

## Research Article

# Data-Driven Analysis of the Chaotic Characteristics of Air Traffic Flow

Zhaoyue Zhang <sup>1</sup>, An Zhang <sup>1</sup>, Cong Sun <sup>1</sup>, Shuaida Xiang <sup>2</sup> and Shanmei Li <sup>2</sup>

<sup>1</sup>School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China

<sup>2</sup>College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China

Correspondence should be addressed to Zhaoyue Zhang; zy\_zhang@cauc.edu.cn

Received 14 June 2020; Revised 31 July 2020; Accepted 3 September 2020; Published 18 September 2020

Academic Editor: Chunjiao Dong

Copyright © 2020 Zhaoyue Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Understanding the chaos of air traffic flow is significant to the achievement of advanced air traffic management, and trajectory data are the basic material for studying the chaotic characteristics. However, at present, there are two main obstacles to this task, namely, large amounts of noise in the measured data and the tedium of existing data processing methods. This paper improves the incorrect trajectory processing method based on ADS-B trajectory data and proposes a method by which to quickly extract the traffic flow through a certain waypoint. Currently, the commonly used theoretical analysis tools for nonlinear complex systems include the classical nonlinear dynamics analysis method and the newly developed complex network-based analysis method. The latter is currently in an exploratory stage because it has just been introduced into the study of air traffic flow. From these two perspectives, the chaotic characteristics of air traffic flow are studied in the present work. From the perspective of nonlinear dynamics, the improved C-C method is used to calculate the reliability parameters, namely, the time delay  $\tau$  and embedding dimension  $m$ , of phase-space reconstruction, and the maximum Lyapunov index is calculated by using the small data volume method to prove the existence of chaos in the system. From the perspective of complex networks, the construction of a visibility graph and horizontal visibility graph is used to prove the existence of chaos in the system, and the goodness-of-fit parameters of the degree distributions of two fitting methods under different time scales are evaluated, which provides support for the air traffic flow theory.

## 1. Introduction

With the growth of China's national economy, its civil aviation industry has also been greatly improved. Air traffic transportation is an important component of the modern transportation system, and the study of air traffic flow characteristics has gradually become a crucial task. Air traffic flow refers to the number of aircraft passing through a certain point or a certain route in a unit of time. China's air traffic flow is characterized by rapid growth and very uneven distribution. The dynamics of air traffic flow depend on the number and the length of air routes, the number of airports, the number of takeoffs and landings, etc. Air traffic flow is mainly concentrated in a few airports located in cities that are political, economic, and tourism centers, especially in the

more developed areas in the eastern part of China. With the rapid increase of domestic flights, the problem of flight delay has become increasingly more obvious, and the study of the chaotic characteristics of air traffic flow is beneficial to its management and control.

In recent years, many scholars have studied the nonlinear mechanism of traffic flow, especially by applying chaos theory to understand the dynamics of transportation systems. Chaos refers to the seemingly random and irregular phenomenon in deterministic nonlinear systems, but this kind of quasi-random phenomenon contains some laws and order. Chaotic systems are very sensitive to initial values, and a small change in the initial value will lead to a great difference in the long-term evolution results of the system. There exist various models for traffic flow prediction,

including the chaos theory, fractal theory, wavelet analysis, and catastrophe theory models. The premise of the application of the chaos theory model is that the system has chaotic characteristics. Thus, when short-term traffic flow is predicted based on the chaos theory model, it must be determined in advance whether there are chaotic characteristics in the traffic flow.

The research on the chaotic dynamic characteristics of ground traffic flow has received substantial attention within the last decade. In 1984, Disbro and Frame first introduced chaos theory into the traffic system [1]. Low and Addison investigated the concept of chaotic behavior in a deterministic car-following model [2]. Tang et al. improved the chaos forecasting method to be effective in forecasting conflicting traffic flows [3]. Based on the maximum Lyapunov index, Liu and Zhang determined the chaotic characteristics of traffic flow and realized the prediction of short-term traffic flow [4].

