Research Article

Probabilistic En Route Sector Traffic Demand Prediction Based upon Statistical Analysis of Error Distribution Characteristics

Wen Tian ![ORCID]**, Huiqing Xu, Yixing Guo, Bin Hu, and Yi Yao**

National Key Laboratory of Air Traffic Flow Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Correspondence should be addressed to Wen Tian; tianwen0665@qq.com

In China, air traffic congestion has become increasingly prominent and tends to spread from terminal areas to en route networks. Accurate and objective traffic demand prediction could alleviate congestion effectively. However, the usual demand prediction is based on conjecture method of flying track, and the number of aircraft flying over a sector in a set time interval could be inferred through the location information of any aircraft track. In this paper, we proposed a probabilistic traffic demand prediction method by considering the deviations caused by random events, such as the change of departure or arrival time, the temporary change in route or altitude under severe weather conditions, and unscheduled cancellation for a flight. The probabilistic method quantifies these uncertain factors and presents numerical value with its corresponding probability instead of the deterministic number of aircraft in a sector during a time interval. The analysis results indicate that the probabilistic traffic demand prediction based on error distribution characteristics achieves an effective match with the realistic operation in airspace of central and southern China, which contributes to enhancing the implementation of airspace congestion risk management.

1. Introduction

As an important part of the Next Generation Air Transportation System (NextGen) program, traffic flow management based on probabilistic traffic demand prediction by considering uncertainty has received much attention in recent years due to the stochastic nature of the predictions. It provides more accurate prediction results that will benefit realistic decisions than deterministic prediction.

Intelligence and dynamic of traffic environment has been deepened with the complexity since the 90s of last century. The method of original static and deterministic traffic demand prediction has gradually been unable to meet the requirements of optimizing the air traffic management. Many scholars began to study the characteristics of the uncertainties in the air traffic system from the perspective of both the trajectory uncertainty and the uncertainties of departure/arrival time [1–10]. In 2002, Larry A. Meyn [11] established a probabilistic approach to predicting traffic demand and made a scientific estimate on demand of the airport through Monte-Carlo simulation. In the same year, K. T. Muller et al. [12]

studied the uncertainty models of the factors in the strategic trajectory prediction and simulated the future operating environment by controlling the distribution of these factors. In 2003, Sandip Roy [13] established an aggregate dynamic stochastic model based on Poisson distribution to quantify the uncertainty of traffic demand. For the disadvantages of the method, Craig R.Wanke et al. [14] extracted and measured the characteristics of uncertainty in demand predictions based on the deviation analysis between predicted values and actual values. In 2004, Craig R. Wanke et al. [15] improved the algorithm and established a statistical model of error distribution applied to Monte-Carlo simulation under special traffic samples. In 2007, E. Gilbo and S. Smith [16] improved the ETMS model by proposing a new model to improve aggregate air traffic demand predictions. In 2009, E. Gilbo and S. Smith [17] analyzed probabilistic prediction of aggregate traffic flow in the terminal area of the airport using uncertainty in individual flight predictions. In 2011, E. Gilbo et al. [18] proposed a new method for probabilistic traffic demand predictions for en route sectors based on uncertain predictions of individual flight events. Wen Tian

(a) Quadrant I   (b) Quadrant II   (c) Quadrant III   (d) Quadrant IV
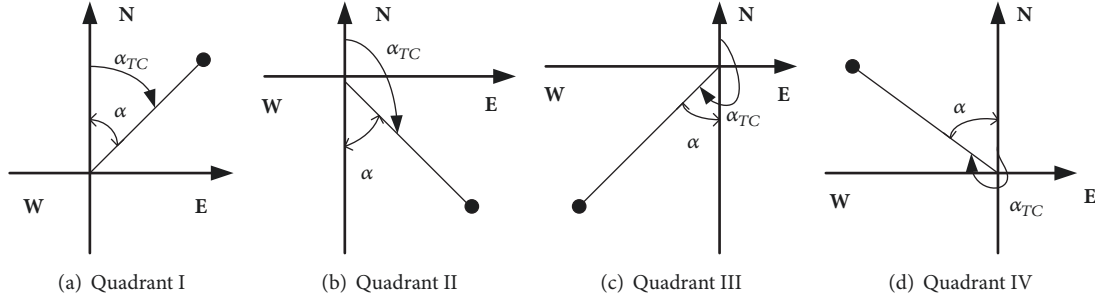
FIGURE 1: Relationship between α_{TC} and α.

and Minghua Hu [19] of Nanjing University of Aeronautics and Astronautics explored the random variation of traffic demand in the airspace sectors in 2011 and established a probabilistic traffic demand prediction model for tactical operations. The proposed model has made contributions to the further study of uncertainty and measurement in air traffic demand prediction. Also in 2011, Chao Wang and Le Yang [20] of Civil Aviation University of China proposed a prediction method of congestion prediction based on the Monte-Carlo simulation. They established the probabilistic distribution model based on aircraft sector entry time, sector exit time, and stay time in the airspace sectors. In 2013, from the perspective of demand uncertainty, Shanmei Li et al. [21] analyzed the influence of aircraft arrival and departure time on the airport traffic demand prediction and established a probabilistic distribution model of multiperiod method. In the same year, Wen Tian [22] further optimized the probabilistic traffic demand prediction method for airspace sectors and applied it to the risk management of airspace congestion. In addition, more and more ATFM researches are probability/uncertainty-based on, which have made contributions to network optimization and decreased congestion. F. Gonze et al. [23] proposed a probabilistic framework for modeling air traffic occupancy count to provide ATC with more precise and valuable information. Dan Chen et al. [24] established a model to reduce en route delays by characterizing realistic dynamics and uncertainty of en route airspace system. Prediction uncertainty is explicitly used in demand and capacity balancing tools to develop effective and efficient congestion resolution actions [25]. YH Chang et al. [26] mainly focus on reduced sector volume problems caused by weather uncertainty and proposed effective solutions to avoid sector congestion.

As mentioned, the majority of traditional research has focused on the measurement of uncertain traffic demand prediction. For more and more study based on aircraft, these segments need to be considered based on actual data. According to the prediction deviation of the time that aircraft went through the en route sector boundaries, we could obtain time randomness characteristics of the single aircraft entering and leaving the sector. Therefore, probabilistic traffic demand of the en route sector could be calculated. A large number of targeted aircraft performance data of particular periods and fixes are required, which made high requirements on accumulation and organization of historical data (including

radar data and prediction data). In our traffic prediction and management system, the latitude ϕ and longitude λ of any waypoint could be provided easily. We assume that fix $A(\phi_A, \lambda_A)$ and fix $B(\phi_B, \lambda_B)$ represent origin and destination separately; method of spherical trigonometry was applied to derive the corresponding parameter as (1) and (2). True course α_{TC} depends on which quadrant it is in (Figure 1).

$$\tan \alpha = \frac{\lambda_B - \lambda_A}{\ln \tan \left(45° + \phi_B/2\right) - \ln \tan \left(45° + \phi_A/2\right)} \quad (1)$$

$$D = \left(\lambda_B - \lambda_A\right) \cos \overline{\phi} \cdot \sec \alpha \quad (2)$$

where

α is the angle between the rhumb line track and longitude

$\overline{\phi}$ is the average latitude of fix A and B, i.e., $\overline{\phi} = (\phi_A + \phi_B)/2$

D is the distance from fix A to B (unit: radian)

In addition, the unit of $\lambda_B - \lambda_A$ is radian, α takes absolute value of calculation results and we should converse D's unit from radian to nautical mile. The cruise speed varies with the aircraft types. Then the estimated time of the arrival time at fix B (T_B) since the flying time from A to B could be calculated. During the process of track conjecture, the system scans radar data and telegraph data (such as DEP, departure message; ARR, arrival message; and OVFLY, overfly message) at a regular interval to obtain the latitude, longitude, and speed of the aircraft. The total aircraft in time interval would be the results of deterministic prediction.

