

Machine Learning Aided Air Traffic Flow Analysis Based on Aviation Big Data

Guan Gui, *Senior Member, IEEE*, Ziqi Zhou, *Student Member, IEEE*, Juan Wang, *Student Member, IEEE*, Fan Liu, *Student Member, IEEE*, and Jinlong Sun, *Member, IEEE*

Abstract—Timely and efficient air traffic flow management (ATFM) is a key issue in future dense air traffic. The emerging demands for unmanned aerial vehicles and general aviation aircraft aggravate the burden of the ATFM. Thanks to the advanced automatic dependent surveillance-broadcast (ADS-B) technique, the aerial vehicles can be tracked and monitored in a real-time and accurate manner, providing possibility for establishing a more intelligent ATFM architecture. In this paper, we first form an aviation big data platform by using the distributed ADS-B ground stations and the obtained ADS-B messages. By exploring the constructed dataset and mapping the extracted information to the routes, the air traffic flow between different cities can be counted and predicted, where the prediction task is implemented on the basis of two machine learning methods, respectively. The experimental results based on real-world data demonstrate that the proposed traffic flow prediction model adopting long short-term memory (LSTM) can achieve better performance, especially when abnormal factors in traffic control are considered.

Index Terms—Air traffic flow management, automatic dependent surveillance-broadcast (ADS-B), aviation big data, machine learning, flight route.

I. INTRODUCTION

DUE to the unprecedented development of the civil aviation industry, efficient and comfortable air travel has become the priority choice for more and more passengers. The rising demand for air travel and the accompanying rapid growth of flights will inevitably lead to the air traffic congestion in limited air space, which will burden the work of air traffic management (ATM), and pose formidable challenges to air traffic surveillance systems. Therefore, it is urgent for establishing the more intelligent and high-performance air traffic surveillance system, where real-time and high-precision air traffic flow management (ATFM) is a key issue [1]. Traditional air traffic surveillance techniques such as primary surveillance radar (PSR) and secondary surveillance radar (SSR) need to be evolved, as their relatively low tracking accuracy cannot satisfy the future crowded space [2]. The emergence of automatic dependent surveillance-broadcast (ADS-B) has brought

tremendous changes to traditional radar-based surveillance technologies [3], [4]. The ADS-B has been considered as a crucial part of the next generation of air traffic surveillance system. The ADS-B technology can be used in many fields such as target recognition [5], multilateration (MLAT), and ATFM (as this study focuses on) [6]. The main work of ATFM is to assist the air control department to make timely and appropriate evaluations whether the air traffic is close to the upper limit, so as to ensure the rationality of the air traffic and to maximize the utilization of the airspace capacity [7].

As the prerequisite for subsequent processing, data analysis and visualizations, reliable and accurate air route flow information plays a crucial role in ATFM system [1]. Until now, most air-traffic controllers and relevant departments still rely on traditional technologies like radar, which is more expensive and less accurate in locating aircrafts. Besides, the two-dimensional radar screens cannot characterize the air route flow information quite well. The accuracy and comprehensiveness of the ADS-B message make it a high-quality data source for air route flow statistics. By using the millions of ADS-B data continuously uploading from our aviation big data platform, we can carry on the statistics of the air route flow in arbitrary period between arbitrary two airports as well as some visualization tasks. Furthermore, precise flow prediction model can help the air traffic control departments to understand the possible trend of the flow in the near future. Thus, the surveillance system can provide support for more reasonable scheduling strategies. The work of flow prediction can be implemented by using historical flow data obtained from previous statistic tasks. Conventional work concerning prediction methods for time series analysis includes Kalman filter-based method [8], auto-regressive moving average (ARMA) method [9], auto-regressive integrated moving average (ARIMA) method [10], etc. However, the conventional methods show their weakness in capturing complex nonlinear time-dependent features, especially in dealing with big volume time series data covering high-dimensional features.

Nowadays, machine learning methods have been applied in time series analysis as its great power in searching complex structures and matching inner relationships between objects. A multilayer perceptron neural network was applied for wind speed prediction for intelligent wind power generation [11]. For example, lots of machine learning based methods were utilized to heterogeneous network traffic control [12]–[15], radio resource assignment techniques [16]–[19], and physical-layer wireless techniques [20]–[22]. A model combining stacked auto-encoders and neural networks was employed in passenger

This work was supported by the Project Funded by the National Science and Technology Major Project of the Ministry of Science and Technology of China under Grant TC190A3WZ-2, National Natural Science Foundation of China under Grant 61901228, the Jiangsu Specially Appointed Professor under Grant RK002STP16001, Innovation and Entrepreneurship of Jiangsu High-level Talent under Grants CZ0010617002 and CZ002SC19001, Summit of the Six Top Talents Program of Jiangsu under Grant XYDXX-010, 1311 Talent Plan of Nanjing University of Posts and Telecommunications. (Corresponding author: Jinlong Sun)

The authors are with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mails: {guiguan, 1218012005, 1219012920, 1018010402, sunjinlong}@njupt.edu.cn).