With the development of ground traffic flow theory, scholars began to explore the nonlinear characteristics of air traffic systems. The time series of air traffic flow based on measured data is an effective method by which its nonlinear characteristics can be studied. Frank et al. found that the Lyapunov exponent calculation provides a clear indication of chaos [5]. Time series are closely related to time scales, and there are great differences between time series extracted from different time scales. If the time scale is too small, it is difficult to determine the fluctuation of traffic flow; in contrast, if the time scale is too large, the macro-characteristics of traffic flow will be more reflected, and it is difficult to capture the detailed features. These are both mentioned in Zheng's research on the macro time series of air traffic flow under different time scales [6] and Wang's empirical analysis of time-dependent air traffic flow based on the Hurst index [7]. At present, the tool used most often for the analysis of the chaotic dynamic characteristics of air traffic flow is the classical nonlinear dynamic analysis method; however, there also exist new analysis methods based on complex networks.

*1.1. Nonlinear Dynamics Analysis Method.* Li et al. proposed an improved maximum Lyapunov exponential algorithm based on the small data volume method and wavelet noise reduction theory, identified the chaotic characteristics in flight conflict time series, and proved the feasibility of the application of chaos theory to flight conflict prediction [8]. Cong and Hu used time-series data of traffic flow in the airspace sector to study the chaotic characteristics of the air traffic system; they solved the correlation dimension and the maximum Lyapunov index of the traffic flow time series via the G-P algorithm and the small data volume method, thereby proving the existence of chaos and fractal characteristics in the air traffic flow time series [9]. Zheng used real trajectory data to study the chaotic characteristics of air traffic flow from both the macro- and microperspectives via the maximum Lyapunov index and recursive graphs [6]. Yang proposed a dynamic identification method for the chaotic characteristics of flow time series to solve the

timeliness problem of judging the chaotic characteristics of flow time series; he proved that, with the increase of data, flow time series gradually possess chaotic characteristics [10].

*1.2. Complex Network-Based Analysis Method.* Complex networks are a new network research theory characterized by partial or total self-organization, self-similarity, and attractors and by being small-world and scale-free. In recent years, with the rise of network science, the study of time series from the perspective of complex networks has attracted widespread attention. Some scholars began to use complex networks to analyze nonlinear time series. The analysis method based on complex networks is often used in the research on the fluctuation of gold prices and vegetable species prices. Researchers have also explored the degree distribution characteristics of networks constructed from air traffic flow time series. Yang and Yang proposed a reliable procedure for the construction of complex networks from the correlation matrix of a time series [11]; they found that the degree distribution of the original series can be well fitted with a power law, while that of the return series can be well fitted with a Gaussian function. Gao and Jin proposed a reliable method for the construction of complex networks from a time series with each vector point of the reconstructed phase space represented by a single node and edge determined by the phase space distance [12]. Donner et al. investigated in detail the corresponding potentials and limitations of networks based on recurrences in the phase space [13]. Pan et al. introduced a novel method for the condition diagnosis of complex systems in the chemical process industry with complex network-based time series analysis [14]. Scholars also gradually introduced complex networks into the study of traffic flow. For example, Tang et al. converted a traffic flow time series into a complex network and then analyzed the characteristics of traffic flow states from the perspective of the complex network [15]. In 2018, Wang and Zhu first analyzed air traffic flow time series based on complex network theory [16]; he examined the mechanical characteristics of air traffic flow from a new perspective, laying a foundation for the future application of other research results of complex network theory in the air traffic field.

This remainder of this paper is structured as follows. Section 2 explains the processing of trajectory data, proposes four types of incorrect trajectory data, and introduces the process of automatic dependent surveillance-broadcast (ADS-B) trajectory data collection. Moreover, it presents an improvement of the efficient identification and processing method of incorrect trajectory data, and the results of clustering and nonclustering when extracting the air traffic flow passing through a waypoint are compared. Section 3 extracts time series of different scales. Section 4 analyzes the chaotic characteristics of air traffic flow based on the classical nonlinear dynamic analysis method, while Section 5 analyzes the chaotic characteristics of air traffic flow based on complex networks. Finally, the conclusions of this work are drawn in Section 6.