2. Data Preparation

An aircraft flying over the en route boundary fixes is defined as a random event. In order to obtain the probability density distribution function of the event, the distribution parameter is usually obtained according to the historical data for a given distribution characteristic, which is a subjective method [19, 21]. In this paper, a direct statistical method was adopted to predict the time error of the large number of aircraft the en route boundary in a certain period of time for a more realistic, objective reflection of random events regularity of the discrete distribution function.
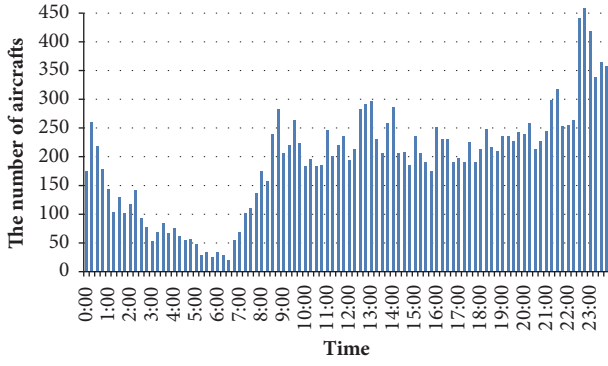
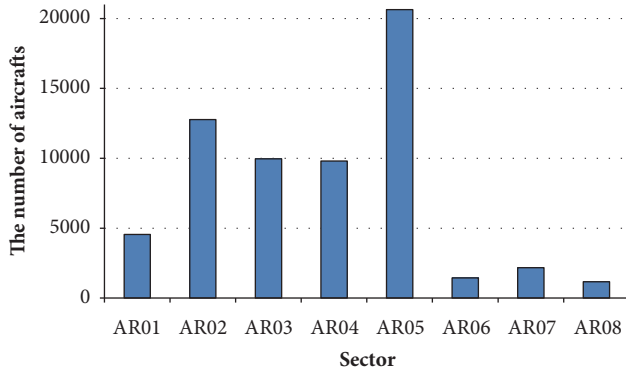FIGURE 2: Bar chart of aircraft in Sector AR05 in central and southern China in different periods in May.

FIGURE 3: Bar chart of aircraft from Sectors AR01 to AR08 in central and southern China in May.

2.1. Analysis of the Prediction Error Characteristics. The relative error of the predicted time of the aircraft through the sector boundary is calculated by comparing the historical predicted time with the corresponding historical actual time. Some regularity can be found after statistical analysis of a large number of predicted values and actual ones, and it may vary depending on the busy level of the predicted periods and the location of targeted sectors. Details are as follows.

(1) *The Busy Level of the Predicted Periods.* As shown in Figure 2 (for details of this analysis see Supplemental Material [S1]), majority of data is concentrated in the time interval of 09: 00-24: 00; it is easier to find the regularity. On the contrary, sectorAR05 seems relatively free during the period of 01: 00-08: 00.

(2) *The Location of Targeted Sectors.* Figure 3 (for details of this analysis see Supplemental Material [S1]) shows the results that the aircraft numbers vary significantly among different sectors. Usually, the more aircraft in the sector there are, the easier error regularity of the predicted time could be found.

2.2. Influencing Factors of Prediction Error and Model Description. It is convenient to describe the problem as a simplified network model composed of four elements: airspace, en route, sectors, and sector boundary fixes. In addition, the

four-dimensional space (one-dimensional time and three-dimensional space) of the aircraft flight is simplified as one-dimension time and two-dimensional space. The entire airspace is divided into two parts: targeted airspace and nontargeted part. As illustrated in Figure 4, targeted airspace consists of those airspace, en route, sectors, and sector boundary fixes which are within predicted space. Others belong to nontargeted part. The aircraft departs from the take-off airport, then enters the targeted airspace through the sector boundary fixes, and finally travels along the en route and continues to fly over several sector boundary fixes until it leaves the targeted airspace.

During the predicted period T, we assume that the total amount of aircraft passed through a targeted sector is N. Aircraft's predicted time error of flying over several sector boundary fixes can be expressed as

$$\Delta t_i = t_i^{etp} - t_i^{atp} \tag{3}$$

where

f_i is aircraft i $(1 \le i \le N)$

t_i^{dep} is departure time of aircraft i

t_i^{etp} is estimated time of aircraft i $(1 \le i \le N)$ flying over several sector boundary fixes

t_i^{atp} is actual time of aircraft i $(1 \le i \le N)$ flying over several sector boundary fixes

The samples that estimated time of aircraft flying over several sector boundary fixes are divided into two subsets. All data before day mth belong to the subset I, data from day $m +$ 1th to the day Mth belong to subset II, and the former class is used to find the distribution characteristics of the prediction deviation. The latter one is used to verify the validity of the statistical regularity. After successful verification, sector enter time of any aircraft on any day can be predicted probabilistically in the future. The analysis process in this paper is based on typical route sectors at typical run time. When the sector changes or emergencies happen (such as bad weather and major events), restatistics and reanalysis on the changed situation are required by method mentioned above to guarantee the accuracy and validity of the error distribution regularity. The location of targeted sectors, the busy degree of the predicted periods, and the difference of the predicted time scale are considered as the main reasons for the time prediction deviation. Two-dimensional probability distribution $f(T_k P_j)$ is used to describe the variation regularity of the prediction error characteristic over time.

(1) T_k represents the time zone. The prediction error results often depends on the busy levels in different periods of the sector. Therefore, according to the air traffic flow management controllers' habits, a day was divided into several time zones for every 15 minutes for analyzing the typical predicted time curve trend and error statistics. k $(k = 1, 2, \ldots, K)$ means time period of error statistics and K indicates the sum of time periods.

(2) P_j is a feature quantity that describes the aircraft's time passing through the boundary fixes. In this paper, a bivariate cumulative distribution function is proposed to describe the predicted time of the aircraft passing through these fixes.
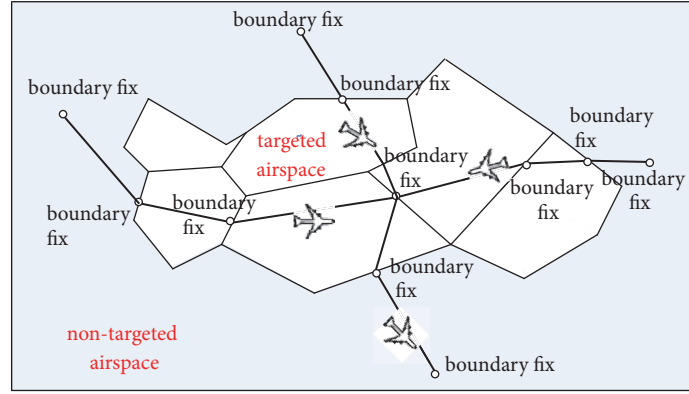
FIGURE 4: Diagram of en route network.

3. Statistical Methods for Predicting Error Distribution

According to the probability statistics management [13], the event frequency function (empirical distribution) can be used as the probability density in big sample analysis due to the minor errors presence compared to the overall distribution density. General solution can be summarized as follows.

(1) We take the 15min time window as the interval unit for the typical operation day of an en route sector and the aircraft's time passing through the boundary fixes could be predicted. Each unit is defined as T_k $(k = 1, 2, \ldots, K)$.

(2) The prediction error sample in interval T_k is divided twice according to the uniform distribution of the number of the prediction error and its size. The steps of the procedure are described below.