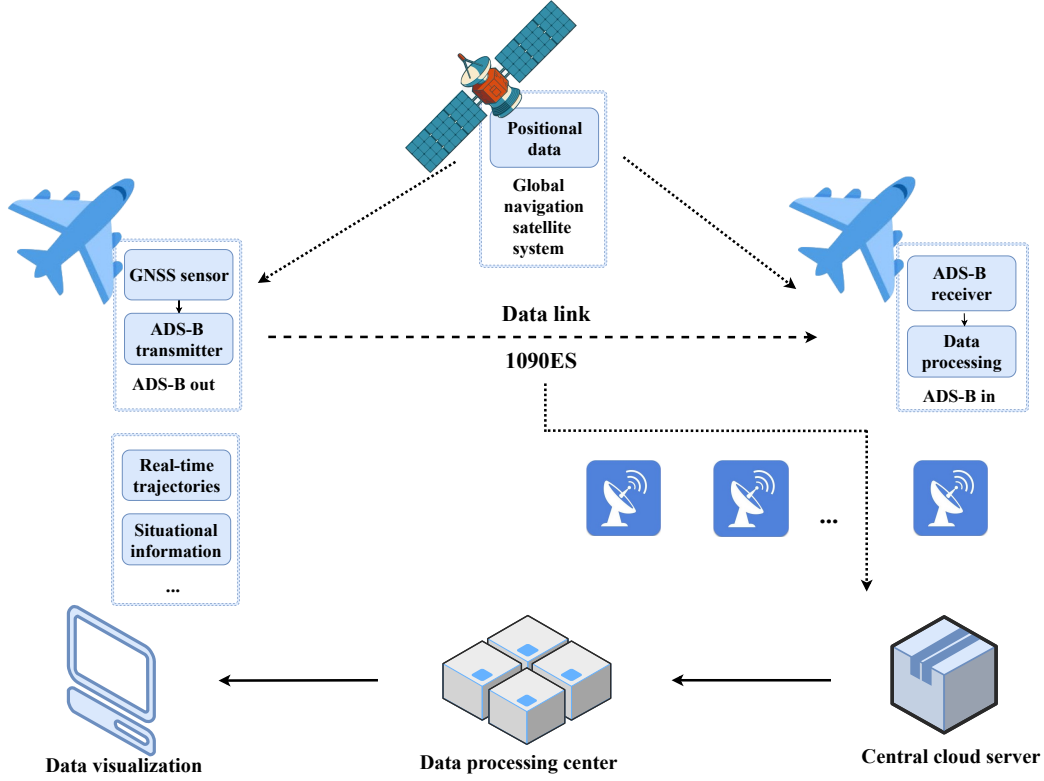


Fig. 1. Aviation big data platform based on the ADS-B data links and ground facilities.

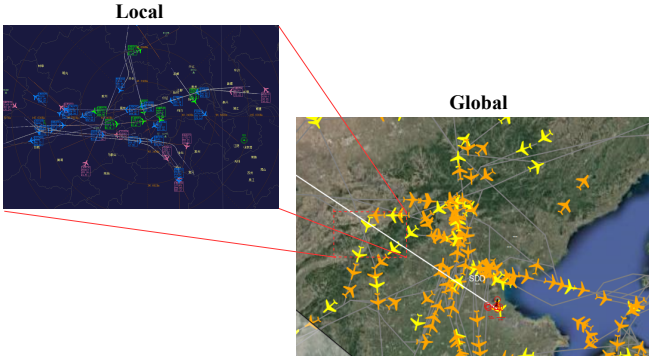


Fig. 2. Real-time visualization of the aircrafts based on the ADS-B messages.

flow prediction [23]. This paper explores the researches on statistics and prediction of the air route traffic flow based on big volume real ADS-B message. The flow prediction task can be implemented based on two machine learning methods: support vector regression (SVR) [24] and long short-term memory (LSTM) [25]. We find that both the SVR and LSTM-based prediction models can adequately predict the air route flow, and better performance can be obtained by the LSTM-based model using big volume dataset. The main contributions of this paper include the following points.

- The geometrical model for air traffic flow statistics is constructed, which can be used for statistic and visualization

tasks according to different granularity.

- To facilitate the follow-up applications of air traffic flow management, two machine learning-based methods are used and compared in predicting the future air traffic flow information.
- The proposed model and algorithm are validated by using real world ADS-B messages, and credible evaluation results are obtained.

II. SYSTEM ARCHITECTURE

A. ADS-B Based Aviation Big Data Platform

Over the last decade, the number of aircrafts boosts in limited airspace, which has inevitably caused the problem of air congestion. Thus, the requirement of ensuring the safety and efficiency of air transportation has brought new challenges to the ATM [26], [27]. Aircraft collision may occur in crowded airspace, which poses serious safety concerns to millions of air passengers. The next generation of ATM system equipped with ADS-B platform can handle the air traffic congestion problem more efficiently, and thus can provide safer air traffic networks [28], [29].

The ADS-B technology is an aircraft operation surveillance technology based on ground-air and air-air communication data links and the global navigation satellite system (GNSS). The overall ADS-B based surveillance architecture can be divided into two subsystems: ADS-B OUT and ADS-B IN. The ADS-B OUT subsystem continuously and automatically broadcasts ADS-B messages consisting of positional data provided by

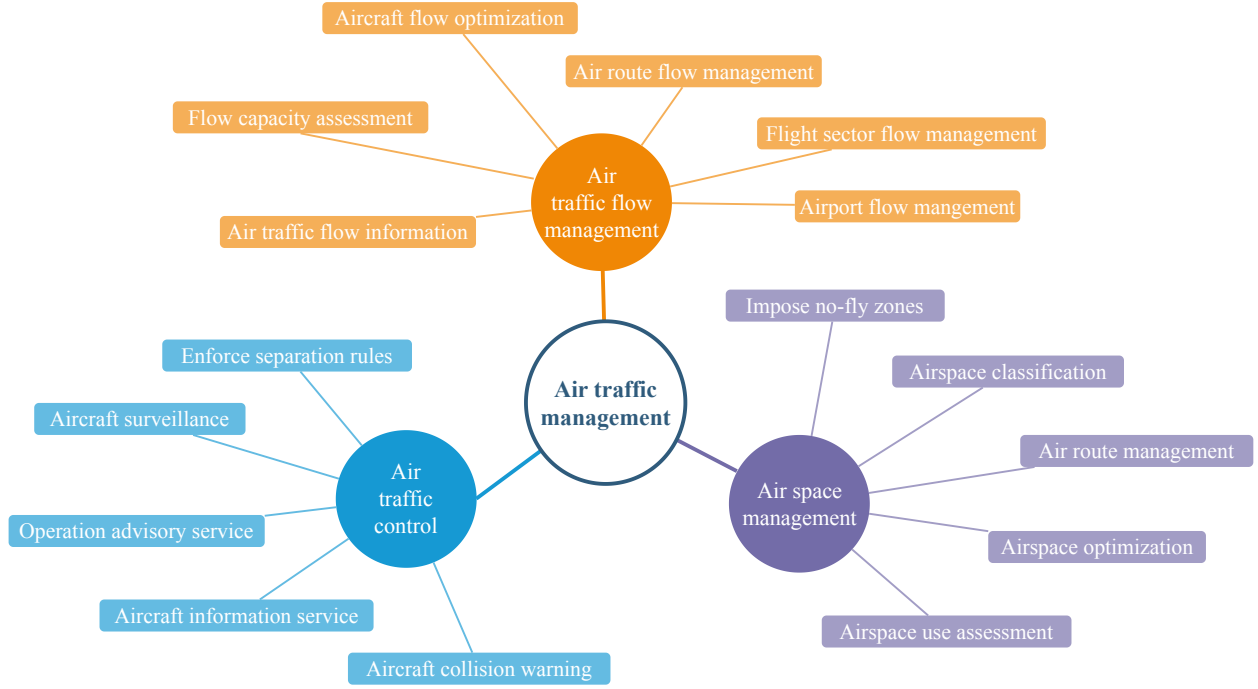


Fig. 3. The branches of the ATM.

the GNSS and other basic information (e.g. ICAO number) of each flight. While the ADS-B IN subsystem provides operation support and enhanced situational awareness, such as conflict warning information, collision avoidance strategy, and meteorological information [30], [31].

