



**HAL**  
open science

# Flow based spatio-temporal graph network model for predicting airport network delays

Zhihan Xu, Ziming Wang, Yanjun Wang, Daniel Delahaye

► **To cite this version:**

Zhihan Xu, Ziming Wang, Yanjun Wang, Daniel Delahaye. Flow based spatio-temporal graph network model for predicting airport network delays. International Conference on Research in Air Transportation, Eurocontrol FAA NTU, Jul 2024, Singapore, Singapore. hal-04636706

**HAL Id: hal-04636706**

**<https://enac.hal.science/hal-04636706>**

Submitted on 5 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Flow based spatio-temporal graph network model for predicting airport network delays

Zhihan Xu, Ziming Wang, Yanjun Wang  
 College of Civil Aviation  
 Nanjing University of Aeronautics and Astronautics  
 Nanjing, China  
 ywang@nuaa.edu.cn

Daniel Delahaye  
 ENAC Research Lab  
 Ecole Nationale de l'Aviation Civile  
 Toulouse 31400, France  
 daniel@recherche.enac.fr

**Abstract**—With the continued advancement of the air transport industry, ensuring the safety and efficiency of air traffic operations has become a major concern. Various efforts are being made to achieve this objective, among which the prediction of delays in air traffic is of significance. Current deep learning methodologies for predicting network wide air traffic delays typically rely on historical delay data as the primary input, often neglecting the inclusion of air traffic flow data. In this paper, the FSTGMAN model is developed to explore the efficacy of incorporating traffic flow data as an additional input for predictive modeling, contrasting its performance with a model that does not utilize flow data. The findings reveal that the incorporation of flow data marginally enhances the overall accuracy of the predictions. Furthermore, the performance of our model is compared with that of baseline models such as MLP, LSTM, Transformer, and Seq2Seq, demonstrating notable advantages.

**Keywords**—flight delay; graph neural networks ; flight delay prediction; airport network; machine learning

## I. INTRODUCTION

According to the International Air Transport Association (IATA) in 2020, there were a total of 46.1 million flights worldwide, transporting over 4.4 billion passengers. The International Civil Aviation Organization (ICAO) predicts in its 2019 Aviation Benefits Report that air traffic will more than double in the next 20 years. Despite the significant impact of the epidemic on the industry, the IATA's statistics show that traffic in November 2023 is at 99.1% of November 2019 levels. This data highlights the rapid development of air traffic and its increasing significance in our transportation system. However, despite the growth of the aviation industry, problems such as safety and flight delays persist. Flight delays are a global challenge that directly impacts operational efficiency and indirectly poses risks to operational safety and environmental protection. Therefore, accurate prediction of air traffic delays is a crucial area of research in the field of air transportation. For passengers, precise delay prediction can help make better flight choices, while for aircraft operators,

This research was supported by the National Natural Science Foundation of China (Grant Nos. U2033203, 61773203, 61903187)

delay information can help them prepare in advance and mitigate risks. However, due to the complex and diverse reasons behind delays, predicting flight delays is a complex task.

In recent years, with the emergence of machine learning, researchers have started to explore the use of machine learning methods to predict flight delays, with promising results. In particular, graph-based networks have shown superior performance in predicting network-wide flight delays. Graph-based networks can capture the graph structure formed by airports and flights. Several studies have shown the capability of improving the performance of graph-based networks by introducing additional information (e.g. meteorological data). Existing studies focused on understanding spatial-temporal dependencies, i.e., delay impacts between different time periods and surrounding nodes. However, fewer studies have focused on understanding the relationship between delays and traffic flow. The fact that air traffic is dynamic and is highly correlated with the network-wide flight plan should be considered. Therefore, we propose a network-wide delay prediction model that considers external factors such as weather, airport network structure, and flight plan data.

Currently, most existing models are based on experimental data from the United States. In the context of air traffic delay prediction in China, Zanin et al. [1] analyzed delay patterns in Chinese airports, focusing on the comparison of factors such as weather and traffic flow in predictions. By examining the correlation between traffic flow and delays during the high-traffic summer seasons over multiple years, the study reveals that traffic flow is not the main cause of delays. Further investigations using the random forest technique elucidated the influence of various factors on the accuracy of the predictions. In particular, traffic flow factors significantly affect the accuracy of the prediction at several airports, suggesting that traffic flow data retain their relevance in predicting delay. These insights offer valuable direction for our research on the incorporation of traffic factors into networks and adapting them for Chinese datasets.

The main contributions of this paper are represented as

follows:

- We propose a Flow based Spatial-Temporal Graph Multiple Attention Network (FSTGMAN) framework to predict the network-wide airport delays, which considers both historical delay data and flight plan data.
- The model uses flow-volume data that other models rarely consider.
- Our model is based on graph networks and attention mechanisms to capture spatial-temporal dependencies in the network-wide delays prediction task.
- A real-world dataset is built to evaluate the performance of the proposed model. Our model performs better for network-wide delay prediction than all baseline models.

