become insignificant.

Figure 8 shows the difference between the simulation data and the empirical data for both $r_d$ and $d_{avg}$, representing the validity of the model on the operation of aviation networks. We can see that the model fits the data well on both September 8 and December 4. Apart from this, while there is a constant gap (10% to 15%) between simulation and reality on September 8 [Figs. 8(c) and 8(d)], this gap is not obvious on December 4, which is an indication of the delay caused by other factors.

## B. Cascading effect of delay at airport

In order to determine the impact of initial delay (the number of flights that have been delayed at an airport at the beginning of a time period) on the propagation intensity in different time periods, we divide a day into four parts and set a 2-hour delay for all flights at each airport in each period, so as to simulate the large-scale effects of extreme weather, terror threats, etc. Then, we define $D_x = \sum_{i=1}^{f_x} d_{xi}$ as a variable quantifying the cascading effect of delay at an airport, where $D_i$ represents the total delay time caused by the failure of airport $x$, $d_{xi}$ indicates the delay time of the flight $i$ at airport $x$, and $f_x$ indicates the number of flights at airport $x$.

As shown in Fig. 9, almost all flights are scheduled after 06:00am. Large airports such as Beijing Capital Airport, Guangzhou Baiyun Airport, Kunming Changshui Airport, and Chengdu Shuangliu Airport had a large number of initial delayed flights in all three time periods, and all led to large total delay at the end of the day. This is because large airports also accommodate more flights, and the delay of these flights cannot be absorbed very well, which leads to the delay in subsequent flights and may trigger large systematic delay. In addition, there are a large number of shallow-colored nodes with large size in all three segments, indicating airports with small initial delays resulted in significant systemic delay, including Hohhot Baita International Airport, Yinchuan Hedong International Airport, and Tianjin Binhai International Airport. Overall, the effect of airport on flight delays is strongly dependent on the number of flights operating during that period, since a large number of delayed flights will increase the queuing time of other flights and make the incoming and outgoing flow of the airport exceed the planned thresholds.

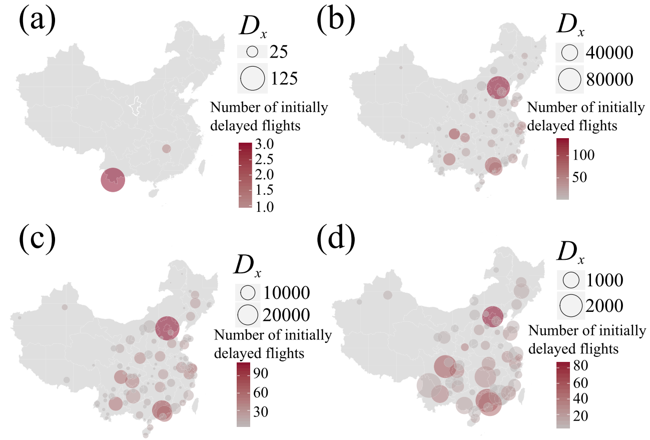


**FIG. 9.** Spatial and temporal distributions of airport delay in CAN during four time periods. (a) 0:00–6:00, (b) 6:00–12:00, (c) 12:00–18:00, and (d) 18:00–24:00. The color of each circle is proportional to the initial delay of the airport, and the circle size represents the total delay that has propagated at the end of all time periods.

We then define four metrics of temporal characteristics of airports as the following to investigate the relationship between overall delay (intensity of cascading effect) and the temporal characteristics of airports: (1) ${Td_x}^{out}$ is the set of nodes that can be reached by temporal paths from node $x$, which is similar to out-degree in an aggregated network. (2) ${Td_x}^{in}$ is the source set of $x$ that can reach through temporal paths, which is similar to in-degree in a network. (3) ${C_x}^T$ is the temporal closeness centrality of node $x$ as the measurement of the speed at which nodes in the temporal network can reach other nodes and is defined as

$$ {C_x}^T = \frac{1}{N-1} \sum_y (1 - \tau_{xy}), \tag{8} $$

where $\tau_{xy}$ represents the average temporal distance between node $x$ and node $y$, and $N$ is the number of nodes. (4) ${B_x}^T$ is the temporal betweenness of $x$ as the metrics of the importance of node $x$ on the temporal shortest paths.