(a) The first division is based on prediction error value of interval T_k. Take the number of intervals corresponding to the prediction error value $T_k + i$ $(i = 0, 2, \ldots, 59)$ as the abscissa and the prediction error value as the ordinate to establish coordinate system for the first division. The vertical axis is separated in unit of λ (e.g., 2min).

(b) The second division is necessary to avoid that the deviation trend of prediction error cannot be reflected due to lack of maximum and minimum in the sample. The sample reference interval is set as $[l - \delta, l + \delta]$, where l indicates the appropriate number of sample (e.g., $l = 8$) and δ is fluctuation range (it could be defined approximately 12% of l). The initial divided intervals are merged from both sides of the predicted value to satisfy the number of prediction error samples in each interval as much as possible, and the boundary values of each interval are recorded as well. If the number of samples cannot meet the requirement after merging process, then a smaller reference interval should replace the original one.

Following the steps mentioned above, total W layer partitions corresponding to the interval T_k can be calculated easily. $L_{k,w}$ $(w = 1, 2, \ldots, W)$ is defined as each partition and the sample size is $R_{k,w}$ for partition $L_{k,w}$. As results showed in Figures 5 and 6, the historical data can be classified from the horizontal and vertical angles, respectively. The results reflect the predicted time error distribution of passing through boundary fixes in different intervals.
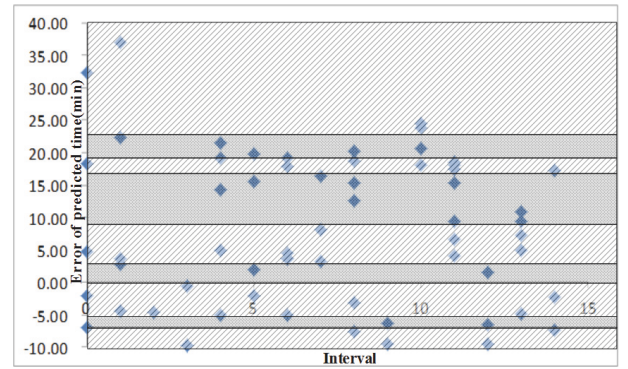
FIGURE 5: Prediction error distribution of $L_{1,1}$.
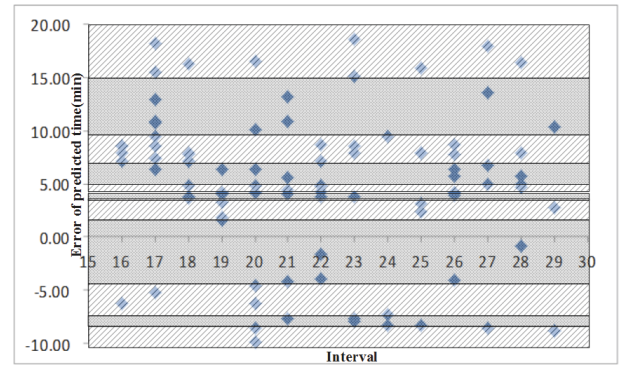
FIGURE 6: Prediction error distribution of $L_{1,2}$.

(3) Calculate prediction errors v_r $(r = 1, 2, \ldots, R_{k,w})$ of $L_{k,w}$ $(k = 1, 2, \ldots, K; w = 1, 2, \ldots, W)$ one by one and the distribution is shown in Figure 7. Note that the abscissa is the percentage of predicted error.

The number of samples in each prediction error interval is calculated as m_s $(s = 1, 2, \ldots, S)$ and $\sum_1^S m_s = R_{k,w}$. In addition, $f_s = m_s/R_{k,w}$, as $R_{k,w}$ is sufficiently large and f_s is regarded as discrete exact probability distribution of the prediction error of layer partition $L_{k,w}$ in interval T_k.
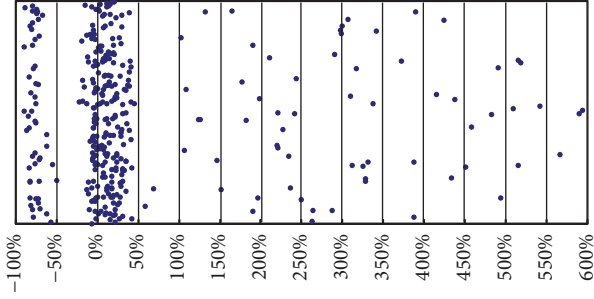
FIGURE 7: Diagram of prediction error distribution.

```
P_0 [0] = 1;
For    (i = 1;  i ≤ M;  i + +)
{   P_i[0] = [1 − P(i)] · P_{i−1}[0];
    P_i[i] = P(i) · P_{i−1}[i − 1];
    For    (k = 1;  k ≤ (i − 1);  k + +)
       {   P_i[k] = P(i) · P_{i−1}[k − 1] + [1 − P(i)] · P_{i−1}[k];}
}
```

PSEUDOCODE 1

(4) Finally, upon overall historical data, we can get exact probability cumulative distribution functions of a total of MW prediction error, which is the table of error statistics. Furthermore, efficiency test should be conducted to ensure that the statistical method of the predicted error distribution characteristic is valid. Assuming that the prediction data of the same time interval in the first day after the historical statistical sample is the predicted sample data, the proposed statistical method is used to calculate the discrete probability distribution predicted time error of the aircraft passing through the boundary fixes and the distribution test is carried out.

(5) According to the time that aircraft passing through the boundary, we divided the aircraft into two categories; one is time of aircraft entering the sector from boundary fixes and the other is time about leaving the sector. Research on probabilistic en route sector traffic demand prediction is based on the error distribution rules of these two types time.

4. Probabilistic En Route Sector Traffic Demand Prediction

4.1. Probabilistic Methods for En Route Sector Traffic Demand Prediction. The probabilistic traffic demand prediction is a derivative of the conventional deterministic traffic demand prediction. After verifying the discrete probability distribution of the prediction error Δt_i of the aircraft's time passing through the boundary fixes, the error statistics can be used to analyze the possible distribution of the aircraft entering and leaving the en route sector. Then the probable number of aircraft in a certain period of one sector can be computed, that is, the probabilistic results of the traffic demand value of the sector. It reflects the implied risk factors in the prediction of traffic demand and provides the prerequisites and basis for the study of the sector congestion risk.

We assume that there are N_k aircraft passing through the sector within targeted prediction interval $[t_k, t_k + 14]$; the predicted time error Δt_i^{in} of these aircraft through the boundary fixes going into the sector can be obtained according to the aforementioned method (i.e., predicted time error of aircraft entering the sector). $p(\Delta t_i^{in})$ and $P(\Delta t_i^{in})$ represent the corresponding probability density function and cumulative probability function separately. Similarly, Δt_i^{out} indicates the predicted time error that aircraft leave the sector through

the boundary fixes and the corresponding probability density probability density function and cumulative probability function are $p(\Delta t_i^{out})$ and $P(\Delta t_i^{out})$.

The predicted time for aircraft i ($1 \leq i \leq N_k$) entering the sector could be calculated by equation $t_i^{in} = t_i^{atp} + \Delta t_i^{in}$, where t_i^{atp} is a constant which means boundary fix arrival time. It is determined by take-off time, route length, and flight performance. Therefore, the probability for aircraft f_i entering and leaving the sector in interval $[t_k, t_k + 14]$ is calculated by the following equations:

$$
\begin{aligned}
P\left(t_k \leq t_i^{in} \leq (t_k + 14)\right) \\
= P\left(t_i^{in} \leq (t_k + 14)\right) - P\left(t_i^{in} \leq t_k\right) \\
= \int_{t_k}^{t_k+14} \left[t_i^{atp} + p\left(\Delta t_i^{in}\right)\right] d\Delta t_i^{in}
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
P\left(t_k \leq t_i^{out} \leq (t_k + 14)\right) \\
= P\left(t_i^{out} \leq (t_k + 14)\right) - P\left(t_i^{out} \leq t_k\right) \\
= \int_{t_k}^{t_k+14} \left[t_i^{atp} + p\left(\Delta t_i^{out}\right)\right] d\Delta t_i^{out}
\end{aligned}
\tag{5}
$$

Based on (4) and (5), the probability for aircraft f_i staying in the sector in interval $[t_k, t_k + 14]$ is shown in

$$
\begin{aligned}
P(i) = P\left(t_k \leq t_i^{in} \leq (t_k + 14)\right) \\
\cdot \left[1 - P\left(t_k \leq t_i^{out} \leq (t_k + 14)\right)\right]
\end{aligned}
\tag{6}
$$

If the deterministic demand prediction value for aircraft f_i existing in the sector in interval $[t_k, t_k + 14]$ on someday is M, then $P_M[m]$ ($0 \leq m \leq M$) represents that there will be m aircraft in the sector of that interval probably. It can be expressed in Pseudocode 1.

