The analysis of the influence of delay absorbing sequence on flight delay propagation

Yutong Wang1*, Jinfu Zhu1 and Chang Liu2

1College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 211106, China
2College of Mathematics and Statistics, Guangxi Normal University, Guilin, Guangxi, 541006, China
*m18551759938@163.com

Abstract: Due to the connection between the flights, flight delay will be spread to the downstream flights via the flight string. Therefore, we need to take some corresponding measures to mitigate flight delay propagation. Firstly, based on the sequences of flight delay absorption by buffer time, the paper analyzes the law of delay propagation among different nodes from three different perspectives. Then, based on the random forest algorithm, the factors affecting delay propagation are sorted by order of importance, and the key influencing factors of delay propagation are obtained. Finally, through a case study of an airline's flight operation data, some measures to mitigate flight delays are proposed from the perspective of adjusting buffer time.

1. Introduction
In recent years, the air transportation industry has developed rapidly. Flight delays appear quite frequently. Flight delays not only cause huge losses to the airlines, but also affect the normal operation of the airports and cause dissatisfaction of passengers. Therefore, the airlines often set a long period of time in the route path to ensure the normal execution of the flight schedule. For example, overnight parking at a certain airport or a longer ground turnaround time during off-peak hours are usually set. One of the main causes of flight delays is delay propagation. Therefore, it is of significant importance to explore the delay propagation’s law of different nodes on the flight string and to find out the key factors affecting the delay propagation, so as to control and mitigate the delay propagation.

At present, relevant scholars at home and abroad have conducted relevant research on flight delay propagation. Jianli Ding et al. [1] determined whether there is flight delay in the downstream airport caused by the initial delay airport based on the time Petri net theory. And a flight delay warning mechanism was established through the calculation of the flight delay time. Yujie Liu, Tao Xu [2-3] used Bayesian networks to analyze flight delay propagation. Weidong Cao et al. [4] established a continuous flight delay propagation model based on the improved Bayesian network to analyze flight delay propagation. PYRGIOTIS N et al. [5] established the flight delay propagation model and calculated the “chain effect” of delay on the other airports caused by the flights’ congestion of a single airport. DUNBAR M et al. [6] established an integrated model of aircraft route and aircrew to quantitatively analyze the impact of aircraft routes and aircrew on flight delay propagation. Weiwei Wu et al. [7] established a modified delay propagation tree model and added the influence of the airports’ random delay factors, and then the prediction accuracy of flight delay propagation was improved. Fan Wu [8]
studied the mechanism of flight delay propagation in the complex network of airports from the qualitative and quantitative levels. Weiwei Wu et al. [9] predicted the overall delay status of key airports based on the weighted Markov chain model, and analyzed the impact of airport random factors on the delay propagation of the flight string.

At present, most researches on flight delay propagation at home and abroad have certain limitations: 1. Most of the studies only consider the propagation laws among consecutive flights, but lack the research on the propagation laws among different nodes on the flight string. 2. The existing literature lacks the comparison of influence degrees of ground buffer time and flight buffer time on the delay absorption respectively, when studying the buffer time for flight delay absorption. Therefore, this paper analyzes the delay propagation laws among different nodes on the flight string, and explores the importance of ground buffer time and flight buffer time for delay absorption respectively, and then proposes corresponding measures to control and mitigate flight delay propagation.

2. Analysis of flight delay propagation process

2.1. Flight delay classification
Flight delays can usually be divided into two categories, one is independent delay, which refers to delays caused by weather, ground support, flow control, and mechanical support. Such delay is caused by the flight's own reasons, and has nothing to do with the flight route; the other type is propagated delay, which refers to the impact on downstream flights due to the late arrival of the aircraft or aircrew of upstream flight. Such delay is caused by the delay propagation of the upstream flight, which is related to the flight route.

Propagated delay is caused by the connecting resources between the flight with independent delay and its downstream flights (including aircraft, aircrew, passengers and airport resources, etc.). For example, the same aircraft flies multiple flights in one day. The independent delay of the initial departure flight may be spread to subsequent flights through the same aircraft. In this way, it will not only affect the normal departure and arrival of flights, but also affect the connection of passengers. In general, flight delays will increase in one day. In other words, the independent delays of upstream flights are small, which may result in greater delays of downstream flights on the same string. Therefore, when there is an independent delay in the flight, a reasonable setting of the buffer time can effectively reduce its propagating impact on the downstream flights.