In Fig. 10, we have presented the relationship between the overall delay and the temporal characteristics of airports. As the

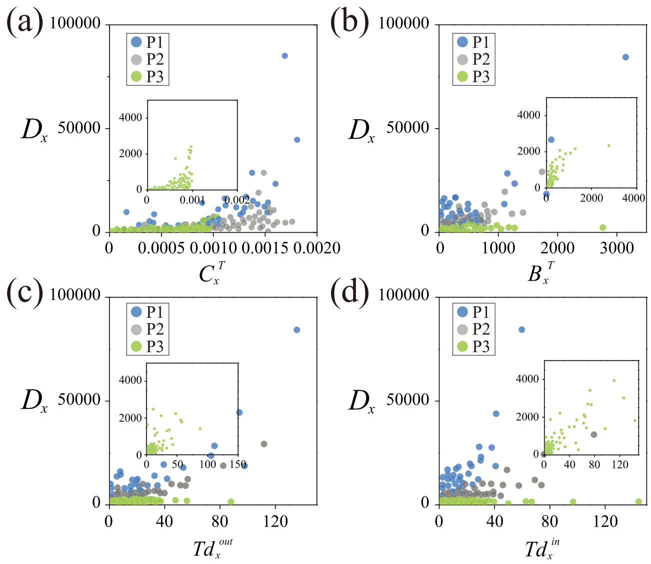


**FIG. 10.** The relationship between the overall delay and four temporal characteristics of airports (a) $C_x^T$, (b) $B_x^T$, (c) $Td_x^{out}$, and (d) $Td_x^{in}$ when all flights are delayed in the corresponding time period. $P_1$, $P_2$, and $P_3$ represent the time periods 6:00–12:00, 12:00–18:00, and 18:00–24:00, respectively; the insets are redrawings for 18:00–24:00.

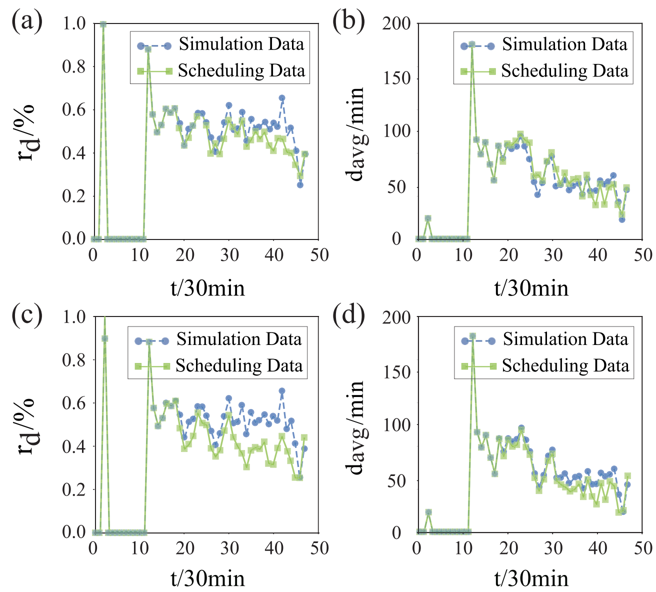


**FIG. 11.** Comparison of the best-fit model and the model with greedy strategy on (a) $r_d$, (b) $d_{avg}$, and the model with strategy based on improved genetic algorithm on (c) $r_d$, (d) $d_{avg}$.

temporal network characteristics have rarely been examined for their impact on flight delay, our motivation was to find potential undiscovered temporal network effect in scenarios of airport failure, e.g., extreme weather, etc. We can tell from the figure that the delay intensity of airports shows weak correlation with $B_x{}^T$, $Td_x{}^{out}$, and $Td_x{}^{in}$ [Figs. 10(b)–10(d)], indicating that the sharing resources (passengers, crew, etc.) from the former flight and the sharing of aircraft between different missions are all positively correlated with the spread intensity. The overall delay shows a more obvious exponential relationship with $C_x{}^T$ [Fig. 10(a)], indicating that the delay for airports with higher temporal closeness centrality has a faster diffusion rate.