4.2. Efficiency Test. In order to verify the validity of the proposed method further, we conduct efficiency test from the perspective of cumulative probability on the basis of the distribution test in the third section. We use $P(i)$ to indicate the probability that aircraft f_i is located within the sector during the interval $[t_k, t_k + 14]$ for full description of error distribution regularity.

We assume that a day is divided into n intervals, the probable demand prediction value is x_i in interval i, and

the corresponding calculated actual demand value based on historical data is y_i; then the correlation coefficient is denoted as

$$\rho = \frac{\sum_1^n (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_1^n (x_i - \overline{x})^2 \sum_1^n (y_i - \overline{y})^2}} \qquad (7)$$

where $\overline{x}$ and $\overline{y}$ denote the mean of x_i and y_i, respectively; therefore, $\overline{x} = \sum_1^n x_i/n$ and $\overline{y} = \sum_1^n y_i/n$. The larger the correlation coefficient ρ is, the better the prediction value fluctuation simulation of the day is. In addition, it is of great practical value and the prediction error laws are equally effective for the same interval of another day.

5. Case Study

5.1. Numerical Examples. In this section, we carried out a case study based on operation data from May 1, 2014, 00:00:00, to May 24, 2014, 23:59:59, of the sectors AR01-AR08 in central and southern China. A total of 19624 time samples of passing through these sectors' boundary fixes from every 07:00 (a relative free level) to 11:59 (a relative busy level) were obtained as historical sample for statistical analysis. Central Limit Theorem (CLT) establishes that, in most situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed. When the sample size is equal to or greater than 30, and the variables are approximately normally distributed [15]. We can assume that the sample space is subject to the normal distribution; the calculation results are shown in Tables 1 and 2 (for details of analysis from 07:00-07:14 in sectors AR01-AR08 see Supplemental Material [S3], which also applied to the results of the rest 19 intervals).

We use the Quantile-Quantile plot (Q-Q plot) of Normal Probability Paper Test to test actual error sample distribution in this paper. It relies on the relationship between the quantile of the sample data and the quantile of the specified distribution to verify whether the data obey the normal distribution. If the relationship curve is a straight line, then the data obey the normal distribution, and otherwise it is not. A total of 320 hypothesis tests were performed on the above samples, and the results were found to be consistent with the normal distribution hypothesis. Due to space limitations, only 16 of the AR01-AR08 sectors test results for a 15-minute interval were selected and shown in Figure 8 (for details of this analysis see Supplemental Material [S2]).

5.2. Results Analysis. Probability distributions of every typical operation day from 07:00 to 11:59 of the sectors AR01-AR08 in central and southern China can be computed according to the prediction error distribution characteristics, probabilistic en route sector traffic demand prediction, and time predicted error parameter. Taking the predicted results in interval 07: 00-07: 14 from May 1, 2014, to May 24, 2014, of sector AR01 as an example, according to the maximum number of aircraft in sector AR01 which is denoted as M, the probability of number m aircraft occurring in the sector is shown in Table 3 (for details of this analysis see Supplemental Material [S3]).

As is shown in Table 3, the abscissa M represents the number of aircraft to be presented of the same interval in the future, which is demand predicted results that used deterministic method referred in Section 4.1. The ordinate m represents the possible number of aircraft and the values in the table mean the corresponding probability of m aircraft present in sector AR01 for M aircraft has been predicted in interval 07: 00-07: 14 of the same sector. Similar to Table 2, a total of 160 en route sector traffic demand and its probability distribution tables can be obtained. However, only 8 test results in interval 07: 00-07: 14 of sectors AR01-AR08 were selected due to space limitations, as shown in Figure 9 (for details of this analysis see Supplemental Material [S3]).

We calculated that the average flight time is 12 minutes of the central and southern regions based on the sample. According to the results, we enlarge each interval by increasing 12 minutes both forward and backward from 07: 00 to 11: 59. Then the corresponding traffic demand and its probability of each sector on May 25, 2014, could be obtained. The maximum predicted traffic demand is picked as the final values shown in Table 4 (for details of this analysis see Supplemental Material [S4]).

Correlation test results of the above predicted results and the actual ones are shown in Table 5 (for details of this analysis see Supplemental Material [S4]). It can be found that the correlation coefficient is more than 70% in busy period (such as 09: 00-11: 59) or sectors with large traffic flow (such as AR01-AR05), the highest or even 99%, which proved that the statistic law is quite effective. In contrast, the correlation coefficient of the idle sector or the smaller flow is not as ideal as the former one because of the small sample size, but it is acceptable overall.

Based on the predicted results above, we take the AR05 sector (the capacity is 15 per minute) as an example; both the deterministic and probabilistic traffic demand predicted results and actual value for each interval of period 07: 00-11: 59 are shown in Figure 10 (for details of this analysis see Supplemental Material [S4]). We also found the following.

(1) Because the deterministic prediction approach is limited by the prediction scale, the result is more accurate when the scale is small. The probabilistic prediction approach is more accurate in the peak period when the sample size is larger than normal.

(2) In the period 08: 45-08: 59, the probabilistic demand prediction results reflected that the traffic demand exceeded volume accurately, which would make the controllers predict the sector congestion easier. From 09:15 to 09: 29, the probabilistic results showed that demand did not exceed the volume, thus avoiding the false alarm caused by the deterministic results.

In general, the accuracy rate of probabilistic traffic demand prediction is 80%, which is 23.1% higher than the accuracy rate of deterministic traffic demand prediction.

6. Conclusions

Since most mature probabilistic traffic demand methods cannot be applied to data of current air traffic management system in China, in this paper, an approach for predicting

TABLE 1: Mean μ and variance σ of aircraft's predicted time error Δt_i^{in} of entering sector AR01-AR08 (min).