Fig. 1 shows the graphical illustration of our practical aviation big data platform based on the ADS-B system. As shown in Fig. 1, the positional information obtained from the GNSS and other situational information of each aircraft can be broadcasted by the ADS-B OUT transmitters. The surveillance stations deployed on the ground and aircrafts equipped with the ADS-B IN subsystems can receive these ADS-B messages over the 1090 ES data link [28]. The ground surveillance center consists of three parts: the central cloud server for storing ADS-B data, the data processing center for data mining, and the data visualization devices for characterizing the obtained information.

As it is shown in Fig. 2, visualization tasks can be accomplished after a series of processes in our data processing center. The visualization includes maps, trajectories of the aircrafts, and real-time flight information. The utilized aviation big data platform based on the ADS-B system can play an important role in establishing a more advanced ATM system.

B. Key Issues in ATM and ATFM

The main aim of the ATM is to make full use of the existing airspace and routes, and to ensure the safety and efficiency of the flights. Thus, the ATM consists of three main tasks: air traffic control (ATC), air space management (ASM), and ATFM [32]. The three branches of the ATM are shown in Fig. 3.

ATC: It is defined as a service provided by air traffic control department. The service includes flight operation advisory

service, aircraft information service, and warning service (e.g. aircraft collision warning). Additionally, aircraft surveillance and enforcement of relevant rules such as separation rules are also within the scope of the ATC.

ASM: It mainly manages and optimizes the given airspace conditions in order to maximize the use of the airspace. The work of ASM contains airspace classification, airspace use assessment, airspace optimization, and no-fly zones setting [33]. The narrower separation standard brought by more accurate localization and tracking technique can improve the capacity of the limited airspace.

ATFM: It is defined as the regulation of the amount of aircrafts in certain airspace. Because of the strict separation distance requirement between two arbitrary aircrafts, the reasonable supervision and regulation is the guarantee of efficient utilization of the available airspace. The tasks of the ATFM include air traffic flow information service, certain airspace (such as air route and airport sector) management, and air traffic flow optimization [34].

The original form of the ATM is based on procedural control system in which the aircrafts must strictly follow the flight schedules. And with the introduction of radar technology (i.e., PSR and SSR), it evolved to a radar-control based system that immensely improved the flexibility of the aircrafts and the control system. However, the traditional radar-control based ATM system cannot meet the requirements of the gradually crowding airspace with the vigorous development of the civil aviation industry.

Continuous decline in punctuality rate of the airports makes air traveling experience bad, and the delays and cancellations of the flights have resulted in huge economic losses. The solutions to these problems necessitate a more accurate and intelligent ATM system.

Air traffic flow statistics and prediction, the main two parts of this work based on the aviation big data platform, play a significant role in improving the accuracy and intelligence of the ATFM, respectively. The real-time messages provided by our ADS-B based aviation big data platform can be an accurate data source for air traffic flow statistic task, which is helpful for optimizing the flight scheduling strategy.

Furthermore, with the application of deep learning in our aviation big data platform, we are able to obtain a more accurate prediction model based on the statistical data of the air traffic flow. Thus, the statistical and predicted information can provide an assistance to support future more intelligent flight scheduling strategies and ATM.

III. AIR TRAFFIC FLOW STATISTIC BASED ON ADS-B MESSAGES

The air traffic flow is defined as the number of aircrafts in a specific airspace within a certain time period. Although the total air traffic flow is relatively low in China, the distribution of the air traffic is extremely uneven due to the regional development difference. The aircraft traffic mainly concentrates in relatively developed regions and some tourist cities. Therefore, we can expect much more congestion and delay in these cities airspace with the continuous growth of flights. The congestion and delay will cost a lot of manpower and resources.

The ADS-B messages collected by our aviation big data platform can provide more accurate and comprehensive general information, including ICAO number, position, velocity, and angle information. As the aircrafts equipped with ADS-B transmitters automatically broadcast the ADS-B messages, there are sufficient data over time to facilitate the air traffic flow statistic task.

The nearly a million pieces of data from the ADS-B ground stations we have deployed can support the subsequent data processing and analyses, and thus to capture the features and trends of the air traffic. Fig. 4 shows the main workflow of the air traffic flow statistic task which can be divided into four modules, namely ADS-B receiver module, pre-processing module, statistic module, and display module. First, we collect the data by using the distributed ADS-B receivers and store them in the data center. Regarding to the pre-processing module, we extract data from the data center and deal with them with a series of operations: data clean, information extraction, outliers processing, chronological order, data partitioning, and data indexing. The processed data is eventually re-stored to the data center waiting for subsequent operations in the statistic module; the operations include data extraction, coordinate transformation, air routes generation, and positional and existential verification. Finally, we can get the air route traffic information and render it on the display module.

A. Pre-processing Module

To create an available dataset for our further study, some data pre-processing methods have been implemented on the obtained big volume ADS-B messages. The processes are as follow.

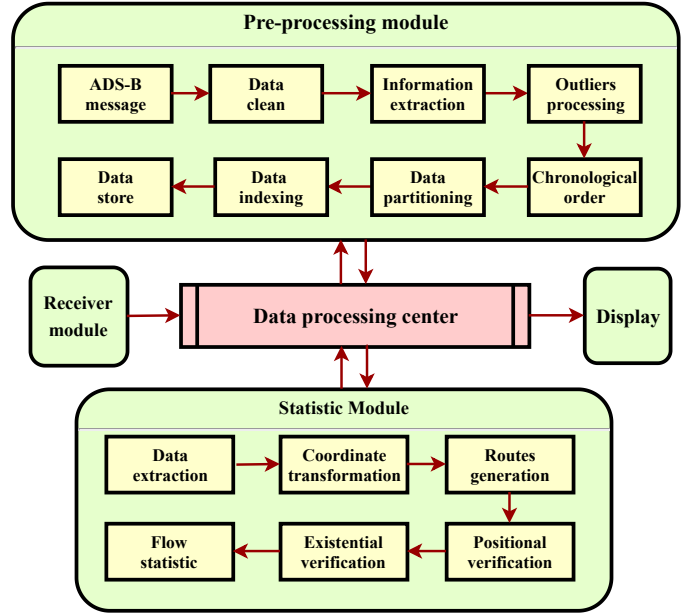


Fig. 4. Workflow of the air traffic flow statistic task.