## V. DELAY CONTROL

The delay propagation model then can be used to simulate the aviation system and to guide the design of controlling strategies for effectively reducing flight delay propagation. In the following, we discuss two controlling strategies: (1) A greedy strategy based on the delay spreading model and (2) an optimal strategy based on an improved genetic algorithm.

### A. Greedy strategy based on delay spreading model

When prioritizing flights, we assume that the status of other flights in the network remains unchanged and that all flights in each unit of time delay for the same length of time (the takeoff takes about 10 min for a single-runway airport and 5 min for two-runway airports). Then, we calculate the total delayed time that the flight may cause and used it as a measure of priority for flight scheduling. In order to eliminate the simulation error of the model itself on the actual operation data of the air traffic system and simultaneously to test the optimization effect of the greedy strategy, we compare the scheduling data generated by the control strategy of the delay cascade model with the simulation data generated by the scheduling process of the model with the same initial delay. Figures 11(a) and 11(b) show the control effect of initial delays on all aircraft on December 4, and the degree of optimization of the greedy strategy is measured by the distance between the scheduling data and the simulation data sequence. We can see that, with the delay spreading model-based greedy strategy, both $r_d$ and $d_{avg}$ are weakly optimized at the expense of huge computing resource consumption. On average, the ratio of delayed flights is reduced by only 1.09% in each 30 min.

### B. Strategy based on improved genetic algorithm (GA)

In this section, we develop an improved genetic algorithm that takes the objective function and the constraints as inputs and generates an optimized scheduling solution. Delayed flights bring substantial losses to airlines, airport operators, and passengers, most of which are unquantifiable. To this end, the aim of flight delay control is to consider $r_d$ and $d_{avg}$. The objective function is

$$\min \sum_{i}^{K} D_i, \tag{9}$$

where $K$ is the number of flights and $D_i$ denotes the delayed time of flight $i$. According to the flow constraint, the actual takeoff and landing traffic inside the airport must be less than a threshold of the planned capacity for the airport,

$$\begin{aligned} SAF_x(t) &> (1+\gamma)AAF_x(t), \\ SDF_x(t) &> (1+\gamma)ADF_x(t), \end{aligned} \tag{10}$$

in which $SAF_x$, $AAF_x$, $SDF_x$, and $ADF_x$ represent the number of scheduled arrival flights, actual arrival flights, scheduled departure flights, and actual departure flights at airport $x$, respectively, and $\gamma$ is the proportion of flow threshold and the number of scheduled flights. Former flights of consecutive flights (multiple flights operated on the same aircraft) must arrive prior to the departure of subsequent scheduled flights, namely,

$$AL_{i_k} < AD_{i_{k+1}}\ (i_k, i_{k+1} \in S_i), \tag{11}$$

where $S_i$ denotes the sequence of flights operated by the flight $i$, $AL_{i_k}$ is the actual arrival time of preceding flight $i_k$, and $AD_{i_{k+1}}$ is the actual departure time of the next flight $i_{k+1}$. In addition, the transfer of crew and passengers between different flights also needs to meet a certain time interval. We assume that the transfer of passengers only occurs when there is no direct flight between the two airports and that the temporal gap between the departure time of the subsequent flight and the landing time of the former flight is more than 1 h and less than 3 h, that is,

$$180 \geq PD_n - PL_i \geq 60\ (A_{in} \neq 1), \tag{12}$$

where $A_{in}$ is the adjacent matrix of the aviation network, $PD_n$ is the planned departure time of the next transfer flights, and $PL_i$ is the planned landing time of preceding flights. Finally, the flight itself has constraints

$$\begin{aligned} AD_i &> PD_i, \\ D_i &= AD_i - PD_i. \end{aligned} \tag{13}$$

In the algorithm, we set the delay time for each flight as a variable, all flight delays as a gene sequence, and the planned flight data and the initial delay for each aircraft as the set of inputs. We then calculate the earliest departure time and the latest departure time of each flight by the role of aircraft rotation so as to determine the range of each variable and reduce the search space of the set of solutions.