## 2. Processing of ADS-B Trajectory Data

In this paper, an analysis method based on the measured data of the operating environment is adopted, and the data source is ADS-B trajectory data. ADS-B mainly carries out air-to-air and air-to-ground surveillance. In general, only airborne electronic equipment is needed; the GPS (Global Positioning System) receiver, data link transceiver, and antenna, and CDTI (cockpit display of traffic information) can complete the relevant functions without any ground support equipment, and an aircraft equipped with ADS-B can broadcast its precise location and other data (such as its speed and height and whether the aircraft is turning, climbing, falling, etc.) through the data chain. The ADS-B receiver, in combination with air traffic control systems and airborne ADS-B from other aircraft, provides accurate, real-time conflict information both in the air and on the ground. ADS-B is a new technology that redefines the three key elements of modern air traffic control, namely, communication, navigation, and surveillance.

The ADS-B provides detailed data on individual flights, such as the aircraft serial number, registration number, ICAO, flight call sign, latitude, longitude, altitude, speed, heading, vertical speed, receiver IP address, and flight time. Because the research object of this paper is air traffic flow, it is necessary to extract the time series of traffic flow from the trajectory data.

In this work, it is found that there are some incorrect trajectory data in the dataset. Incorrect trajectories originate from system errors in recording the trajectory data, which leads to the recording of noise. Erroneous trajectory data are different from abnormal trajectory data; anomalous trajectory data refer to the trajectory deviating from the prevailing traffic flow, not an incorrect trajectory. The existence of an abnormal trajectory reflects the randomness of air traffic operation and the influence of control measures on air traffic flow. If the trajectory data are not processed, it will result in difficulties in the subsequent analysis, as presented in Figure 1.

Therefore, before traffic flow identification, it is necessary to identify and process the incorrect trajectory data. In this section, based on the types and characteristics of the incorrect trajectory data found in the research process, related identification and processing methods are studied. There were mainly four types of incorrect trajectory data found in the research process of the present work, namely, scattered, multiconnected, misplaced, and defective trajectories. The scatter diagram and trajectory diagram of incorrect trajectories are presented in Figure 2.

As shown in Figure 2(a), the scattered trajectory is composed of a series of scattered points that may be far away from each other, and the connecting lines may misrepresent the movement state of the aircraft. There are few track points, which are far less than the average of the number of normal track points. Such trajectory data can be either point or noise data consisting of several straight lines.

As shown in Figure 2(b), the multiconnected trajectory is the erroneous connection of two trajectories. The original data may be confused with the same track during

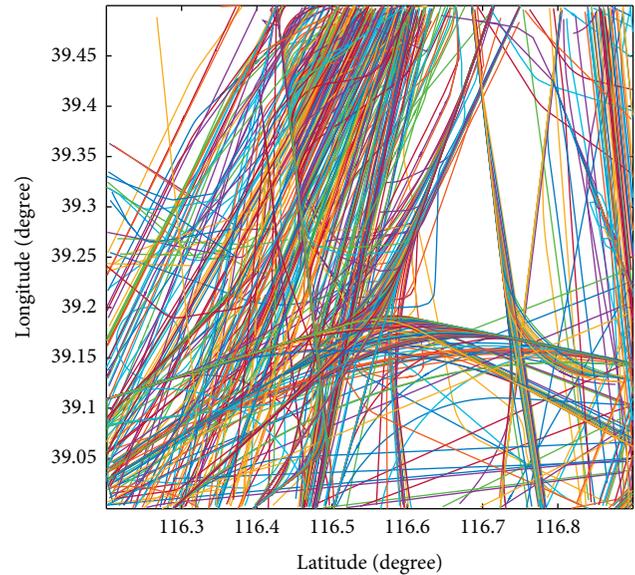


FIGURE 1: Original trajectory data.

transcription and encoding. Such trajectory data will create an extra line between two unrelated track points. Even if the points are filled by linear interpolation, the trajectory diagram will not change.