## II. RELATED WORKS

There are two main types of methods to predict air traffic delay: Simulation-based or Operations Research-based approaches, and data-driven approaches [2].

### A. Simulation based or OR based approaches

Numerous methods have been proposed in the field of operations research to predict flight delays. These methods are basically based on modeling simulation and queuing theory. One such method is the National Airspace System Performance Analysis Capacity (NASPAC) developed by MITRE CAAS [3]. NASPAC is a discrete-time simulation model that calculates flight delays, identifies capacity constraints, and analyzes the adjusted system performance of new airports, runway closures, and new flights. Another example is LMINET, which is based on queuing theory [4]. Unlike machine learning methods, operations research methods employ explainable theoretical models for forecasting, whereas machine learning methods are considered a “black box” with no interpretability. Although machine learning methods can achieve higher prediction accuracy, operations research methods still hold practical value. Additionally, in recent years, complex network theory, particularly the propagation model, has been applied to research on air traffic delay. For instance, Baspinar et al. have established an airport delay and flight delay propagation model based on the infectious disease model [5].

### B. Data-driven approach

With the advancement of machine learning and data science, a growing body of research has focused on the use of big data and machine learning techniques to predict flight delays.

1) *Traditional machine learning*: Bayesian networks and random forests are commonly used in traditional machine learning models to predict air traffic delays [6]. A Bayesian network is a graphical representation of the probability relationship among a set of random variables. It is a Directed Acyclic Graph (DAG) consisting of nodes that represent variables and directed edges that connect these nodes. The nodes represent random variables,

and the directed edges depict the relationships between nodes, with the parent node influencing the child node. The strength of the relationship is expressed using conditional probability, while nodes without a parent node are expressed using prior probability. Bayesian networks offer a concise way to describe the internal behavior and interdependencies of subsystems, making them valuable for studying complex systems with multiple factors. In 2005, Xu et al. developed the earliest Bayesian airport network model [7]. This model focused on three airports: Chicago O’Hare International Airport (ORD), New York LaGuardia Airport (LGA), and Atlanta Jackson International Airport (ATL). It used flight operation data from the Federal Aviation Administration (FAA) of the United States from November 2003 to January 2004. Although this model only considered three airports and weather factors, it effectively demonstrated the propagation effect of delays, serving as an important study for subsequent research based on Bayesian networks.

In subsequent studies, the model has been gradually improved. Liu uses Bayesian networks to investigate flight delay prediction in terms of running time and accuracy. They proposed an improved Bayesian network model to improve the accuracy of prediction and running speed [8]. Wu et al. studied the importance of delay propagation models and concentrated on resource connections, including aircraft, crew, and passenger connections. They developed a delay propagation model by merging delay propagation trees with Bayesian networks [9]. Rodriguez-Sanza et al. focused on the operational details of individual nodes in the air traffic network. They took into account factors such as weather, flight and airport configuration, then established a Bayesian network arrival delay and congestion prediction model based on airport nodes [10]. Although Bayesian networks can effectively capture network propagation effects, constructing them requires a substantial amount of prior knowledge.

The Random Forest (RF) is an extension of the Bagging method, which combines multiple decision trees to make predictions. It is specifically designed for decision tree classifiers and has been shown to effectively avoid overfitting by converging the prediction error to the generalization error [11]. In a study conducted by Rebollo and Hamsa Balakrishnan in 2014, the random forest classification and regression algorithms were used to predict delays at key airport pairs in the National Airspace System (NAS) [12]. The study considered a predefined system delay state and incorporated spatial-temporal explanatory variables. The prediction results demonstrated that the accuracy of the two-hour prediction interval in classified delay prediction was 81%, and the error in continuous delay prediction (i.e., regression modeling) was 21 minutes. Random forests have the ability to automatically determine the importance of variables and exhibit low sensitivity to outliers in training data. Additionally, they can still perform well even when the number of variables exceeds the sample size.

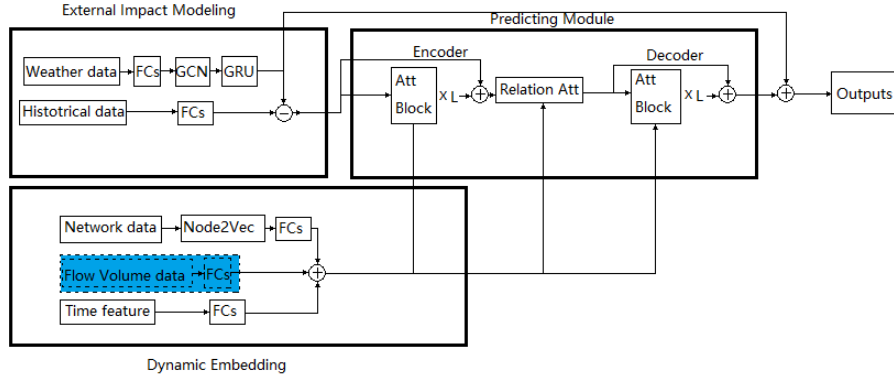


Figure 1. Structure of FSTGMAN.