Because of the large scale and the large number of variables within the aviation system, we improve the GA via the following aspects to solve the optimization problem of large-scale scheduling:

(1) First, due to the large solution space and strong constraints, the number of feasible solutions in the process of population evolution is limited. To solve this problem, we need to repair the unfeasible solutions and convert them into feasible solutions to speed up the convergence. The straightforward operation is to postpone the departure time of flights, which is subjected to takeoff and landing flow control for the most recent time period.
(2) Second, in order to make the algorithm converge faster to satisfactory solution, we adopt the intergenerational reserved Elitism Genetic Algorithm. This approach selects the best individual from parents to ensure that individual fitness will not reduce.
(3) Finally, in order to avoid the impact of uniform crossover and mutation probability on the population evolution of individuals with different fitness, the adaptive crossover and mutation probability is used to speed up the search of satisfactory solutions.

By comparing the difference between the scheduling data and the simulation data, we show the performance of the improved genetic algorithm in reducing the extent of delay spreading, including average delayed time of flights and the number of delayed flights, in Figs. 11(c) and 11(d). We can see that although the improved GA strategy has a weak influence on reducing the total delayed time of flights (about 210.7 min in this period), it can reduce the average ratio of delayed flights to a large extent, by about 9.67% for each 30 min.

## VI. CONCLUSION

In this study, we first propose an agent-based data-driven model focusing on four factors, including aircraft rotation, flight connectivity, scheduling process, and disturbance, to create a simulator for reproducing the delay propagation in aviation networks. We then run the simulator not only to evaluate the effect of the four individual factors on delay propagation separately but also to analyze the synthesized model with combinations of different factors. Apart from this, we analyze the impact of initial delay on propagation intensity in different time periods and investigate the relationship between overall delay and temporal characteristics of airports. Finally, we discuss two rescheduling strategies to guide the design of controlling strategies for effectively reducing flight delay propagation.

Results show that the PD model can reproduce the delay propagation on days with moderate and serious delay very well, while the FCFSD model performs better on days with serious delay. When we investigate the impact of initial delay on propagation intensity in different time periods, we find that the effect of airport on flight delays is strongly dependent on the number of flights; in addition, the propagation intensity shows a weak positive correlation with ${Td_x}^{out}$, ${Td_x}^{in}$, and ${B_x}^T$, while the overall delay shows a more obvious exponential relationship with ${C_x}^T$. When delay occurs, the planned flight schedule may become unfeasible, and we discuss two controlling strategies for reducing flight delay propagation: a greedy strategy based on the delay spread model and a strategy based on an improved genetic algorithm. We see that, although the improved GA strategy has a little effect on reducing the total delayed time of flights, it can reduce the average ratio of delayed flights to a large extent.

With a special focus on the temporal pattern and a substantial effort in building a systematic model incorporating interactions of all flights and airports, we believe that this study is crucial for the understanding of delay propagation in airport networks, and it provides optimized rescheduling strategies for minimizing delay. In this work, we have considered as many factors as possible for delay propagation and delay absorbing, and we have taken different scheduling mechanisms into account in scenarios of airport congestion, providing an important supplement to existing models.

## ACKNOWLEDGMENTS

The authors would like to thank Professor Zhe Liang from the Joint Algorithm Laboratory of Tongji University & Xiamen Airlines for helpful discussions and suggestions in improving the study. S.Q. and J.M. are partially supported by the Natural Science Foundation of China (NSFC) (Grant Nos. 71771213 and 91846301). S.C. is supported by the Natural Science Foundation of China (Grant No. 71522014). X.L. acknowledges the Natural Science Foundation of China (Grant Nos. 71790615 and 91846301) and the Hunan Science and Technology Plan Project (Grant Nos. 2017RS3040 and 2018JJ1034).