As shown in Figure 2(c), the misplaced trajectory is represented as an acute turning point, and the aircraft will not turn at an acute angle when it turns. When the three adjacent track points are misaligned, it will appear as an acute angle, which is represented as an obvious sharp bulge in the trajectory diagram.

As shown in Figure 2(d), the defective trajectory shows that the distance between two adjacent track points is large, which is caused by data loss.

Incorrect trajectory data cannot represent the correct trajectory of the aircraft; this will increase the number of trajectories and affect the data processing efficiency, which is not conducive to the identification of prevailing traffic flow. According to the characteristics of the incorrect trajectory data, four processing methods for this data are proposed in this paper. Because each piece of data is information acquired by radar, it has the significance of existence. To retain the original trajectory characteristics, it is necessary to deal with the incorrect trajectory data according to the principle of “more dismantling and less deletion;” the scattered trajectory data must be deleted, the multiconnected trajectory data must be split, the misplaced trajectory data must be smoothed, and the defective trajectory data must be supplemented.

The process diagram is presented in Figure 3.

Figure 4 presents a comparison of the effects before and after the processing of incorrect trajectory data from the entire day of December 14, 2019, over LASAN (a waypoint) in Shanghai. By comparison, it can be found that there are many lines on the left side of the figure that do not pass through the position of the prevailing traffic flow. These lines are the lines that connect two unrelated trajectories in multiple pieces of data. When multiple pieces of data are