	AR01		AR02		AR03		AR04		AR05		AR06		AR07		AR08	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
07:00 -07:14	1.7	1.9	8.1	12.4	7.6	9.7	-70.7	112.2	6.3	18.4	6.4	1.1	6.2	0	3.3	7.5
07:15 -07:29	0.5	1.5	2.5	7.6	3.1	9.8	-23.0	70.7	4.7	7.5	5.6	0	1.5	10.3	4.2	0.9
07:30 -07:44	0.7	0.4	4.9	5.7	2.4	8.2	-13.4	24.0	8.4	17.7	9.5	3.5	2.4	10.9	5.6	3.6
07:45 -07:59	0.7	0.5	-0.1	11.4	5.6	10.7	-29.7	28.4	2.2	23.8	7.3	2.9	1.3	4.7	3.6	4.4
08:00 -08:14	0.6	1.3	4.7	5.2	6.0	10.2	-9.1	43.5	-2.1	24.2	0.3	6.3	2.3	9.8	-5.8	5.9
08:15 -08:29	1.8	1.5	0.7	10.4	3.0	9.8	5.6	13.1	2.6	19.2	-4.9	7.8	6.1	6.6	2.6	6.6
08:30 -08:44	2.2	1.8	8.5	18.3	3.2	9.4	-1.5	18.3	5.6	15.9	3.1	5.9	9.5	8.3	-0.4	5.5
08:45 -08:59	2.9	3.7	2.7	13.4	6.6	9.8	-6.1	25.9	2.8	20.9	4.0	4.7	3.8	11.6	-0.6	6.7
09:00 -09:14	6.8	4.3	3.2	10.1	7.8	8.5	0.6	15.4	-0.2	27.9	1.9	8.6	6.7	10.7	2.8	1.9
09:15 -09:29	4.4	4.9	6.6	9.1	6.7	8.7	3.4	17.0	3.6	15.5	3.8	1.8	8.0	11.6	1.0	1.4
09:30 -09:44	10.4	7.6	6.7	7.2	6.1	8.9	2.6	13.1	5.7	13.0	8.4	3.8	2.8	13.9	3.1	2.3
09:45 -09:59	3.7	5.7	4.5	9.0	4.3	8.6	-4.5	15.9	1.4	16.5	6.9	3.9	1.2	13.0	7.0	4.4
10:00 -10:14	1.3	5.9	1.6	9.7	4.63	8.3	1.6	12.9	5.3	15.7	7.5	4.7	3.3	13.4	5.8	3.7
10:15 -10:29	1.60	0.97	1.49	8.62	6.94	8.65	5.8	29.6	3.39	10.57	7.46	1.83	3.22	9.48	0.26	1.83
10:30 -10:44	7.7	7.7	1.5	11.1	4.3	6.9	2.7	9.3	4.7	24.8	-1.1	6.2	1.7	8.6	0.3	5.6
10:45 -10:59	4.2	4.5	0.7	9.9	3.5	7.9	1.1	12.4	2.9	11.5	2.1	4.4	4.2	9.4	-2.8	4.2
11:00 -11:14	1.88	2.56	-0.2	8.57	8.67	8.27	8.5	26.1	4.99	13.00	0.21	4.6	8.58	11.3	3.4	2.6
11:15 -11:29	3.1	3.0	1.0	9.3	4.6	8.3	-4.2	20.8	5.2	13.6	5.9	2.45	-2.99	18.09	6.59	7.16
11:30 -11:44	3.7	4.8	20.4	8.7	5.9	8.8	5.4	13.3	5.3	13.6	6.1	5.3	5.4	5.4	4.9	6.1
11:45 -11:59	2.0	1.4	2.2	10.0	4.9	8.4	5.7	20.5	4.8	13.4	5.4	5.8	0.7	7.6	-1.2	9.9

the time of flying over several sector boundary fixes and probabilistic en route sector traffic demand was proposed to analyze time predicted error distribution characteristics and influencing factors. Furthermore, the traffic demand probability distribution and the change regularities in a certain period of the sectors are obtained based on the actual operation data. In this way, the accuracy of probabilistic demand prediction is estimated to be 23.1% higher than that

TABLE 2: Mean μ and variance σ of aircraft's predicted time error Δt_i^{out} of leaving sector AR01-AR08 (min).

	AR01		AR02		AR03		AR04		AR05		AR06		AR07		AR08	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
07:00 -07:14	3.3	2.0	10.3	14.1	8.3	9.9	-70.8	112.6	6.5	18.4	8.6	1.6	9.9	0	6.2	1.1
07:15 -07:29	1.7	2.0	3.8	8.6	1.1	11.5	-23.0	70.7	5.5	7.9	6.8	0	1.6	10.3	6.4	1.4
07:30 -07:44	1.6	1.1	7.4	5.7	0.1	10.2	-13.3	24.1	7.30	14.3	10.9	3.5	2.4	10.9	8.2	3.5
07:45 -07:59	1.2	1.4	1.5	11.9	4.5	12.5	-29.5	28.7	1.2	21.2	8.8	2.6	2.5	4.9	5.1	5.4
08:00 -08:14	1.4	1.8	8.3	6.6	5.8	12.3	-9.1	43.6	-1.3	24.8	0.5	6.7	2.6	9.7	-5.7	5.9
08:15 -08:29	3.6	2.5	2.3	11.4	2.4	11.1	6.6	12.9	3.1	19.1	-4.8	8.0	6.9	6.8	2.7	6.6
08:30 -08:44	2.8	2.2	8.8	18.5	2.2	10.7	-1.2	18.5	5.8	15.8	3.1	5.9	9.7	8.4	-0.1	5.5
08:45 -08:59	3.77	3.84	3.17	13.7	5.66	11.2	-5.9	25.9	2.8	20.9	5.0	4.9	4.2	11.2	1.3	7.4
09:00 -09:14	7.2	4.4	4.2	10.4	7.3	9.6	0.9	15.5	-0.2	28.0	3.4	9.2	6.8	10.7	4.1	1.9
09:15 -09:29	4.9	4.7	7.7	9.7	5.9	9.7	3.5	16.9	3.7	15.7	5.3	2.0	7.9	11.4	2.2	1.6
09:30 -09:44	10.7	7.4	9.0	8.9	5.0	10.3	3.0	13.2	6.5	13.7	9.2	3.2	2.3	14.2	4.7	2.6
09:45 -09:59	4.2	5.4	5.6	10.0	2.6	10.2	-4.4	15.9	1.9	17.0	7.7	3.9	1.4	12.9	7.9	3.6
10:00 -10:14	1.9	6.1	1.7	9.8	3.9	9.2	1.8	12.9	4.9	13.4	10.7	8.2	3.0	12.1	8.2	4.3
10:15 -10:29	2.1	1.2	1.7	8.8	5.9	9.6	6.2	29.6	3.5	10.7	8.7	1.9	3.5	9.4	0.8	1.9
10:30 -10:44	8.2	7.3	1.9	11.5	4.9	8.2	2.8	9.4	4.9	24.9	-0.7	6.6	1.8	8.8	1.1	6.1
10:45 -10:59	5.4	4.6	1.0	10.0	2.5	9.7	1.4	12.4	3.1	11.7	2.9	4.9	4.3	9.7	-1.7	5.1
11:00 -11:14	2.8	3.0	0.5	9.4	6.1	9.2	8.8	26.1	5.3	13.3	1.3	5.2	8.2	11.9	4.9	2.4
11:15 -11:29	4.1	3.0	1.7	9.9	4.1	9.1	-4.0	20.9	5.6	13.9	7.3	2.5	-3.2	18.2	7.1	6.5
11:30 -11:44	4.4	4.5	21.1	8.7	5.7	9.3	5.7	13.2	5.6	13.9	6.1	4.9	5.8	5.54	5.8	5.9
11:45 -11:59	2.9	2.1	2.8	10.4	4.5	9.7	6.0	20.7	5.0	13.6	5.5	5.3	1.9	8.1	-0.8	10.1

of the traditional deterministic traffic demand prediction, which shows that this method can provide more proper and accurate analysis of traffic demand prediction for air traffic flow management.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Table 3: Traffic demand and probability distribution in interval 07: 00-07: 14 of sector AR01.