## REFERENCES


[1] D. L. Bryan and M. E. O'kelly, "Hub-and-spoke networks in air transportation: An analytical review," J. Reg. Sci. **39**, 275–295 (1999).
[2] W.-B. Du, B.-Y. Liang, C. Hong, and O. Lordan, "Analysis of the Chinese provincial air transportation network," Phys. A Stat. Mech. Appl. **465**, 579–586 (2017).
[3] C. Hong, J. Zhang, X.-B. Cao, and W.-B. Du, "Structural properties of the Chinese air transportation multilayer network," Chaos Solitons Fractals **86**, 28–34 (2016).
[4] S. A. R. Khan, D. Qianli, W. SongBo, K. Zaman, and Y. Zhang, "Travel and tourism competitiveness index: The impact of air transportation, railways transportation, travel and transport services on international inbound and outbound tourism," J. Air Transp. Manag. **58**, 125–134 (2017).
[5] R. Guimera, S. Mossa, A. Turtschi, and L. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," Proc. Natl. Acad. Sci. U.S.A. **102**, 7794–7799 (2005).
[6] A. M. Churchill and D. J. Lovell, "Coordinated aviation network resource allocation under uncertainty," Transp. Res. E Logist. Transp. Rev. **48**, 19–33 (2012).
[7] L. Bengtsson, J. Gaudart, X. Lu, S. Moore, E. Wetter, K. Sallah, S. Rebaudet, and R. Piarroux, "Using mobile phone data to predict the spatial spread of cholera," Sci. Rep. **5**, 8923 (2015).
[8] C. Taylor and O. D. Weck, "Coupled vehicle design and network flow optimization for air transportation systems," J. Aircr. **44**, 1479–1486 (2007).
[9] D. Alexander, "Volcanic ash in the atmosphere and risks for civil aviation: A study in European crisis management," Int. J. Disast. Risk Sci. **4**, 9–19 (2013).
[10] W. Cong, M. Hu, B. Dong, Y. Wang, and C. Feng, "Empirical analysis of airport network and critical airports," Chin. J. Aeronaut. **29**, 512–519 (2016).
[11] R. Guimera and L. A. N. Amaral, "Modeling the world-wide airport network," Eur. Phys. J. B **38**, 381–385 (2004).
[12] W. Li and X. Cai, "Statistical analysis of airport network of China," Phys. Rev. E **69**, 046106 (2004).
[13] X. Sun, V. Gollnick, and S. Wandelt, "Robustness analysis metrics for worldwide airport network: A comprehensive study," Chin. J. Aeronaut. **30**, 500–512 (2017).
[14] X. Sun, S. Wandelt, and F. Linke, "Temporal evolution analysis of the European air transportation system: Air navigation route network and airport network," Transportmetrica B Transp. Dyn. **3**, 153–168 (2015).
[15] W. Wang and C. Wang, "Temporal-spatial network of flight schedule in hub airport: A case study of Beijing International Airport," Acta Geogr. Sin. **68**, 762–774 (2013).
[16] J. Yin, M. Hu, Y. Ma, K. Han, and D. Chen, "Spatial-temporal topology and performance analysis of airport taxi network," in *Transportation Research Board 97th Annual Meeting* (TRB, 2018), pp. 1–8.
[17] D. Baker, R. Merkert, and M. Kamruzzaman, "Regional aviation and economic growth: Cointegration and causality analysis in Australia," J. Transp. Geogr. **43**, 140–150 (2015).
[18] D. Brockmann and D. Helbing, "The hidden geometry of complex, network-driven contagion phenomena," Science **342**, 1337–1342 (2013).
[19] S. AhmadBeygi, A. Cohn, Y. Guan, and P. Belobaba, "Analysis of the potential for delay propagation in passenger airline networks," J. Air Transp. Manag. **14**, 221–236 (2008).
[20] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, "Data-driven modeling of systemic delay propagation under severe meteorological conditions," in *Tenth USA/Europe Air Traffic Management Research and Development Seminar* (ATM Seminar, 2013), pp. 1–9.
[21] P. Fleurquin, J. J. Ramasco, and V. M. Eguíluz, "Characterization of delay propagation in the US air-transportation network," Transp. J. **53**, 330–344 (2014).