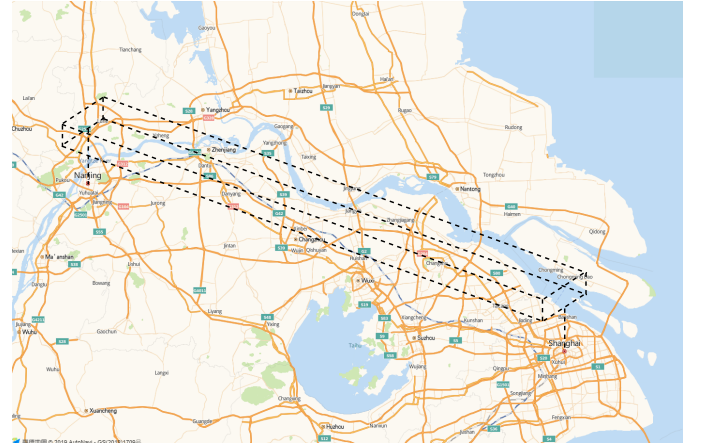


Fig. 5. Illustration of the defined cuboid route.

First, we divide the collected ADS-B messages by date and flight number, so we can get data arrays of different flights in every single day. Each item of the data arrays contains the content of the complete ADS-B information including ICAO identity number, latitude, longitude, speed etc. To eliminate the impact of some invalid data, we also need to filter some abnormal data due to possible unstable reception, for example, garbled characters and coordinate values exceeding the actual range. And we remove all zero values existing in the ADS-B messages. Then we extract the information needed for flow statistic, including flight numbers, latitude and longitude information, and etc. And the truncated data items are sorted by time. Thus, the data items are separated by the index of time, date, and flight number, respectively. After that, they are stored in the data center waiting for subsequent processing.

B. Statistic Module

To obtain comprehensive air route information and to enrich the dataset for a better air traffic prediction task, we define each air route as a corridor with a width of 10 km that connects one specified airport to another, which is shown in Fig. 5. The height of the corridor is assumed stretching from the mean sea level to an infinity altitude. Thus, we can calculate the number of flights within a specific air route over certain time period, namely, the statistical traffic flow of the route. The specific steps of the statistic task are as follows.

Initially, we determine the evaluation area of a specific air route by the specific range of longitude, latitude and altitude. Those parameters are defined by the positional information of the ends of this air route. Also, the defined ranges are prepared for subsequent calculation and verification.

Next, we retrieve the data in a certain period from the data center and divide them into hourly separated intervals for making reasonable comparisons at different time intervals on a single air route.

Finally, we traverse the data slice with hourly period separation to count the flights in this air route. Before we add the ICAO number of a flight appearing in air route to obtain a set called flight-in-route, there are still two verification steps: existential verification for checking whether this flight number is already added; positional verification for checking whether the position of this flight is within the air route.

IV. AIR TRAFFIC FLOW PREDICTION METHODS

The study of prediction task is an important part of the field of time series. Traditional time series prediction models can be divided into two categories: mathematical statistics based prediction methods such as ARIMA [35], and nonlinear prediction methods such as chaotic model [36].

With the development of machine learning, a series of artificial intelligence algorithms such as neural networks [37] and support vector machine (SVM) [38] have been widely used in the field of time series prediction. These modern algorithms show superior performance in exploring data with hidden features.

In this section, we propose two prediction models based on the SVR and LSTM, respectively, which are trained by the massive ADS-B data obtained from our aviation big data platform.

A. Features Selection

The input dataset plays a significant role in any predictor based on machine learning, and the performance of the prediction model can be broadly determined by the selection of the data features. According to the related work, the selected features mainly include time-series features such as date and time [39], [40].

As the air route flow is a time series, the time-related basic features are considered by us alongside some factors that may affect the air route flow. Thus, we can form an input vector \mathbf{x} containing these features. We first define two vectors denoting

two types of features, namely time vector \mathbf{t} and impact vector \mathbf{p} , which are written as

$$\mathbf{t} = [t_1, t_2, t_3] \quad (1)$$

$$\mathbf{p} = [p_1, p_2, p_3] \quad (2)$$

where t_1 , t_2 and t_3 denote the hour of day, the day of week, and the day of month, respectively. p_1 , p_2 , and p_3 denote the holiday index, season index, and average flow, respectively. Hence, the input vector \mathbf{x} can be denoted as:

$$\mathbf{x} = [f, r, \mathbf{t}, \mathbf{p}] \quad (3)$$

where f denotes the flight number within the air route, r denotes the air route, \mathbf{t} and \mathbf{p} are the time vector and impact vector, respectively.

Generally, it is hard to extract the complex nonlinear relationship between the air route flow and the above features by using traditional time series prediction models, let alone to deal with the big dataset. Therefore, it is necessary to apply some artificial intelligence algorithms to better use the massive data and obtain more accurate prediction results.

B. SVR-based Flow Prediction

The SVR is an extension of the SVM, and it can bring in efficient solutions for regression problems. In summary, the SVR converts a nonlinear problem into a linear problem in a high-dimensional space, and calculates the complex features of the high-dimensional space through the kernel functions [41], [42]. For a pair (x, y) , traditional regression models typically calculate the loss by the computing the difference between the predicted $f(x)$ and the true value y , and the loss is zero only if they are identical. In contrast, the SVR assumes that we can tolerate a threshold ϵ as the maximum discrepancy between them [43]. The SVR problem can be formulated as follows:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m \ell_{\epsilon}(f(\mathbf{x}_i) - y_i) \quad (4)$$

where C denotes a regularization constant, and ℓ_{ϵ} the ϵ -insensitive loss function expressed as:

$$\ell_{\epsilon}(z) = \begin{cases} 0, & \text{if } |z| \leq \epsilon \\ |z| - \epsilon, & \text{otherwise} \end{cases} \quad (5)$$

Thus, the loss is calculated only when the absolute value of the difference between the $f(x)$ and y is greater than ϵ . This is equivalent to constructing a band with a width of 2ϵ centered on $f(x)$. If the sample falls into this zone, the regression result is correct.