m	M															
	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0	15.0
0.0	1.00	0.98	0.79	0.65	0.54	0.49	0.46	0.37	0.30	0.25	0.23	0.16	0.13	0.09	0.08	0.06
1.0	—	0.02	0.21	0.31	0.37	0.38	0.39	0.40	0.40	0.38	0.37	0.33	0.30	0.25	0.22	0.19
2.0	—	—	0.00	0.04	0.09	0.11	0.13	0.18	0.22	0.25	0.26	0.29	0.30	0.30	0.29	0.28
3.0	—	—	—	0.00	0.01	0.01	0.02	0.04	0.07	0.10	0.10	0.15	0.18	0.21	0.23	0.24
4.0	—	—	—	—	0.00	0.00	0.00	0.01	0.01	0.02	0.03	0.05	0.07	0.10	0.12	0.14
5.0	—	—	—	—	—	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.03	0.04	0.06
6.0	—	—	—	—	—	—	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02
7.0	—	—	—	—	—	—	—	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8.0	—	—	—	—	—	—	—	—	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9.0	—	—	—	—	—	—	—	—	—	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10.0	—	—	—	—	—	—	—	—	—	—	0.00	0.00	0.00	0.00	0.00	0.00
11.0	—	—	—	—	—	—	—	—	—	—	—	0.00	0.00	0.00	0.00	0.00
12.0	—	—	—	—	—	—	—	—	—	—	—	—	0.00	0.00	0.00	0.00
13.0	—	—	—	—	—	—	—	—	—	—	—	—	—	0.00	0.00	0.00
14.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.00	0.00
15.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.00

Table 4: Probabilistic traffic demand prediction results in interval 07: 00-11:59 on May 25, 2014, of Sectors AR01-AR08 (number and probability abbreviated as N and P separately).

Interval	Sector															
	AR01		AR02		AR03		AR04		AR05		AR06		AR07		AR08	
	N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P
07:00-07:14	2	30%	0	28%	2	37%	1	33%	5	27%	1	19%	0	27%	1	18%
07:15-07:29	2	45%	1	25%	5	14%	3	48%	3	34%	1	27%	1	31%	2	19%
07:30-07:44	0	47%	2	33%	6	22%	0	34%	8	15%	1	28%	2	25%	1	23%
07:45-07:59	2	35%	3	17%	8	19%	4	29%	13	23%	2	31%	1	31%	0	31%
08:00-08:14	1	15%	3	21%	10	28%	2	25%	13	11%	0	25%	2	21%	1	28%
08:15-08:29	0	27%	4	39%	11	40%	5	37%	8	38%	2	37%	1	26%	3	31%
08:30-08:44	0	25%	6	14%	13	17%	7	56%	12	23%	3	22%	2	29%	0	17%
08:45-08:59	2	31%	7	21%	14	31%	12	28%	20	19%	1	27%	3	31%	2	27%
09:00-09:14	2	47%	6	35%	13	42%	10	35%	21	29%	1	38%	3	26%	2	30%
09:15-09:29	4	39%	8	41%	16	49%	8	24%	13	31%	2	23%	3	18%	1	23%
09:30-09:44	5	42%	9	11%	18	19%	8	26%	14	29%	1	13%	0	17%	1	32%
09:45-09:59	3	22%	9	26%	14	39%	3	29%	11	21%	1	29%	1	31%	0	16%
10:00-10:14	3	44%	11	30%	11	23%	6	34%	9	32%	1	34%	3	28%	1	40%
10:15-10:29	0	15%	7	18%	6	19%	8	26%	9	18%	0	37%	1	27%	2	19%
10:30-10:44	3	39%	10	42%	7	24%	4	21%	10	20%	1	29%	2	19%	1	23%
10:45-10:59	3	51%	9	47%	8	32%	1	15%	12	18%	2	31%	4	25%	0	38%
11:00-11:14	3	29%	10	28%	8	29%	4	23%	14	34%	1	27%	1	30%	4	18%
11:15-11:29	3	47%	11	18%	10	24%	3	34%	20	28%	2	34%	2	41%	2	32%
11:30-11:44	0	13%	8	31%	8	11%	6	19%	16	19%	1	23%	3	33%	2	27%
11:45-11:59	0	15%	11	23%	5	29%	6	38%	17	17%	1	21%	3	29%	1	13%

Table 5: Correlation test of traffic demand forecast of each sector in interval 07:00-11:59 on May 25, 2014.

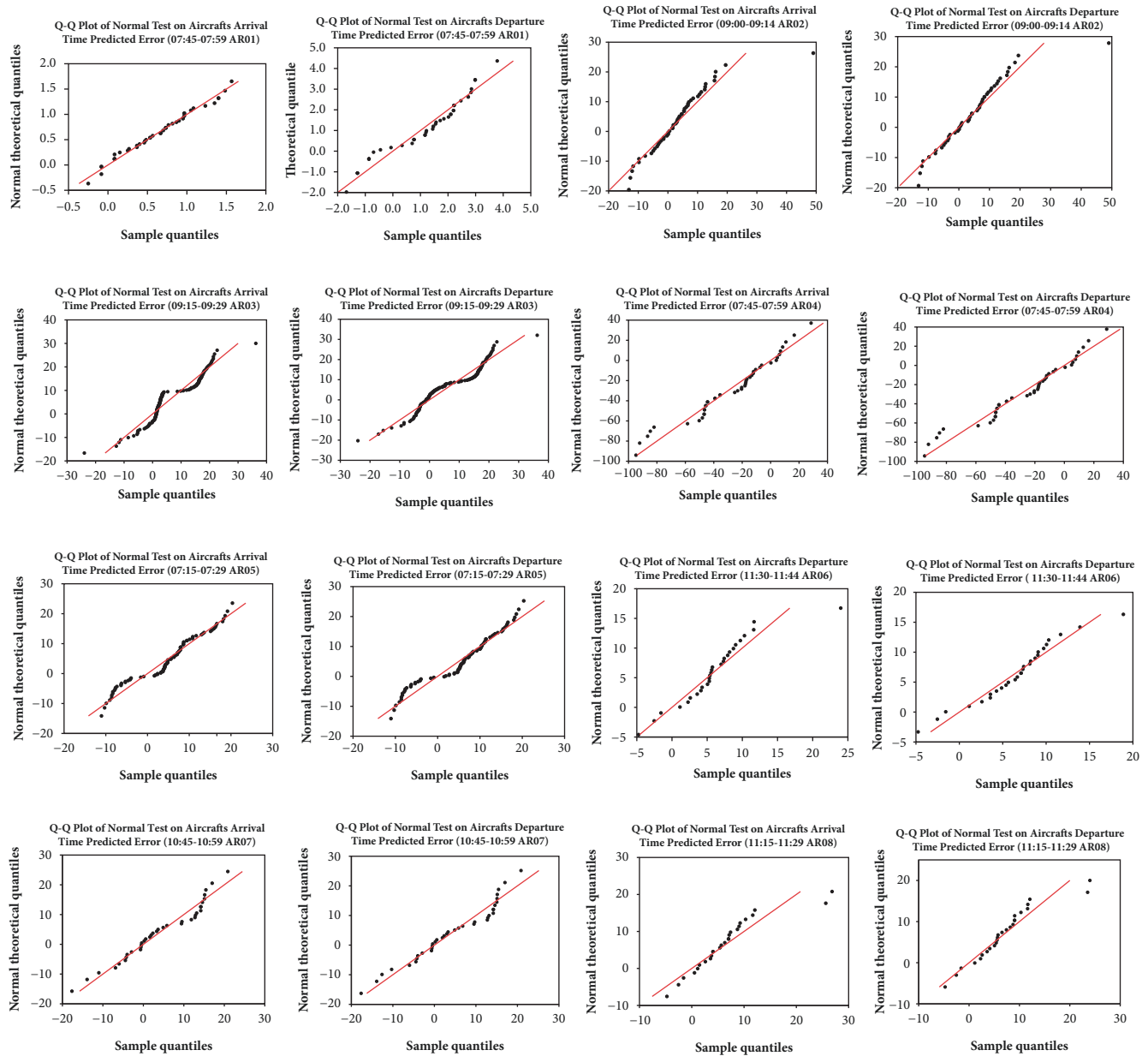| Interval	Sector							
	AR01	AR02	AR03	AR04	AR05	AR06	AR07	AR08
07:00-08:59	0.703146	0.916477	0.992406	0.984911	0.983829	0.667256	0.700649	0.695608
09:00-11:59	0.933251	0.982346	0.992271	0.989755	0.99632	0.87519	0.920918	0.915726

FIGURE 8: Q-Q plots of normal test on aircraft' arrival / departure time predicted error.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Wen Tian and Huiqing Xu conceived and designed the experiments. Wen Tian performed the experiments. Yixing Guo and Yi Yao analyzed the data. Bin Hu contributed analysis tools. Wen Tian and Huiqing Xu wrote the paper.