2) *Deep learning methods*: Traditional machine learning models are known as “white box models” because their internal structure can be understood and explained. This allows for better control over the model’s details but also requires manual input and significant experience, making them less effective for processing large amounts of data. On the other hand, deep learning models are often referred to as “black box models” because they are more difficult to interpret. However, they are able to process massive amounts of data more effectively by combining low-level features to form abstract high-level representations of attributes or features. Deep learning is rooted in the concept of artificial neural networks, specifically multi-layer perceptrons with multiple hidden layers. Significant progress has been made in areas such as image recognition and speech recognition using deep learning. While simple deep learning models such as Artificial Neural Networks (ANNs) can handle various prediction problems, they are not well suited for processing time series data, such as air traffic delay data. To address this, more advanced models such as Long Short Term Memory (LSTM) [13], Recurrent Neural Networks (RNN) and Gated Recurrent Unit (GRU) [14] have been proposed and considered.

The above models lack the ability to effectively adjust the output length, which has led to the consideration of the Sequence-to-Sequence (seq2seq) model [15]. The seq2seq model consists of an encoder and a decoder. The encoder converts the input into a fixed length vector and passes it to the decoder. The decoder then receives the data and the flight delay information to be predicted, and decodes it to obtain the desired output length. Although the seq2seq model solves the issue of output length, it does not consider the physical relationships between airports in the airport system. To address this, Convolutional Neural Networks (CNNs) have been explored, as they can capture the propagation of flight delays through the network system [16]. Additionally, the Graph Convolutional Networks (GCNs) model has been proposed to

better represent air traffic models [17]. The studies based on graph-to-sequence normally use historical delay data, weather data and spatial-temporal information. Most of them did not take traffic flow data into account when predicting future delays. Since a major cause for flight delay is the imbalance between traffic demand and capacity. In fact, the demand for air traffic at the airport is influenced by the flight plan. Therefore, it is worth considering the use of flight flow data and exploring the potential differences it may make. This will be discussed in the upcoming sections.

### III. FLOW BASED SPATIAL-TEMPORAL GATED MULTI-ATTENTION GRAPH NETWORKS MODEL

This work focuses on the Flow Based Spatial Temporal Gated Multi Attention Graph Networks Model (FSTGMAN), which is an extension of the pure STGMAN model without flow input [18]. STGMAN is specifically designed to predict airport network delays. The STGMAN model is based on the GMAN model [19] and has the ability to accurately predict future airport network delays over longer time periods. For example, it uses data from past airport delays of 24 hours to predict future delays of 24 hours with high accuracy. Figure 1 provides the structure of the FSTGMAN structure in this study. The solid blocks represent the original STGMAN model, while the dashed blocks represent additional modifications that take into account flow data. The original model will be used as a reference for comparison with the modified model. The following sections will provide detailed explanations of the different modules in the FSTGMAN model.

#### A. External Impact Modeling

Flight delays are often influenced by meteorological factors, and historical data also includes the impact of these external factors. Therefore, it is important to preprocess the historical data by incorporating meteorological information before feeding it into

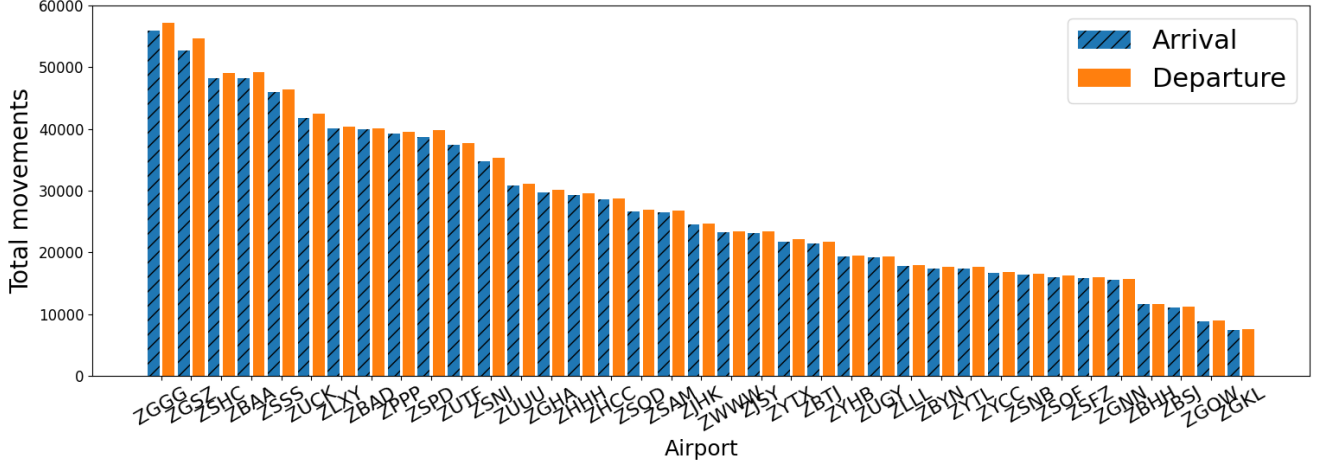


Figure 2. Total volume of flights in 37 airports.