2.2. Graphic analysis of flight delay propagation process
In order to clearly express the calculation method of the propagated delay, Figure 1 illustrates the relationship between the flight arrival and departure times in the flight plan database and the actual operation database.

![Figure 1. Delay propagation process between consecutive flights](image)
Table 1. Symbols and their meanings

| Symbol | Meaning                     | Symbol | Meaning                     | Symbol | Meaning                     |
|--------|-----------------------------|--------|-----------------------------|--------|-----------------------------|
| STD    | Schedule departure time     | SFT    | Schedule flight time        | PD     | Propagated delay            |
| STA    | Schedule arrival time       | MFT    | Minimum flight time         | F buffer | Flight buffer time         |
| ATD    | Actual departure time       | ATA    | Actual arrival time         | G buffer | Ground buffer time         |
| IAD    | Independent arrival delay   | TDD    | Total departure delay       | TAD    | Total arrival delay         |
| SGT    | Schedule ground turnaround time | MGT    | Minimum ground turnaround time | IDD    | Independent departure delay |

3. Calculation method of flight propagated delay

Ground buffer time and flight buffer time play an important role in the absorption and mitigation of flight delays. In the development of the flight plan, the buffer time is usually set at the airport ground phase and the air flight phase. Since the specific data of ground buffer time and flight buffer time cannot be accurately obtained, this paper uses the method \[^{[10]}\] proposed by Jetzki (2009) to define the buffer time of this paper, and uses the flight plan database and flight operation database to determine the buffer time of the ground phase and air flight phase.

In view of the uncertainty of the sequences of the buffer time for the propagated delay and independent delay absorption, this paper uses the idea \[^{[11]}\] proposed by Nabin Kafle (2016), and assumes that the buffer time absorbs the propagated delay from the upstream nodes proportionally, so the propagated delay of the node \(i\) is the sum of the propagated delays from the upstream nodes to the node \(i\). Therefore, the following three delay absorption sequences are considered:

1. Buffer time first absorbs independent delay, and then absorbs propagated delay
   - Case 1: buffer time is less than independent delay
     Since the buffer time is less than the independent delay, so the buffer time only absorbs partial independent delay, and can’t absorb propagated delay. The propagated delay at node \(i\) is equal to the total delay time of the previous node \(i-1\).
   - Case 2: buffer time is greater than or equal to independent delay
     Since the buffer time is greater than or equal to the independent delay, so the independent delay is completely absorbed by the buffer time. The propagated delay of the node \(i\) is equal to the total delay time.

2. Buffer time first absorbs propagated delay, and then absorbs independent delay
   - Case 1: buffer time is less than propagated delay
     Since the buffer time is less than propagated delay, so the buffer time only absorbs partial propagated delay, and can’t absorb independent delay. The propagated delay at node \(i\) is equal to the remaining delay of the total delay time of the previous node \(i-1\) after absorption by the buffer time.
   - Case 2: buffer time is greater than or equal to propagated delay
     Since buffer time is greater than or equal to propagated delay, the propagated delay is completely absorbed by the buffer time.

3. Proportional simultaneous absorption of propagated delay and independent delay
   Assuming that \(x\) represents the propagated delay at node \(i\) absorbed by the buffer time, the independent delay at node \(i\) absorbed by the buffer time \(B_{j-1,i}\) is \(B_{j-1,i} - x\). According to the proportional simultaneous absorption of propagated delay and independent delay by the buffer time, it comes to the following formula:

   \[
   x = \frac{B_{j-1,i} * D_{i-1}}{B_{j-1,i} + D_{i}} \tag{3.1}
   \]

In summary, the total propagated delay of the initial delay \(D_i\) to the downstream nodes \(T_{delay}\) is
\[ T_{delay} = \sum_{i=2}^{n} p_{i,i} = D_i \times \sum_{i=2}^{n} \left( \prod_{s=2}^{i} \rho_s \right) \]  

(3.2)

Where, \( p_{i,i} \) represents the propagated delay of the initial delay to the downstream node \( i \); \( D_i \) represents the flight delay at node \( i \); \( \rho_s \) represents the propagated delay probability at node \( m \) caused by the initial node.

\[ \rho_s = \begin{cases} 
\min\left(1 - \frac{D_s}{D_{s+1}}\right) & \text{Buffer time first absorbs independent delay, then absorbs propagated delay} \\
\max\left(1 - \frac{D_s}{D_{s+1}}\right) & \text{Buffer time first absorbs propagated delay, then absorbs independent delay} \\
\frac{D_s}{D_s + D_{s+1}} & \text{Buffer time proportionally and simultaneously absorbs propagated delay and independent delay}
\end{cases} \]

(3.3)

4. Case study

The data in this article comes from the actual flight delay data of China Eastern Airlines (company code MU) from January 1, 2014 to December 31, 2015. To verify the significant factors affecting the propagated delay, this article selects the flight data of a certain route with initial delay at its original airport—Shanghai Pudong International Airport from January 1, 2014 to December 31, 2015.