Acknowledgments

Supplementary Materials

The main dataset and corresponding results are provided as follows: [S1] Dataset:Fig.2 Fig.3.xlsx. provides the raw data of Figures 2 and 3. [S2] Results:Fig.8 AR(01-08).spv provides normal test results on aircrafts arrival/departure time predicted error for a 15-minute interval of AR01-AR08 sectors. [S3] Dataset and Results:Fig.9 AR(01-08)-0700-0714.xlsx provide raw data and statistical results in interval 07: 00-07: 14 of AR01-AR08 sectors. [S4] Results:Fig.10.xlsx provides correlation test results of AR01-AR08 sectors and probabilistic/actual values for 20 intervals from 07:00 to 11:59 of AR01-AR08 sectors. *(Supplementary Materials)*
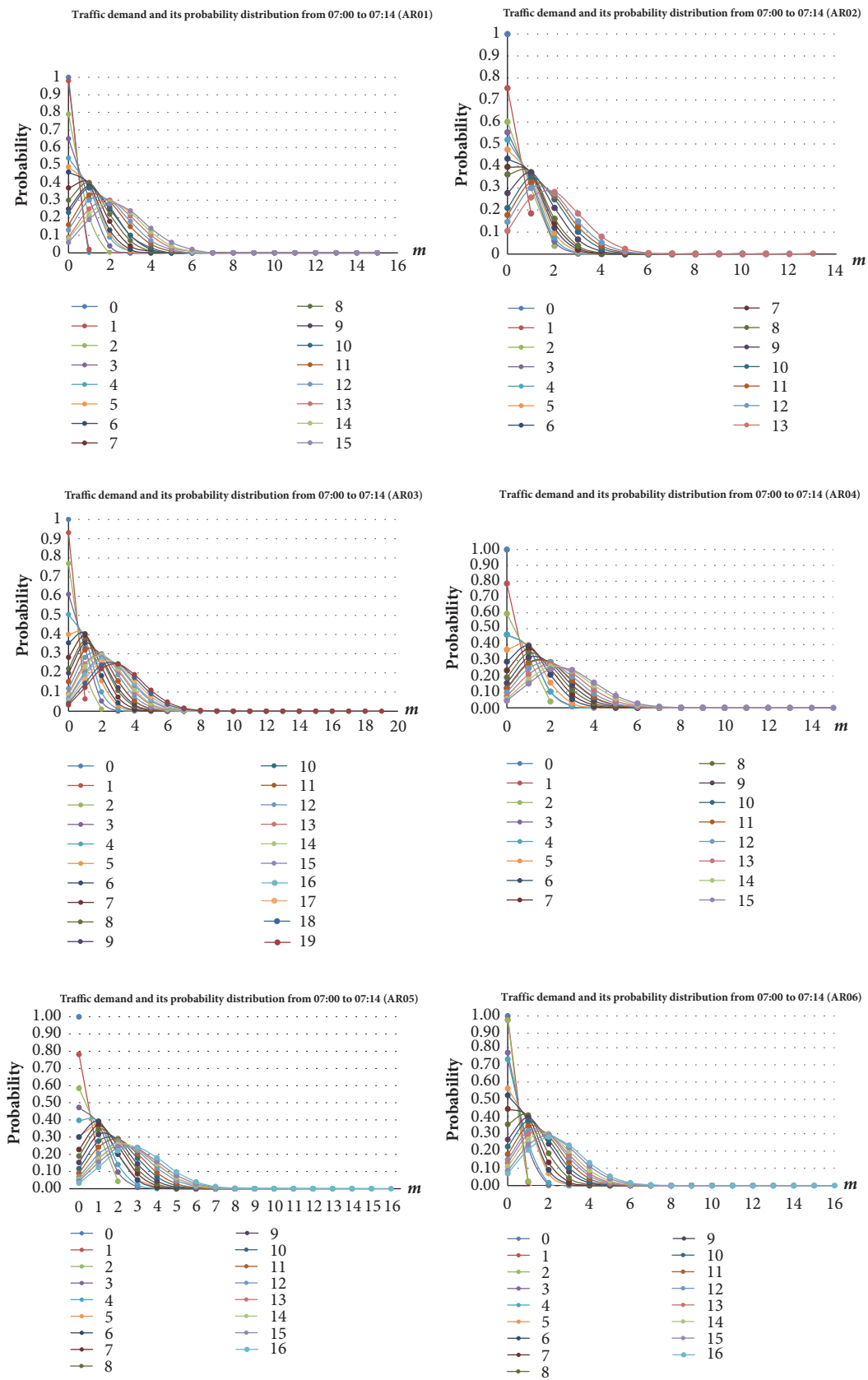
Traffic demand and its probability distribution from 07:00 to 07:14 (AR01)

Traffic demand and its probability distribution from 07:00 to 07:14 (AR02)

Traffic demand and its probability distribution from 07:00 to 07:14 (AR03)

Traffic demand and its probability distribution from 07:00 to 07:14 (AR04)

Traffic demand and its probability distribution from 07:00 to 07:14 (AR05)

Traffic demand and its probability distribution from 07:00 to 07:14 (AR06)

FIGURE 9: Continued.

FIGURE 9: Traffic demand and probability density function of each sector in interval 07:00-07:14.

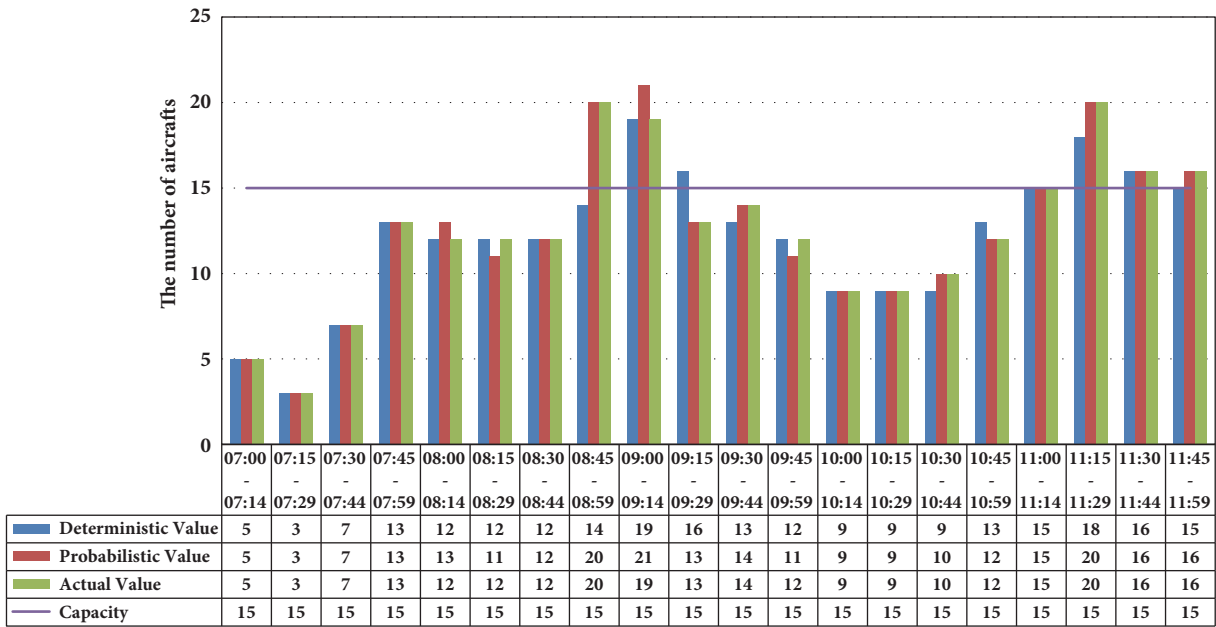|	07:00 - 07:14	07:15 - 07:29	07:30 - 07:44	07:45 - 07:59	08:00 - 08:14	08:15 - 08:29	08:30 - 08:44	08:45 - 08:59	09:00 - 09:14	09:15 - 09:29	09:30 - 09:44	09:45 - 09:59	10:00 - 10:14	10:15 - 10:29	10:30 - 10:44	10:45 - 10:59	11:00 - 11:14	11:15 - 11:29	11:30 - 11:44	11:45 - 11:59
Deterministic Value	5	3	7	13	12	12	12	14	19	16	13	12	9	9	9	13	15	18	16	15
Probabilistic Value	5	3	7	13	13	11	12	20	21	13	14	11	9	9	10	12	15	20	16	16
Actual Value	5	3	7	13	12	12	12	20	19	13	14	12	9	9	10	12	15	20	16	16
Capacity	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15