the prediction module. To achieve this, the weather data is passed through multiple fully connected layers, followed by GCN and GRN networks, which extract spatial and temporal information separately. These extracted information are then subtracted from the historical data that has also been processed by multiple fully connected layers. As a result, the meteorological impact in the historical data is effectively eliminated. However, since the weather factors have been removed, they need to be reintroduced before obtaining the final output of the entire algorithm.

### B. Dynamic Embedding

The correlation between the delay trend and the temporal and spatial characteristics of the airport should be considered. Regarding the spatial attributes, the Node2Vec technique [20] is employed to transform each node (representing an airport) in the airport graph into a vector that encompasses information about the graph structure. Subsequently, this vector is processed through two fully connected layers to obtain a vector of a specified dimension. As for the temporal attributes, day-of-week and time-of-day information are encoded using one-hot encoding. This encoded information is then further processed using two fully connected layers. Finally, the aforementioned results are combined to generate an embedding that incorporates both the graph structure information and the time information.

Furthermore, to consider the effects of flow data, an extra input branch is incorporated. As illustrated by the dashed blocks in Figure 1, flow data pass through the fully connected layers before being combined with the previously mentioned embedding. This procedure can be understood in the following way: Traffic flow exhibits a high degree of regularity over time,

typically being sparse in the early morning and dense throughout the day, with corresponding temporal characteristics. Spatially, major airports typically experience greater traffic volumes, and the number of flights correlates with the geographical location of the airports. This aspect may also hold significant relevance. Experimental results have shown that incorporating traffic data into the dynamic embedding module enhances the model’s accuracy more effectively than integrating it into the external influence module.

### C. Prediction Module

This module utilizes multiple attention mechanisms to make predictions about flight delays. The module consists of an encoder, a decoder, and a relational attention layer that connects the two. Each encoder and decoder is made up of several Att Blocks internally. The Att block uses inter-node attention to handle spatial attention scores and inter-time attention to handle temporal attention scores. Here we have:

$$\alpha_{i,v}^{(k)} = \frac{\langle f_{s,1}^{(k)}(h_{i,t}^{(1-1)} \| e_{i,t}), f_{s,2}^{(k)}(h_{v,t}^{(1-1)} \| e_{v,t}) \rangle}{\sqrt{d}} \quad (1)$$

$$\tilde{\alpha}_{i,v}^{(k)} = \frac{\exp(\alpha_{i,v}^{(k)})}{\sum_{v' \in V} \exp(\alpha_{i,v'}^{(k)})} \quad (2)$$

$$\beta_{t,\tau}^{(k)} = \frac{\langle f_{t,1}^{(k)}(h_{i,t}^{(1-1)} \| e_{i,t}), f_{t,2}^{(k)}(h_{i,\tau}^{(1-1)} \| e_{i,\tau}) \rangle}{\sqrt{d}} \quad (3)$$

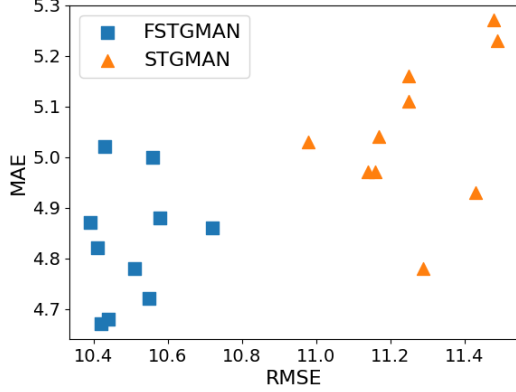


Figure 3. MAE and RMSE of the tests (in minutes).