C. LSTM-based Flow Prediction

The LSTM is evolved from the recurrent neural network [44], and it is suitable for processing and predicting events with relatively long intervals and delays in time series [45]. One of its advantages is that it can avoid the gradient vanishing problem in traditional recurrent neural networks. The LSTM shows gigantic power in the field of natural language

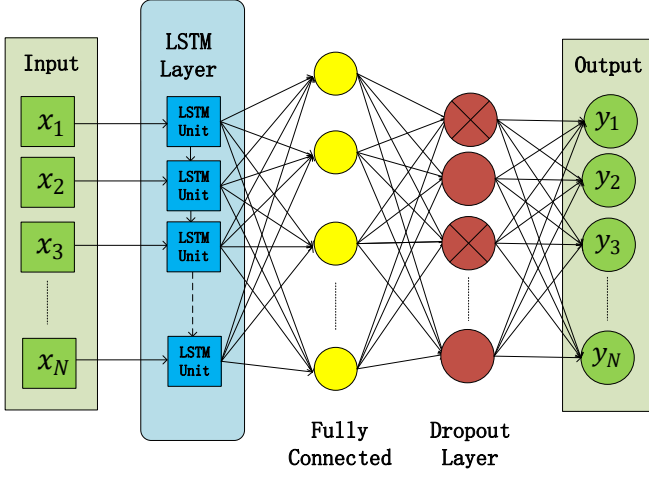


Fig. 6. Architecture of the LSTM-based air traffic prediction model.

processing, target recognition, and sound detection [46]–[48]. Additionally, there is unique structure named gated neuron existing in the LSTM cells. The structure allows short-term and long-term memories be captured, which makes the LSTM suitable for time sequence prediction tasks.

The structures of the LSTM cell can be explained by four forget gates. The output of the LSTM network can be calculated by the following functions [49], [50]:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (6)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (7)$$

$$c_t = f_t c_{t-1} + \tanh(W_{xc}x_t + W_{hc}h_t + b_c) i_t \quad (8)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (9)$$

$$h_t = o_t \tanh(c_t) \quad (10)$$

where $x(t)$ is the input of the model at time t , W and b denote weight matrices and bias vectors, respectively. i_t , f_t , c_t , and o_t denote the four different gates, namely input gate, forget gate, candidate gate and output gate. h_t denotes RNN hidden layer state $h = [h_1, h_2, \dots, h_t]$.

Fig. 6 shows the architecture of the LSTM-based model, which includes three layers: the LSTM layer, the fully connected layer, and the dropout layer. The LSTM layer functions as capturing the time correlation among the air traffic states at different time. The time-steps set in our experiment is 24, which suggests LSTM layer unfold as 24 layers in the time domain. The dropout layer devitalizes neurons with a certain probability, which can improve the generalization ability of the networks. The fully-connected layer is used to reshape the output as the expected form and the active function is rectified linear unit (ReLU) function.

This work is about a regression problem for time series. Therefore, we compare the true statistical flow data y with

the predicted value \tilde{y} by calculating root mean squared error (RMSE) and mean absolute error (MAE). These metrics are defined as follow:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i|^2} \quad (12)$$

Here, smaller values of these metrics mean the better performance of the proposed prediction models.

V. EXPERIMENT RESULTS

A. Setup

The ADS-B omnidirectional antenna in each ADS-B ground station can receive ADS-B signals broadcasted from every flight equipped with an ADS-B transmitter. The coverage of each antenna is within 300km in diameter. We have deployed 14 ADS-B ground stations in Northeast China, North China and East China, and the number is increasing. We have collected all the ADS-B data from December 2018 to May 2019 with a storage of 75 GB in total. Based on the approximately one million items of ADS-B data in six-month period, we have carried out a traffic flow statistic for 240 routes defined by arbitrary two airports. To make a description and analysis without loss of generality, we selected several typical busy air routes from different regions, i.e., Shanghai–Tianjin route, Beijing–Wuhan route, Nanjing–Beijing route, and Guangzhou–Shanghai route.

B. Results of Statistical Air Route Flow

The statistical hourly number of the flights inside the Beijing–Wuhan route is shown in the Fig. 7. the left is the hourly flow value of certain day in May and the right shows the overall traffic flow trend in May. Beijing and Wuhan are one of the most prosperous cities in China, and the air route connecting the two cities is expected to be busy.

To make comprehensive analyses and reasonable comparisons, we choose another three air routes, i.e., Shanghai–Tianjing route, Nanjing–Beijing route and Guangzhou–Shanghai route. The statistical results of the traffic flow are depicted in Fig. 8. As the figure shows, the air traffic in the four air routes follow a similar periodic law, and all the statistical flows of these routes show obvious peaks and valleys.

In the early hours of the morning, the route traffic was very sparse. It begins to rise at 5:00 am and reaches the peak at about 10:00 am and continues to maintain a peak flow within about ten hours. The flow begins to decrease gradually from 9:00 pm. Because of the connection between major cities in the north and south, the Nanjing–Beijing and Shanghai–Tianjin routes both have heavy traffic, exhibiting median peak traffic flow reaching 323 and 278, respectively.

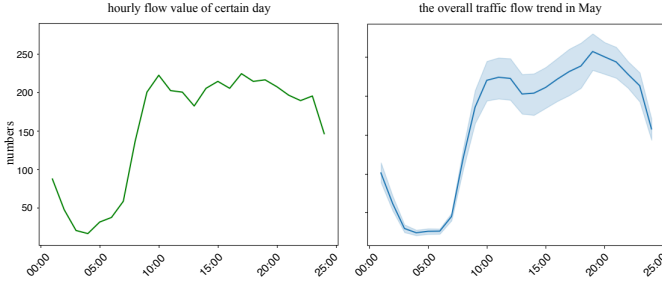


Fig. 7. Statistical hourly number of flights within the Beijing-Wuhan route: values of a certain day (left), and trend of a certain month (right).

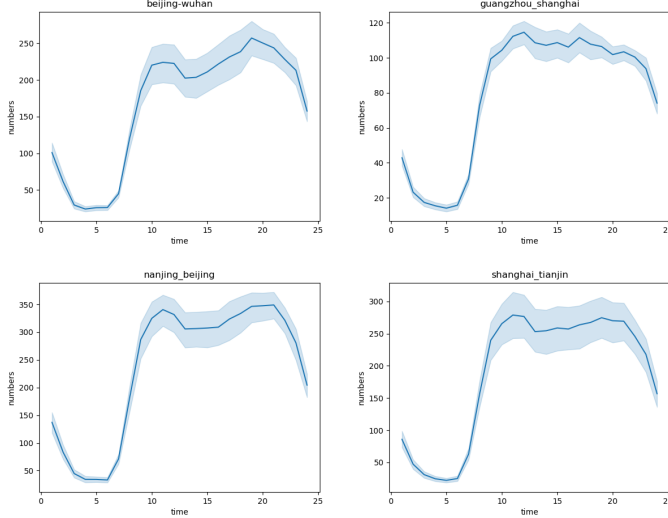


Fig. 8. Statistical hourly number of flights within different routes.