[22] N. Pyrgiotis, K. M. Malone, and A. Odoni, "Modelling delay propagation within an airport network," Transp. Res. C Emerg. Technol. **27**, 60–75 (2013).
[23] P. T. Wang, L. A. Schaefer, and L. A. Wojcik, "Flight connections and their impacts on delay propagation," in *The 22nd Digital Avionics Systems Conference, 2003, DASC'03* (IEEE, 2003), Vol. 1, p. 5B.
[24] J.-T. Wong and S.-C. Tsai, "A survival model for flight delay propagation," J. Air Transp. Manag. **23**, 5–11 (2012).
[25] See http://www.caac.gov.cn for "Statistical Bulletin of Civil Aviation Industry Development in 2017."
[26] Y.-J. Liu and S. Ma, "Flight delay and delay propagation analysis based on Bayesian network," in *2008 International Symposium on Knowledge Acquisition and Modeling* (IEEE, 2008), pp. 318–322.
[27] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, "Systemic delay propagation in the US airport network," Sci. Rep. **3**, 1159 (2013).
[28] A. Abdelghany, G. Ekollu, R. Narasimhan, and K. Abdelghany, "A proactive crew recovery decision support tool for commercial airlines during irregular operations," Ann. Oper. Res. **127**, 309–331 (2004).
[29] L. Lettovskỳ, E. L. Johnson, and G. L. Nemhauser, "Airline crew recovery," Transp. Sci. **34**, 337–348 (2000).
[30] G. Yu and X. Qi, *Disruption Management: Framework, Models and Applications* (World Scientific, 2004).
[31] G. Stojković, F. Soumis, J. Desrosiers, and M. M. Solomon, "An optimization model for a real-time flight scheduling problem," Transp. Res. A Policy Pract. **36**, 779–788 (2002).
[32] S. Lan, J.-P. Clarke, and C. Barnhart, "Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions," Transp. Sci. **40**, 15–28 (2006).
[33] S. Ahmadbeygi, A. Cohn, and M. Lapp, "Decreasing airline delay propagation by re-allocating scheduled slack," IIE Trans. **42**, 478–489 (2010).
[34] M. S. Aktürk, A. Atamtürk, and S. Gürel, "Aircraft rescheduling with cruise speed control," Oper. Res. **62**, 829–845 (2014).
[35] D. Jianli, Y. Yuecheng, and W. Jiandong, "A model for predicting flight delay and delay propagation based on parallel cellular automata," in *2009 ISECS International Colloquium on Computing, Communication, Control, and Management* (IEEE, 2009), Vol. 1, pp. 70–73.
[36] Y. Ding, "Predicting flight delay based on multiple linear regression," in *IOP Conference Series: Earth and Environmental Science* (IOP Publishing, 2017), Vol. 81, p. 012198.
[37] N. Kafle and B. Zou, "Modeling flight delay propagation: A new analytical-econometric approach," Transp. Res. B Meth. **93**, 520–542 (2016).
[38] Y. Liu and S. Ma, "Modeling and estimating for flight delay propagation in a reduced flight chain based on a mixed learning method," in *2008 International Symposium on Knowledge Acquisition and Modeling* (IEEE, 2008), pp. 491–495.
[39] W. Wu and C.-L. Wu, "Enhanced delay propagation tree model with bayesian network for modelling flight delay propagation," Transp. Plan. Technol. **41**, 319–335 (2018).
[40] R. Yao, W. Jiandong, and X. Tao, "Prediction model and algorithm of flight delay propagation based on integrated consideration of critical flight resources," in *2009 ISECS International Colloquium on Computing, Communication, Control, and Management* (IEEE, 2009), Vol. 2, pp. 98–102.
[41] See http://www.variflight.com for more information about Chinese Aviation Networks datasets.
[42] J. Milan, "The flow management problem in air traffic control: A model of assigning priorities for landings at a congested airport," Transp. Plan. Technol. **20**, 131–162 (1997).
[43] P. R. Srivastava, "A strategic prioritization approach to airline scheduling during disruptions," Ph.D. thesis (The University of Texas at Austin, 2015).