$$\tilde{\beta}_{t,\tau}^{(k)} = \frac{\exp(\beta_{t,\tau}^{(k)})}{\sum_{\tau' \in N_t} \exp(\beta_{t,\tau'}^{(k)})} \quad (4)$$

where  $h_{i,t}^{(l-1)}$  is the hidden representations of nodes  $i$  at time step  $t$  in the  $(l-1)$ -th Att Block,  $e_{i,t}$  is the dynamic embedding representations of node  $i$  at time step  $t$ ,  $f_{s,1}^{(k)}, f_{s,2}^{(k)}, f_{t,1}^{(k)}, f_{t,2}^{(k)}$  denote four non-linear functions for the  $k$ -th head attention mechanism that produce  $d$ -dimensional output representations,  $\langle, \rangle$  is the inner-product operation,  $\parallel$  is the concatenation operation,  $\tilde{\alpha}_{i,v}^{(k)}$  is the attention score of the  $k$ -th head attention mechanism that represents the importance of node  $v$  to node  $i$ ,  $\tilde{\beta}_{t,\tau}^{(k)}$  is the attention score between time steps  $t$  and  $\tau$ , indicating the importance of time step  $\tau$  to time step  $t$ .  $N_t$  denotes a set of time steps before time step  $t$ , i.e., only information from time steps earlier than the target step is considered for causality.

Subsequently, the hidden state of node  $i$  at time step  $t$  is revised as follows:

$$hs_{i,t}^{(l)} = \parallel_{k=0}^{K-1} \left\{ \sum_{v \in V} \tilde{\alpha}_{i,v}^{(k)} \cdot f_{s,3}^{(k)}(h_{v,t}^{(l-1)}) \right\} \quad (5)$$

$$ht_{i,t}^{(l)} = \parallel_{k=0}^{K-1} \left\{ \sum_{\tau \in N_t} \tilde{\beta}_{t,\tau}^{(k)} \cdot f_{t,3}^{(k)}(h_{i,\tau}^{(l-1)}) \right\} \quad (6)$$

where  $f_{s,3}^{(k)}, f_{t,3}^{(k)}$  are two non-linear function,  $hs_{i,t}^{(l)}, ht_{i,t}^{(l)}$  are the output representations of the inter-node and inter-time attention mechanism in the  $l$ -th Att Block.

These results are then combined using a gate mechanism, which dynamically controls the flow of inter-node and inter-time correlations at each node and time step. The relational attention layer is responsible for managing the relationship between historical data and predicted results.

## IV. DATA

### A. Air traffic flow and flight delay data

We have obtained four months of flight data from air traffic management authority in China, specifically from March 2023 to June 2023. To ensure the reliability of the data, we selected data from 37 major airports, taking into account factors such as airport traffic volume and the availability of meteorological data. Figure 2 illustrates the total number of inbound and outbound flights at these airports during the selected time period. Notably, Guangzhou Baiyun Airport (ICAO code: ZGGG) is the busiest airport, with a total volume of over 55,000 inbound and outbound flights. This equals to an average daily throughput of more than 1,000 flights. To maintain the integrity of the airport network, we excluded flights that departed from or landed outside of these 37 airports. That is to say, we only considered delays and flight flow between these airports. Experimental results have shown that incorporating the complete airport flow, which includes flights departing and arriving at airports not within the selected network, into this model does not alter the prediction accuracy. To ensure accurate results, we also excluded abnormal data with excessively long delays (greater than 3 hours). After sorting the data, we obtained the delay and flow data for departure and arrival flights at each airport, consisting of a total of 2,640 hours from March 2 to June 20. Following the standard approach in deep learning, we divided this data into three parts for training, validation, and testing. The ratio of these three parts is 7:2:1.

### B. Weather data

Regarding meteorological data, we used the R language package Pmeter to retrieve historical METAR messages from each airport. Subsequently, we evaluated and assigned scores to the data, considering factors such as wind speed, dew point temperature, temperature, and unusual weather conditions specific to each airport. A higher score indicates poorer weather conditions at the airport. These meteorological data will be incorporated into the model as hourly meteorological scores for each airport.

## V. RESULTS

With the input of historical data of the last 24 hours, our prediction range is 24 hours ahead. In other words, the input delay  $x_{in} \in R^{24 \times 37 \times 2}$  (representing the historical arrival/departure delays of 37 airports over the past 24 hours) corresponds to the output  $x_{ou} \in R^{24 \times 37 \times 2}$  (forecasting the arrival/departure delays of 37 airports for the next 24 hours). A 24-hour forecast period is relatively extensive for delay prediction studies. On the one hand, this duration demonstrates the robust capabilities of our model. In contrast, such a long forecast period provides ample data for decision makers to assess trends and formulate long-term strategies. We performed 10 tests each on the original STGMAN model and the modified FSTGMAN model with flow data, and the performance is shown in Figure 3. Among them,

TABLE I. PERFORMANCE OF STGMAM (IN MINUTES)

Model	STGMAN			FSTGMAN		
	Flights	Arr.	Dep.	Total	Arr.	Dep.
MAE	5.77	3.73	4.78	5.66	3.61	4.67
RMSE	13.56	8.17	11.29	12.23	8.02	10.42

TABLE II. HYPERPARAMETERS IN FSTGMAN

Hyperparameters	Description	Value
$L$	The number of Att Blocks	1
$K$	The number of attention heads	8
Alpha	The learning rate	0.01
Optimizer	The selected optimizer	Adam
$D$	The dims of embedding size	64
$d$	The dims of each attention outputs	8

the orange dots represent the original model, and the blue dots represent the modified model with flow data, which can clearly show that the MAE value of the modified model is generally better than that of the original model. Additionally, the RMSE value of the enhanced model is significantly reduced compared to the original, suggesting a more consistent performance from the FSTGMAN model. The detailed data on the optimal performance of the two models are shown in Table I., and Table II. presents the hyperparameters used in the both FSTGMAN and STGMAN model (they are the same).