4.1. Data preprocessing

There are many factors that affect flight delay propagation, and the flight propagated delays are affected by a combination of multiple effects. Therefore, the influencing factors considered in this paper mainly include airport on-time rate, airport weather condition, airport wind speed, ground buffer time and flight buffer time.

In view that the buffer time of the selected route is mainly concentrated in the time period from 18:00 to 24:00. Therefore, the buffer time factors considered in this paper mainly aggregate and distribute the ground buffer time and flight buffer time of downstream flights respectively to each unit time group from 18:00 to 24:00. In other words, each ground/flight buffer time of downstream flights is grouped into one of the seven time groups, and then all the ground/flight buffer time in each time group is superimposed. The grouping of downstream flights’ ground buffer time is subject to the actual departure time of the aircraft, and the grouping of flight buffer time is subject to the actual arrival time of the aircraft.

There are many sources of weather data, such as the “Weather Post Website”. The data on the website is usually downloaded through the web crawler. The quantization of weather categories can be simply described as classification and rating according to the different severities of the weather. For example, different weather conditions (such as clouds, rain or haze) have different effects on flight operation. When weather conditions such as rainstorm, haze or thunderstorm occur at the airport, 1 is taken, and otherwise, 0 is taken. At the same time, rating is required according to the crosswind levels of the departure airport. Different levels of crosswind have different effects on airport flight operation. When the airport has a crosswind of level 6 or higher, 1 is taken, and otherwise, 0 is taken.

4.2. Random forest

Random Forest (RF) is the integration of multiple decision trees. It is an integrated learning method based on Bagging. Since the random forest has a large number of decision trees, the ranking result of the final feature importance can be obtained through the integration of the importance of the variables obtained from each decision tree. This result is more stable and more reliable than the result of a single decision tree. Therefore, this paper chooses the random forest method for feature selection.

The method used in this paper is based on OOB data. If the feature has little effect on the result of the decision tree after being disrupted on the OOB, that is, the difference we calculate is small, it
means that the feature is not important; if the disruption has a great impact on the result of the decision tree, that is, the difference we calculate is large, it indicates that this feature is important.

The ranking results of three different delay absorption sequences are shown in Figure 2-4. It can be seen from the ranking results of feature importance that the most significant impacts on flight propagated delays in three different situations are the flight buffer time either at 18 o'clock or 19 o'clock, and the third is the flight buffer time at 20 o'clock. Furthermore, it is concluded that the flight buffer time has a greater impact on flight delay propagation than the ground buffer time.

Figure 2. The ranking results of feature importance when buffer time first absorbs independent delay, then absorbs propagated delay

Figure 3. The ranking results of feature importance when buffer time first absorbs propagated delay, then absorbs independent delay

Figure 4. The ranking results of feature importance when buffer time proportionally and simultaneously absorbs propagated delay and independent delay

4.3. Delay control and adjustment

The planned buffer time setting plays an important role in absorbing and mitigating initial delay and propagated delay among various nodes on the flight string. Moreover, from the importance ranking results of the features affecting the flight propagated delay based on the random forest algorithm, it can be concluded that the flight buffer time has a greater impact on the propagated delay than the ground buffer time. Therefore, an appropriate increase in the planned flight buffer time can be regarded as a method to effectively suppress delay propagation.

However, if the flight buffer time is only increased during the air flight phase, the total buffer time on the flight string will be increased, which will reduce the efficiency of the aircrafts and aircrew, and will also incur additional costs for the airlines. Moreover, if the ground buffer time is too long, when the aircraft ground handling is completed ahead of time, the aircraft still has to wait in the parking position due to the planned departure time, thereby reducing the efficiency of the use of airport
resources. Therefore, this paper proposes to reduce the planned buffer time during the ground turnaround phase on the flight string by 10%, and to proportionally increase the reduced ground buffer time to the flight buffer time during the air flight phase. The total propagated delay before and after the adjustment is as shown in figure 5.