C. Results of Air Traffic Flow Prediction

The dataset for the proposed two predictors (SVR-based and LSTM-based) is generated from the hourly separated flow information of the 240 routes from November 2018 to May 2019. Since the existence of abnormal data conditions, any traffic flow far exceeds the historical average value (above 1000) will be regarded as an invalid value and will be replaced by the historical average value for this hour. The input of the predictors are some selected features as follows:

- Hour of day, day of the week, day of the month, and season of the year.
- Whether the day is statutory holiday.
- Historical average traffic flow.

The performance of the SVR-based predictor can be greatly influenced by the selection of the kernel function. There are many kinds of kernel functions, e.g., linear function, polynomial function, and radial basis function (RBF) [41]. We compare the models using the three kernel functions with the same air route dataset to select an appropriate kernel function. Based on the tested case, we choose RBF as the kernel function because of its outstanding efficiency in the prediction task compared with other kernel functions.

Then two parameters are considered. The parameter penalty C determines the size of the weight vector, and the parameter gamma from RBF determines the width of the RBF corre-

sponding to each support vector and thus affects generalization performance. In order to obtain an appropriate parameter combination, a grid search method has been implemented and the best combination is selected by calculating RMSE score. The combinations of the parameters are shown in Table I.

TABLE I
PARAMETERS CONSIDERED IN SVR BASED PREDICTOR

Parameter	Setting
penalty C	50, 100, 200, 300
gamma	0.1, 0.01, 0.02, 0.04

The data was divided into training and testing datasets with a ratio of 3:1 for the LSTM-based predictor. There are many parameters that affect the performance of the LSTM-based model, which are also critical to be determined. The parameters and options that need to be considered for the LSTM-based model are as follow in Table II.

TABLE II
PARAMETERS CONSIDERED IN LSTM BASED PREDICTOR

Parameter	Setting
activation function	tanh, soft-sign
recurrent activation	sigmoid, hard-sigmoid
dropout ratio	0, 0.2, 0.5
epochs	200, 400, 600
batch size	10%, 20%, 30%

TABLE III
RESIDUALS DISTRIBUTIONS OBTAINED BY THE TWO PREDICTORS

Residuals Percentage	SVR	LSTM
less 30	62.79%	77.06%
30-50	23.14%	12.55%
50-100	11.23%	8.24%
over 100	2.84%	2.15%

To find the best combination of parameters, we train the model with different hyper-parameters by using a grid search method, and the aim is to obtain the lowest RMSE score.

D. Analysis

When processing a large amount of air route traffic flow data, the SVR predictor spends more time than the LSTM-based predictor. And the RMSE score of the two predictors are 38.31 and 29.73, respectively. The results of the two prediction models are depicted in Fig. 9 and Fig. 10, respectively. In the two figures, the top sub-figures present the predicted air flow information, where the blue line, orange line, and green line are true air flow value, validated value, and tested value, respectively.

As the traffic restrictions caused by some activities may bring in abnormal factors to the air traffic. The residual figures (bottom in Fig. 9 and Fig. 10) show the prediction ability considering abnormal traffic flow, and we can see that the LSTM-based model shows superior performance in mitigating the large residual values caused by the abnormal factors. On

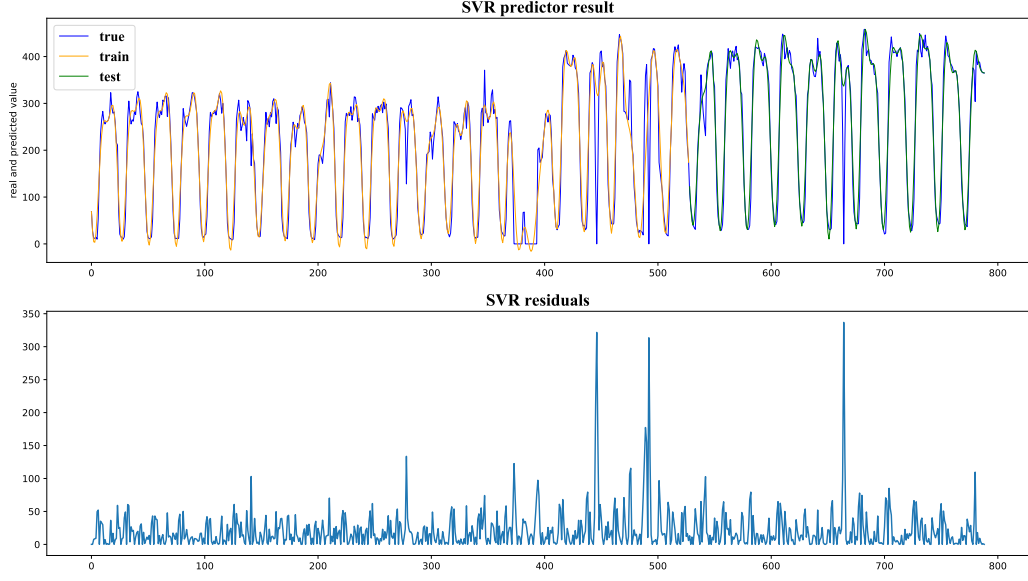


Fig. 9. The predicted traffic flow and the corresponding residual by using the SVR-based predictor.

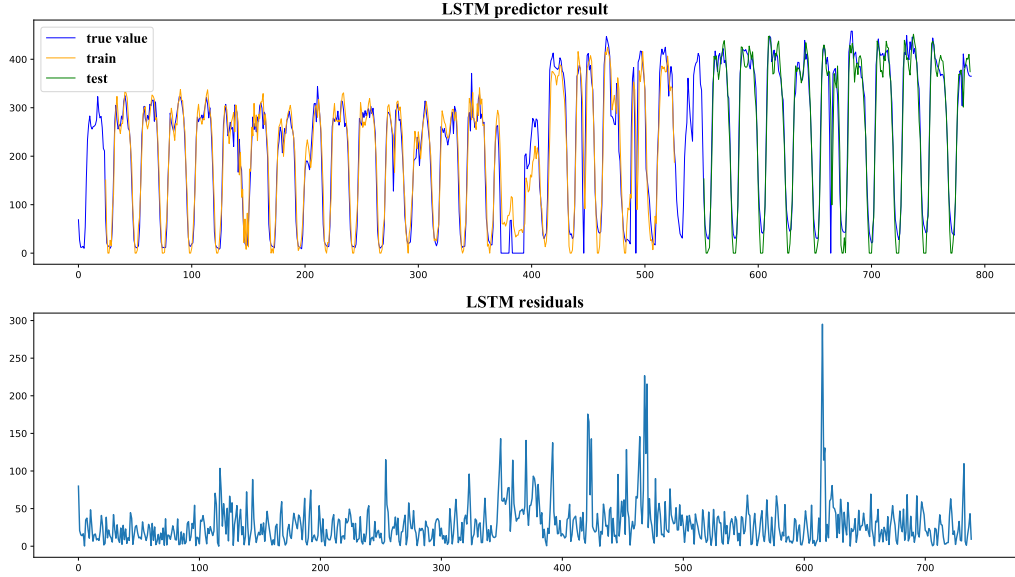


Fig. 10. The predicted traffic flow and the corresponding residual by using the LSTM-based predictor.