FIGURE 10: Comparison of the deterministic and probabilistic demand prediction results and comparison of the actual traffic flow value and capacity (Sector AR05).

References

[1] H. Vandevenne and M. A. Lippert, "Evaluation of Runway-Assignment and Aircraft-Sequencing Algorithms in Terminal Area Automation," *The Lincoln Laboratory Journal*, vol. 7, no. 02, pp. 215–238, 1994.

[2] D. H. Williams and St. M. Green, *Flight Evaluation of Center-TRACON Automation System Trajectory Prediction Process*, NASA, TP-1998-208439, Langley Research Center, Hampton Virginia, Va, USA, 1998.

[3] D. Bertsimas and S. S. Patterson, "The air traffic flow management problem with enroute capacities," *Operations Research*, vol. 46, no. 3, pp. 406–422, 1998.

[4] E. R. Mueller and G. B. Chatterji, "Analysis of aircraft arrival and departure delay characteristics," in *Proceedings of the AIAA's*

Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum, pp. 1–14, Los Angeles, Calif, USA, October 2002.

[5] C. Barnhart, P. Belobaba, and A. R. Odoni, "Applications of operations research in the air transport industry," *Transportation Science*, vol. 37, no. 4, pp. 368–391, 2003.

[6] P. Flener, J. Pearson, M. Ågren, C. Garcia-Avello, M. Çeliktin, and S. Dissing, "Air-traffic complexity resolution in multi-sector planning," *Journal of Air Transport Management*, vol. 13, no. 6, pp. 323–328, 2007.

[7] J. W. Taylor, "A quantile regression neural network approach to estimating the conditional density of multiperiod returns," *Journal of Forecasting*, vol. 19, no. 4, pp. 299–311, 2000.

[8] D. Bertsimas and S. S. Patterson, "Traffic flow management rerouting problem in air traffic control: a dynamic network flow approach," *Transportation Science*, vol. 34, no. 3, pp. 239–255, 2000.

[9] A. Mukherjee and M. Hansen, "A dynamic rerouting model for air traffic flow management," *Transportation Research Part B: Methodological*, vol. 43, no. 1, pp. 159–171, 2009.

[10] L. Yang, S. Yin, M. Hu, K. Han, and H. Zhang, "Empirical exploration of air traffic and human dynamics in terminal airspaces," *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 219–244, 2017.

[11] M. Larry, "Probabilistic methods for air traffic demand forecasting," in *Proceedings of the AIAA, The AIAA Guidance, Navigation, and Control Conference and Exhibit*, pp. 5–8, Monterey, Canada, 2002.

[12] K. T. Mueller, J. A. Sorensen, and G. J. Couluris, "Strategic aircraft trajectory prediction uncertainty and statistical sector traffic load modeling," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit 2002*, 11, 1 pages, Monterey, Canada, August 2002.

[13] S. Roy, B. Sridhar, and G. C. Verghese, "An aggregate dynamic stochastic model for air traffic control," in *Proceedings of the FAA & Eurocontrol, The 5th USA/Europe ATM 2003 R&D Seminar*, pp. 1–10, Budapest, Hungary, 2004.

[14] C. R. Wanke, M. B. Callaham, D. P. Greenbaum, and A. J. Masalonis, "Measuring uncertainty in airspace demand predictions for traffic flow management applications," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit 2003*, pp. 11–14, Austin, Tex, USA, August 2003.

[15] C. Wanke, S. Mulgund, D. Greenbaum, and L. Song, "Modeling traffic prediction uncertainty for traffic management decision support," in *Proceedings of the Collection of Technical Papers - AIAA Guidance, Navigation, and Control Conference*, pp. 2244–2258, Providence, RI, USA, August 2004.

[16] E. Gilbo and S. Smith, "A new model to improve aggregate air traffic demand predictions," in *Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit*, Hilton Head, SC, USA, 2007.

[17] E. Gilbo and S. Smith, "Probabilistic prediction of aggregate traffic demand using uncertainty in individual flight predictions," *Journal of Extracellular Vesicles*, vol. 3, no. 2, pp. 1–7, 2013.

[18] E. P. Gilbo and S. B. Smith, "New method for probabilistic traffic demand predictions for en route sectors based on uncertain predictions of individual flight events," in *Proceedings of the FAA& Eurocontrol, The Ninth USA/Europe Air Traffic Management Research and Development Seminar (ATM2011)*, pp. 1–11, Berlin, Germany, 2011.

[19] W. Tian and M. Hu, "Airspace sector probabilistic traffic demand prediction model," *Journal of Southwest Jiaotong University*, vol. 46, no. 2, pp. 340–346, 2011.

[20] C. Wang and L. Yang, "Probabilistic methods for airspace sector flow and congestion prediction," *Journal of Southwest Jiaotong University*, vol. 46, no. 1, pp. 162–166, 2011.

[21] S. Li, X. Xu, and F. Wang, "Risk prediction model and methodology of airport congestion based on probabilistic demand," *Journal of Southwest Jiaotong University*, vol. 48, no. 1, pp. 154–159, 2013.

[22] T. Wen, *Research on Probabilistic Airspace Congestion Management Based upon Uncertain Traffic Demand Prediction*, Nanjing University of Aeronautics and Astronautics, NanJing, China, 2012.

[23] F. Gonze, A. Simonetto, E. Huens, J. Boucquey, and R. M. Jungers, "Probabilistic occupancy counts and flight criticality measures for ATM," in *Proceedings of the 12th USA/Europe Air Traffic Management Research and Development Seminar, ATM 2017*, Seattle, Wash, USA, June 2017.

[24] D. Chen, M. Hu, H. Zhang, J. Yin, and K. Han, "A network based dynamic air traffic flow model for en route airspace system traffic flow optimization," *Transportation Research Part E: Logistics and Transportation Review*, vol. 106, pp. 1–19, 2017.

[25] C. Wanke and D. Greenbaum, "Incremental, probabilistic decision making for En route traffic management," *Air Traffic Control Quarterly*, vol. 15, no. 4, pp. 299–319, 2007.

[26] Y. Chang, S. Solak, J. B. Clarke, and E. L. Johnson, "Models for single-sector stochastic air traffic flow management under reduced airspace capacity," *Journal of the Operational Research Society*, vol. 67, no. 1, pp. 54–67, 2017.