![Figure 5. Change in the total propagated delay before and after the buffer time adjustment](image)

It can be seen from the results of the total propagated delay variation before and after the buffer time adjustment that the buffer time adjustment has no obvious effect on the first case (buffer time first absorbs independent delay, then absorbs propagated delay), because independent delay is absorbed firstly, and the propagated delay is only absorbed when there is residual buffer time. Therefore, the adjustment of the buffer time has little effect on the total propagated delay in this case. For the second and third cases, total propagated delay is reduced by more than 400 minutes after the buffer time adjustment (red) compared to that before the adjustment (blue), which shows a more obvious effect. Moreover, in the second case (buffer time first absorbs propagated delay, then absorbs independent delay), the initial delay of 25 flights, which have propagated delay before the buffer time adjustment, is completely absorbed after the buffer time adjustment. Therefore, propagated delay no longer exists.

5. Conclusion

According to the law of delay propagation among different nodes on the flight string, this paper analyzes the propagated law of initial flight delay from three different delay absorption sequences, and ranks the importance of the features affecting the delay propagation based on random forest algorithm. The conclusion is obtained that the flight buffer time has a greater impact on the delay propagation than the ground buffer time. Finally, this paper proposes an adjustment plan to mitigate the impact of initial flight delay propagation, and uses the flight operation data of the domestic airlines to verify the adjustment. The data shows that the reasonable adjustment of the buffer time can effectively reduce the propagation of initial delay, which can alleviate the pressure of airports to a certain extent and reduce the costs of airlines.

In the follow-up study, with the expansion of the data scale, it will take into account the propagated impact of the independent delays generated by different nodes on the flight string of major airlines on the subsequent nodes, besides the propagated delays of the initial delays to the subsequent nodes. The follow-up study will also consider the influence of connecting resources of aircrew and passengers on the flight propagated delay, so as to propose a more effective adjustment program.

Acknowledgement

We would like to thank my tutor Jinfu Zhu and the upperclassman Yeqing Liu whose comments have helped us substantially improve the quality of the paper.
Fund project
Funded by Nanjing University of Aeronautics and Astronautics Graduate Innovation Base (Lab) Open Fund Project (kfjj20180723), Air Transportation Network Operation Safety Optimization and Management(71731001) supported by NSFC, Fundamental Research Funds for the Central Universities (NS2016067), Science Foundation of China(61603178)

References
[1] Ding, J., Chen, D., Xu, T. (2008) Construction of Flight Delay Chain Response Model Based on Time Petri Net. Journal of System Simulation , 20(14): 3888-3891.
[2] Liu, Y. (2008) Research on Flight Delay Spread Based on Bayesian Network. Computer Engineering and Applications, 44.
[3] Xu, T. Ding, J., Wang, J. (2009) Flight delay and sweep analysis model based on Bayesian network. Journal of System Simulation , 21(15): 4818-4822.
[4] Cao, W. (2009) Flight Delay Analysis Based on Improved Bayesian Network Structure Learning. Tianjin University.
[5] PYRGIOTIS, N., MALONE, K. M., ODONI, A. (2011) Modeling delay propagation within an airport network. Transportation Research Part C: Emerging Technologies, 27(2): 60-75.
[6] DUNBAR, M., GROYLANG, G., WU, C. L. (2012) Robust airline schedule planning: Minimizing propagated delay in an integrated routing and crewing framework. Transportation Science, 46(2): 204-216.
[7] Wu, W., Zheng, S. (2014) Research on Flight Delay Propagation Based on Modified Delayed Tree. Traffic information and safety, 32(4): 119-124.
[8] Wu, F. (2015) Research on Mechanism of Flight Delay in Airport Network. Nanjing University of Aeronautics and Astronautics.
[9] Wu, W., Meng, T., Zhang, H. (2016) Flight Plan Optimization Based on Airport Delay Prediction. Transportation System Engineering and Information, 16(6):189-195.
[10] Jetzki, M. (2009) The propagation of air transport delays in Europe. RWTH AACHEN UNIVERSITY.
[11] Kafle, N., Zou, B. (2016) Modeling flight delay propagation: A new analytical-econometric approach. Transportation Research Part B: Methodological, vol. 93, pp. 520–542.
[12] You, H. (2016) R language prediction actual combat. Electronic Industry Press, Beijing.
[13] Daniel T. L., Chantal D. L. (2017) Data mining and predictive analysis. Tsinghua University Press, Beijing.