Table I. on the left side displays the outcomes of future delay predictions utilizing the original STGMAN model. It is evident that the prediction accuracy for arrival flight delays is substantially lower compared to departure flight delays, potentially due to the complex nature of inbound flight delay data. The table indicates a total mean absolute error (MAE) of 4.78, deemed to be a satisfactory outcome. Nonetheless, it is important to note that this result might be influenced by the lack of significant long-term variations in the aviation network during the analyzed period. As discussed in the paper [1], flight delays in China typically undergo notable fluctuations during the summer months (around July). Regrettably, the dataset employed in this study does not adequately encompass this timeframe to capture pertinent data.

To the right of Table I., the FSTGMAN model's predictions for future delays are displayed. This model demonstrates improved precision, as evidenced by its RMSE (10.42) and MAE (4.67) scores. Traffic flow data, in comparison to delay data, exhibits more consistency and smaller variations across different days. Therefore, temporally, it acts as the weight for each time period, enhancing the model's ability to assimilate the time-specific traits. Conversely, during early morning hours with fewer flights, the limited data samples might lead to notable delay fluctuations due to changes in traffic flow. Such variability could deteriorate the model's performance by increasing randomness. Introducing

traffic data at this juncture could mitigate these fluctuations. Spatially, this approach also facilitates understanding the distinct features of each airport. Furthermore, by leveraging attention mechanisms, it is anticipated that flow data will assist in discerning the spread of air traffic across the entire airport network. Once assimilated, the model could utilize this data to estimate delays stemming from traffic overloads. Nonetheless, the actual structure of the model might require further specialization to fully realize this capability. Additionally, the current scope of the airport network discussed in this study does not encompass the majority of traffic and fails to include numerous aircraft movements at smaller and medium-sized airports, potentially limiting the anticipated outcomes.

To further examine the performance of the model, we illustrate the average MAE of 24 hours and 37 airports, as depicted in figures 5 and 4. Each prediction cycle during the model's evaluation outputs the predicted delay for 37 airports over a 24-hour period. Given the variable traffic flow at different times of the day, it is crucial to assess the model's performance throughout the day. Figure 4 presents a comparison of the delay predictions between the original and modified models over a full day, where the blue and orange lines indicate the results of the modified model, and the green and red lines indicate those of the original model. It is evident that the new model shows significant improvements primarily in the early morning hours compared to the original STGMAN model. However, during the day, the performances of both models are nearly identical. Enhancements in arrival delay accuracy are observed primarily at 2 AM and between 4-7 AM, while enhancements in departure delay are noted at 1 AM and between 3-4 AM, all during times with fewer flights of the respective type.

The FSTGMAN model partially addresses the accuracy shortcomings of the original model during early morning periods, enhancing the prediction accuracy by almost five minutes. Figure 5 displays the MAE values for both the original and the updated model at various airports, with points indicating the MAE of the original model and bars for the modified model. Variations in arrival delay errors are kept under one minute, whereas departure delay errors remain largely the same. There are minimal differences between the old and new models in this aspect. Random initialization of model parameters might lead to suboptimal initial values, occasionally preventing the model from achieving peak performance at certain times or airports. Overall, the FSTGMAN model has notably improved performance in areas where the original model was lacking, particularly by improving accuracy during early morning hours, thus offering greater utility to professionals and pilots active at these times.

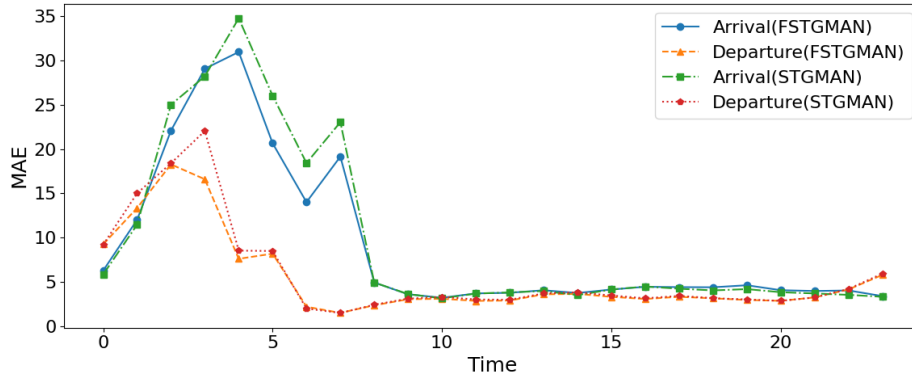


Figure 4. MAE over a day in FSTGMAN (in minutes).