the other hand, the residual distributions are shown in Table III. In detail, the SVR-based prediction model counts 62.79%, 23.14%, 11.23%, and 2.84% for the different intervals, respectively. And the LSTM-based model counts 77.06%, 12.55%, 8.24%, and 2.15%, respectively. The results also demonstrate the better performance of LSTM-based model in predicting the air traffic flow.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a model for calculating the air traffic flow using the big volume ADS-B data obtained by our aviation big data platform. Based on this model, we calculated the flow information of more than 200 air routes, and further made some visualization tasks for the sake of ATFM. Moreover, two prediction models applying SVR and

LSTM were proposed to facilitate a timely surveillance and optimization of the traffic flow. Due to the obvious peaks and valleys of the flow variation, we found that the SVR-based predictor can accurately predict the flow because of its advantage in handling data with recognizable trend. The LSTM-based predictor can provide improved performance over the SVR-based one in the tested case, and it is inferred that the LSTM-based predictor can better deal with abnormal points of the flow variation.

Future work in the field of ATFM could be investigated. For instance, the air traffic statistics and prediction tasks can be executed in airports and control sectors. Features concerning the abnormal change of the flow can be investigated. Additionally, the advantages of different models can be combined to seek a better prediction accuracy.

REFERENCES

- [1] Y. Zhang, R. Su, Q. Li, C. G. Cassandras, and L. Xie, "Distributed flight routing and scheduling for air traffic flow management," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2681–2692, 2017.
- [2] S. Neemat and M. Inngs, "Design and implementation of a digital real-time target emulator for secondary surveillance radar/identification friend or foe," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 27, no. 6, pp. 17–24, 2012.
- [3] A. Yang, X. Tan, J. Baek, and D. S. Wong, "A new ads-b authentication framework based on efficient hierarchical identity-based signature with batch verification," *IEEE Trans. Serv. Comput.*, vol. 10, no. 2, pp. 165–175, 2015.
- [4] M. Leonardi, "ADS-B anomalies and intrusions detection by sensor clocks tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 5, pp. 2370–2381, 2018.
- [5] T. Kacem, D. Wijesekera, and P. Costa, "Ads-bsec: A holistic framework to secure ads-b," *IEEE Trans. Intell. Veh.*, vol. 3, no. 4, pp. 511–521, 2018.
- [6] K.-Q. Cai, J. Zhang, M.-M. Xiao, K. Tang, and W.-B. Du, "Simultaneous optimization of airspace congestion and flight delay in air traffic network flow management," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3072–3082, 2017.
- [7] Y. Yang, "Practical method for 4-dimensional strategic air traffic management problem with convective weather uncertainty," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 6, pp. 1697–1708, 2017.
- [8] C. Guan, P. B. Luh, L. D. Michel, and Z. Chi, "Hybrid kalman filters for very short-term load forecasting and prediction interval estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3806–3817, 2013.
- [9] M. Laner, P. Svoboda, and M. Rupp, "Parsimonious fitting of long-range dependent network traffic using arma models," *IEEE Commun. Lett.*, vol. 17, no. 12, pp. 2368–2371, 2013.
- [10] K. Yunus, T. Thiringer, and P. Chen, "ARIMA-based frequency-decomposed modeling of wind speed time series," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2546–2556, 2015.
- [11] R. Ak, O. Fink, and E. Zio, "Two machine learning approaches for short-term wind speed time-series prediction," *IEEE Trans. Neural Netw. Learn.*, vol. 27, no. 8, pp. 1734–1747, 2015.
- [12] N. Kato, Z. M. Fadlullah, B. Mao, F. Tang, O. Akashi, T. Inoue, and K. Mizutani, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wirel. Commun. Mag.*, vol. 24, no. 3, pp. 146–153, 2017.
- [13] L. Zhao, J. Wang, J. Liu, and N. Kato, "Routing for crowd management in smart cities: A deep reinforcement learning perspective," *IEEE Commun. Mag.*, vol. 57, no. 4, pp. 88–93, 2019.
- [14] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 4, pp. 2432–2455, 2017.
- [15] B. Mao, Z. M. Fadlullah, F. Tang, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "Routing or computing? the paradigm shift towards intelligent computer network packet transmission based on deep learning," *IEEE Transactions on Computers*, vol. 66, no. 11, pp. 1946–1960, 2017.
- [16] Y. Zhou, Z. M. Fadlullah, B. Mao, and N. Kato, "A deep-learning-based radio resource assignment technique for 5G ultra dense networks," *IEEE Network*, vol. 32, no. 6, pp. 28–34, 2018.
- [17] M. Liu, J. Yang, and G. Gui, "DSF-NOMA: UAV-assisted emergency communication technology in a heterogeneous internet of thing," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5508–5519, 2019.
- [18] M. Liu, T. Song, G. Gui, J. Hu, and H. Sari, "Deep cognitive perspective: resource allocation for noma based heterogeneous iot with imperfect sic," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2885–2894, 2019.
- [19] T. K. Rodrigues, K. Suto, H. Nishiyama, J. Liu, and N. Kato, "Machine learning meets computation and communication control in evolving edge and cloud," *IEEE Communications Surveys and Tutorials*, to be published, doi: 10.1109/COMST.2019.2943405.
- [20] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, 2019.
- [21] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2019.2949122.
- [22] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, 2018.
- [23] L. Liu and R.-C. Chen, "A novel passenger flow prediction model using deep learning methods," *Transp. Res. Part C Emerg. Technol.*, vol. 84, no. 11, pp. 74–91, 2017.
- [24] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of pv power output," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, 2014.
- [25] H. Palangi, L. Deng, Y. Shen, J. Gao, X. He, J. Chen, X. Song, and R. Ward, "Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 4, pp. 694–707, 2016.
- [26] Y. Zhou, Y. Wan, S. Roy, C. Taylor, C. Wanke, D. Ramamurthy, and J. Xie, "Multivariate probabilistic collocation method for effective uncertainty evaluation with application to air traffic flow management," *IEEE Trans. Syst., Man, Cybern.*, vol. 44, no. 10, pp. 1347–1363, 2014.
- [27] K. Margellos and J. Lygeros, "Toward 4-d trajectory management in air traffic control: a study based on monte carlo simulation and reachability analysis," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 5, pp. 1820–1833, 2012.
- [28] M. Strohmeier, M. Schafer, V. Lenders, and I. Martinovic, "Realities and challenges of nextgen air traffic management: the case of ads-b," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 111–118, 2014.
- [29] G. Thumbur, N. Gayathri, P. V. Reddy, M. Z. U. Rahman, and A. Lay-Ekuakille, "Efficient pairing-free identity-based ads-b authentication scheme with batch verification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 5, pp. 2473–2486, 2019.
- [30] R. Holdsworth, J. Lambert, and N. Harle, "Inflight path planning replacing pure collision avoidance, using ads-b," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 16, no. 2, pp. 27–32, 2001.
- [31] D. He, N. Kumar, K.-K. R. Choo, and W. Wu, "Efficient hierarchical identity-based signature with batch verification for automatic dependent surveillance-broadcast system," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 2, pp. 454–464, 2016.
- [32] W. Zhang, M. Kamgarpour, D. Sun, and C. J. Tomlin, "A hierarchical flight planning framework for air traffic management," *Proc. IEEE*, vol. 100, no. 1, pp. 179–194, 2011.
- [33] P. Arico, G. Borghini, G. Di Flumeri, S. Bonelli, A. Golfetti, I. Graziani, S. Pozzi, J.-P. Imbert, G. Granger, R. Benhacene, *et al.*, "Human factors and neurophysiological metrics in air traffic control: a critical review," *IEEE Rev. Biomed. Eng.*, vol. 10, no. 4, pp. 250–263, 2017.
- [34] M. Mitici and H. A. Blom, "Mathematical models for air traffic conflict and collision probability estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 1052–1068, 2018.
- [35] J. Guo, H. He, and C. Sun, "ARIMA-based road gradient and vehicle velocity prediction for hybrid electric vehicle energy management," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2019.2912893.
- [36] N. Safari, C. Chung, and G. Price, "Novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 590–601, 2017.
- [37] R. Chandra, "Competition and collaboration in cooperative coevolution of elman recurrent neural networks for time-series prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3123–3136, 2015.

- [38] C.-H. Liao and C. H.-P. Wen, "Svm-based dynamic voltage prediction for online thermally constrained task scheduling in 3-d multicore processors," *IEEE Embedded Syst. Lett.*, vol. 10, no. 2, pp. 49–52, 2017.
- [39] Z. Ma, J. Xing, M. Mesbah, and L. Ferreira, "Predicting short-term bus passenger demand using a pattern hybrid approach," *Transp. Res. Part C Emerg. Technol.*, vol. 39, no. 2, pp. 148–163, 2014.
- [40] X. Jiang, L. Zhang, and X. M. Chen, "Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in china," *Transp. Res. Part C Emerg. Technol.*, vol. 44, no. 7, pp. 110–127, 2014.
- [41] C.-H. Wu, J.-M. Ho, and D.-T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, 2004.
- [42] J. Guo, Z. Xie, Y. Qin, L. Jia, and Y. Wang, "Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM," *IEEE Access*, vol. 7, pp. 42946–42955, 2019.
- [43] J. Cai, J. King, C. Yu, J. Liu, and L. Sun, "Support vector regression-based behavioral modeling technique for RF power transistors," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 5, pp. 428–430, 2018.
- [44] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, 2016.
- [45] Y. Hu, X. Sun, X. Nie, Y. Li, and L. Liu, "An enhanced LSTM for trend following of time series," *IEEE Access*, vol. 7, pp. 34020–34030, 2019.
- [46] M. Sundermeyer, H. Ney, and R. Schlüter, "From feedforward to recurrent LSTM neural networks for language modeling," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 3, pp. 517–529, 2015.
- [47] T. Hayashi, S. Watanabe, T. Toda, T. Hori, J. Le Roux, and K. Takeda, "Duration-controlled LSTM for polyphonic sound event detection," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 11, pp. 2059–2070, 2017.
- [48] J. Liu, G. Wang, L.-Y. Duan, K. Abdiyeva, and A. C. Kot, "Skeleton-based human action recognition with global context-aware attention LSTM networks," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1586–1599, 2017.
- [49] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband rf power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10348–10356, 2019.
- [50] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2019.2954094.



Ziqi Zhou (S'18) is currently pursuing the M.Sc degree in signal processing for wireless communications with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. Her research interest includes deep learning and its applications.



Juan Wang (S'19) Juan Wang is currently pursuing the master's degree in communication and information engineering with the Nanjing University of Posts and Telecommunications, Nanjing, China. Her research interest includes machine learning for wireless communications.



Fan Liu (S'18) is currently pursuing the M.Sc degree in signal processing for wireless communications with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. Her research interest includes deep learning and its applications.



Guan Gui (M'11-SM'17) is a professor with Nanjing University of Posts and Telecommunications, Nanjing, China. He received the Dr. Eng degree in Information and Communication Engineering from University of Electronic Science and Technology of China, Chengdu, China, in 2012. From September 2012 to March 2014, he was supported by Japan society for the promotion of science (JSPS) fellowship as postdoctoral research fellow at Tohoku University. From April 2014 to October 2015, he was an Assistant Professor in Department of Electronics and

Information System, Akita Prefectural University.

He is currently engaged in research of deep learning, compressive sensing and advanced wireless techniques. Dr. Gui has published more than 200 international peer-reviewed journal/conference papers. He received Member and Global Activities Contributions Award in IEEE ComSoc and 9 best paper awards including ICNC 2018, ICC 2017, ICC 2014 and VTC 2014-Spring. He was also selected as for Jiangsu Specially-Appointed Professor (2016), Jiangsu High-level Innovation and Entrepreneurial Talent (2016), Jiangsu Six Top Talent (2018), Nanjing Youth Award (2018). Dr. Gui was an Editor of Security and Communication Networks (2012–2016). He has been the Editor of IEEE Transactions on Vehicular Technology, since 2017, the Editor of IEEE Access, since 2018, the Editor of KSII Transactions on Internet and Information Systems since 2017, the Editor of Journal on Communications, since 2019, and the Editor-in-Chief of EAI Transactions on Artificial Intelligence, since 2018. He is IEEE Senior Member.



Jinlong Sun (M'18) received the M.Sc and Ph.D. degrees from the Harbin Institute of Technology (HIT), Harbin, China, in 2014 and 2018, respectively. He is currently working as an assistant professor with the Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China. His current research interests include signal processing for wireless communications, machine learning, and integrated navigation systems.