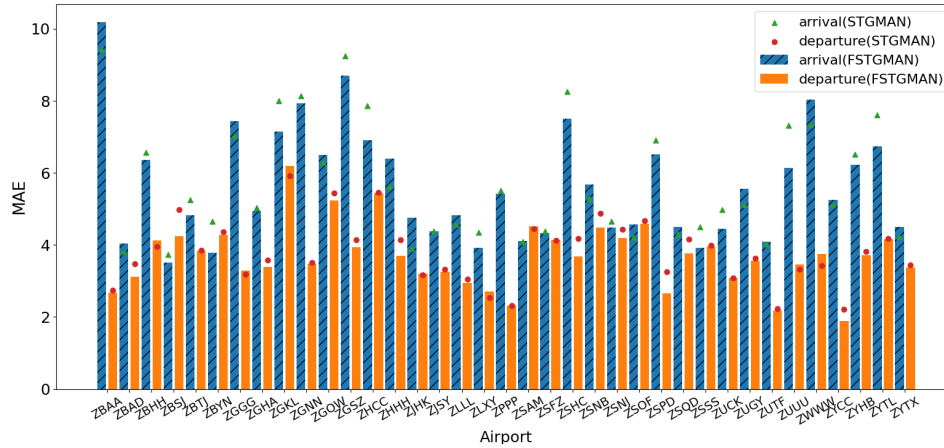


Figure 5. MAE of the 37 airports (in minutes).

TABLE III. PERFORMANCE OF OTHER MODELS USING DELAY AND FLOW

Model	MLP		Seq2Seq		LSTM		Transformer		STGMAN		FSTGMAN	
	arrival	departure	arrival	departure	arrival	departure	arrival	departure	arrival	departure	arrival	departure
MAE	13.79	9.66	11.69	8.39	11.84	8.41	11.68	8.22	5.77	3.73	5.66	3.61
RMSE	23.41	17.37	20.16	15.29	20.33	15.34	20.10	15.08	13.56	8.17	12.23	8.02

A. Comparison with other models

In this section, we compare the performance of delay prediction with other commonly used models for predicting time series. These models include multilayer perceptron (MLP), Long-Short Term Memory Networks (LSTM), Seq2Seq and Transformer. The input of these models is the historical delay data of 24 hours, and the output is the prediction of delay of future 24 hours, same as FSTGMAN. MLP is a fully connected neural network with a forward structure. LSTM, a special type of recurrent neural

network (RNN), is capable of learning long-term dependencies by using gates to add or remove information from the cell state. Unlike LSTM, MLP does not incorporate any additional state information from previous time steps. The Seq2Seq model is an encoder-decoder architecture that can be applied to various tasks. In this paper, the Seq2Seq model consists of an encoder and decoder, each with three LSTM layers. Transformer, on the other hand, utilizes attention mechanisms to capture information in time series data and is the only method that employs feed-forward neural networks. The Transformer model retains loca-



TABLE IV. HYPERPARAMETERS IN OTHER MODELS

Model	Hyperparameter	Description	Value
MLP	$n$	the number of hidden layers	2
LSTM	$n$	number of hidden layers	3
	$H$	the hidden size	128
Seq2Seq	Optimizer	the selected optimizer	Adam
	Alpha	the learning rate	0.001
	$H$	the hidden size	64
Transformer	$d$	the dims of embedding size	64
	$D_{ff}$	the dims of Feed Forward size	128
	$L$	the num of Att Blocks	1
	$K$	the number of attention heads	2
	Alpha	the learning rate	0.0001
	Optimizer	the selected optimizer	Adam

tion information in the input through position encoding. When testing the prediction using flow data in the Transformer model, we simply modify the encoder input to incorporate flow data. Table IV. provides the hyperparameters for the baseline models, and the corresponding results are presented in Table III..

From the results we can see that the performance of FSTGMAN and STGMAN is much better than the baseline models. With delay information, spatial information, weather information, and aircraft flow volume information being considered, it is natural that the model can achieve great result. While the enhancements with input of flow data may appear insignificant in comparison to the substantial advancements made by the original STGMAN, even small improvements are valuable. However, as our method of incorporating flow data is rather straightforward, there is room to further refine the network structure to identify an improved approach for integrating flow data.

## VI. CONCLUSIONS

In this work, we use delay and flow data from April to June 2023 in China to predict airport network delays. By adding the processed flow data into the dynamic embedding module, we successfully incorporate flow data into the model. The results demonstrate that the predictive accuracy of our method surpasses that of the original model that does not incorporate flow input, and is notably superior to other baseline models. Analysis of the prediction results reveals enhancements that occur primarily in the early morning, with a more stable prediction outcome. Nonetheless, significant predictive errors persist during the early morning hours, suggesting that further steps could be implemented to even out the Mean Absolute Error (MAE) across different airports and time intervals. Given the basic nature of our current study, further investigation into a more appropriate framework is warranted in future research endeavors.

## ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (Grant Nos. U2033203, 61773203).

## REFERENCES

- [1] Massimiliano Zanin, Yanbo Zhu, Ran Yan, Peiji Dong, Xiaoqian Sun, and Sebastian Wandelt. Characterization and prediction of air transport delays in china. *Applied Sciences*, 10(18), 2020.
- [2] Chunzheng WANG, Minghua HU, Lei YANG, and Zheng ZHAO. Review on air traffic delay prediction. *Systems Engineering and Electronics*, 44(3):863, 2022.
- [3] Frolow I. and Sinnott J.H. National airspace system demand and capacity modeling. *Proceedings of the IEEE*, 77(11):1618–1624, 1 1989.
- [4] D. LONG, D. LEE, J. JOHNSON, E Gaier, and P Kostiuk. Modeling air traffic management technologies with a queuing network model of the national airspace system. Technical Report NASACR-1999-208988, The National Aeronautics and Space Administration, 1999.
- [5] B. Baspinar and E. Koyuncu. A data-driven air transportation delay propagation model using epidemic process models. *International journal of aerospace engineering*, 2016(Pt.2), 1 2016.
- [6] Leonardo Carvalho, Alice Sternberg, Leandro Maia Goncalves, Ana Beatriz Cruz, Jorge A. Soares, Diego Brandao, Diego Carvalho, and Eduardo Ogasawara. On the relevance of data science for flight delay research: a systematic review. *Transport Reviews*, 41(4):499–528, 1 2021.
- [7] Ning Xu, Kathryn B. Laskey, Chun-Hung Chen, Shannon Williams, and Lance Sherry. Bayesian network analysis of flight delays. In *Transportation Research Record*, 2007.
- [8] LIU Y J. The estimation of flight delay and propagation based on bayesian networks. mastersthesis, Tianjin University, 2009.
- [9] Cheng-Lung Wu and Kristie Law. Modelling the delay propagation effects of multiple resource connections in an airline network using a bayesian network model. *Transportation research, Part E. Logistics and transportation review*, 122(Feb.):62–77, 1 2019.
- [10] Alvaro Rodriguez-Sanza, Fernando Gomez Comendador, Rosa Arnaldo Valdes, Javier Perez-Castan, Rock Barragan Montes, and Sergio Camara Serrano. Assessment of airport arrival congestion and delay: Prediction and reliability. *Transportation research, Part C. Emerging technologies*, 98(Jan.):255–283, 1 2019.
- [11] LEO BREIMAN. Random forests. *Machine learning*, 45(1):5–32, 1 2001.
- [12] Juan Jose Rebollo and Hamsa Balakrishnan. Characterization and prediction of air traffic delays. *Transportation Research Part C: Emerging Technologies*, 44:231–241, 2014.
- [13] Zhiyong Cui, Ruimin Ke, Ziyuan Pu, and Yin Hai Wang. Stacked bidirectional and unidirectional lstm recurrent neural network for forecasting network-wide traffic state with missing values. *Transportation Research Part C: Emerging Technologies*, 118:102674, 2020.
- [14] Yanting Yin, Yajing Wu, Xuebing Yang, Wensheng Zhang, and Xiaojie Yuan. Se-gru: Structure embedded gated recurrent unit neural networks for temporal link prediction. *IEEE Transactions on Network Science and Engineering*, 9(4):2495–2509, 2022.
- [15] Xinglei Wang, Xuefeng Guan, Jun Cao, Na Zhang, and Huayi Wu. Forecast network-wide traffic states for multiple steps ahead: A deep learning approach considering dynamic non-local spatial correlation and non-stationary temporal dependency. *Transportation Research Part C: Emerging Technologies*, 119:102763, 2020.
- [16] Bowen Du, Xiao Hu, Leilei Sun, Junming Liu, Yanan Qiao, and Weifeng Lv. Traffic demand prediction based on dynamic transition convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, 22(2):1237–1247, 2021.
- [17] Yong Chen and Xiqun (Michael) Chen. A novel reinforced dynamic graph convolutional network model with data imputation for network-wide traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 143:103820, 2022.
- [18] Hongfeng Zheng, Ziming Wang, Chuanpan Zheng, Yanjun Wang, Xiaoliang Fan, Wei Cong, and Minghua Hu. A graph multi-attention network for predicting airport delays. *Transportation Research Part E: Logistics and Transportation Review*, 181:103375, 2024.
- [19] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. Gman: A graph multi-attention network for traffic prediction, 2019.
- [20] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks, 2016.