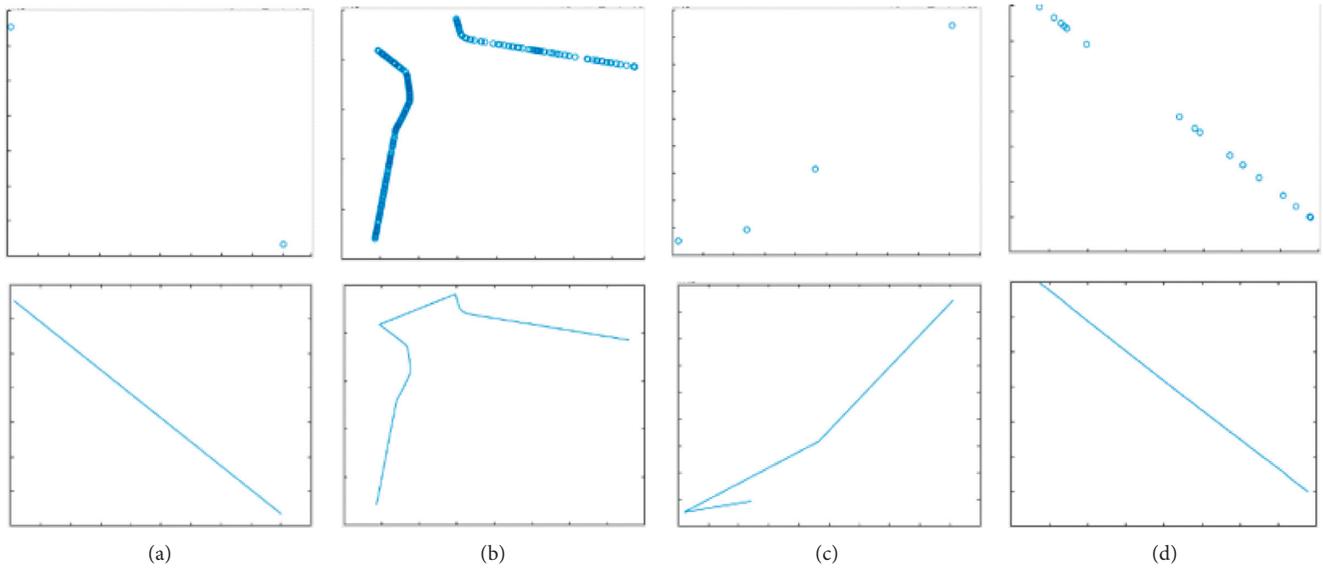


FIGURE 2: Scatter diagram and trajectory diagram corresponding to wrong trajectory. (a) Scattered trajectory. (b) Multiconnected trajectory. (c) Misplaced trajectory. (d) Defective trajectory.

split, the lines disappear. The comparison demonstrates that the traffic flow becomes clearer after the incorrect data are deleted, and the original features are still retained.

### 3. Extraction of Air Traffic Flow Time Series on Different Time Scales

Trajectory clustering is the main means by which traffic flow from chaotic trajectory data is identified; at present, the prevailing trajectory clustering method is adopted in most domestic and international research on air traffic flow identification. In this paper, the object of study for the traffic flow is on the air route, so the clustering of traffic flow in terminal areas need not be considered. Therefore, in this paper, a traffic flow identification method based on the air route network structure proposed by Zheng [6] is adopted; the air route network structure of air traffic flow is used for automatic identification, and the traffic flow is simultaneously determined via the waypoint sequence.

Taking the traffic flow over Laiyang City, Shandong Province, as an example, the trajectory data in the research area are processed and the traffic trajectory through P449 are extracted, as shown in Figure 5.

The complete clustering steps were found to be redundant for traffic flow extraction. For all the traffic flow data in Figure 5(a), if only extract the trajectory data whose closest distance to P449 is less than 10 km, it can be extracted as the traffic flow passing through the route point, as shown in Figure 6. The total traffic flow is 12% greater than the total traffic flow after clustering. Compared with the manual sorting of traffic flow sequences after pattern recognition, this not only realizes complete automation but is also more practical.

The air traffic flow time series  $T_f = [t_{f_1}, t_{f_2}, \dots, t_{f_i}, \dots, t_{f_{n'}}]$  can be obtained by calculating the number of

flights in a certain time scale, and the sequence length is  $n'$ . The traffic flow passing through P449 on December 14 is used as an example. Due to the low traffic flow during the period of 0:00 to 8:00 (PEK time), the traffic flow characteristics cannot be studied as a research sample. Therefore, the period of 8:00 to 24:00 is selected as the observation time, and the time series is constructed with four different time intervals of 2 min, 5 min, 10 min, and 15 min as the time scales, as shown in Figure 7.

By comparing the subgraphs in Figure 7, it can be determined that, although the time scales are different, similar peaks or troughs appear at the same time; however, the detailed characteristics of the time series are different. The reason for this is that larger time scales can smooth the data, but the resulting flow time series loses the internal details of the traffic flow. In this paper, a suitable time scale is selected to analyze the chaotic dynamics of air traffic flow.

### 4. Chaotic Characteristics of Air Traffic Flow Based on the Classical Nonlinear Analysis Method

**4.1. Judgment of Chaotic Characteristics.** In recent years, the chaotic time series analysis method has been widely used in many scientific studies and engineering fields. Phase-space reconstruction is the basis of chaotic time series analysis; according to this theory, the calculation of the time delay  $\tau$  and embedding dimension  $m$  [17] is the key problem of nonlinear time series phase-space reconstruction. For a given time series, such as the macroscopic traffic flow time series  $T_f = [t_{f_1}, t_{f_2}, \dots, t_{f_i}, \dots, t_{f_{n'}}]$ , it is assumed that the delay time is  $\tau$  and the embedding dimension is  $m$ . The  $m$ -dimensional vector  $X_n$  is the phase point produced by the phase space reconstruction as follows:

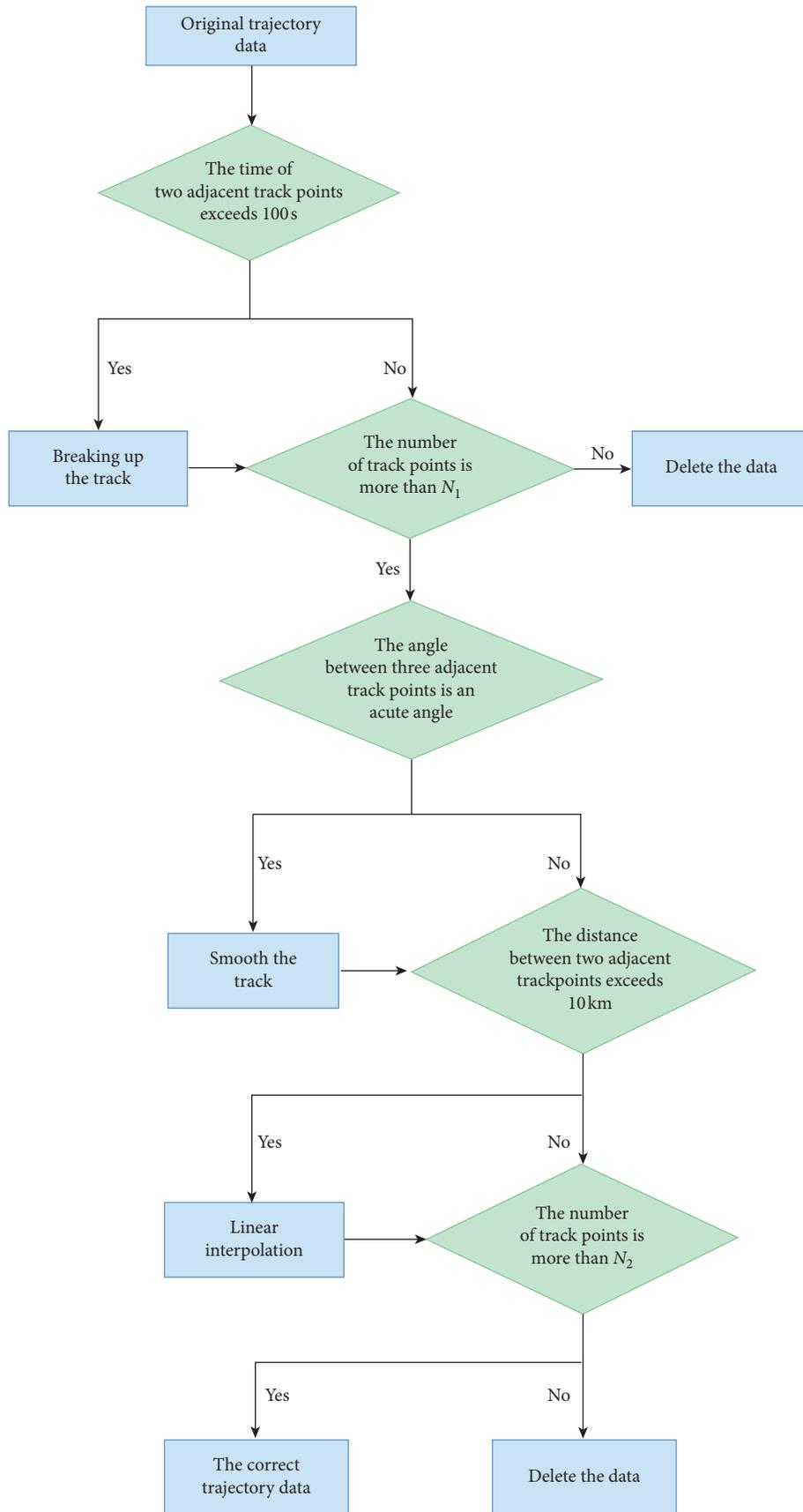


FIGURE 3: Wrong trajectory data processing algorithm flowchart.

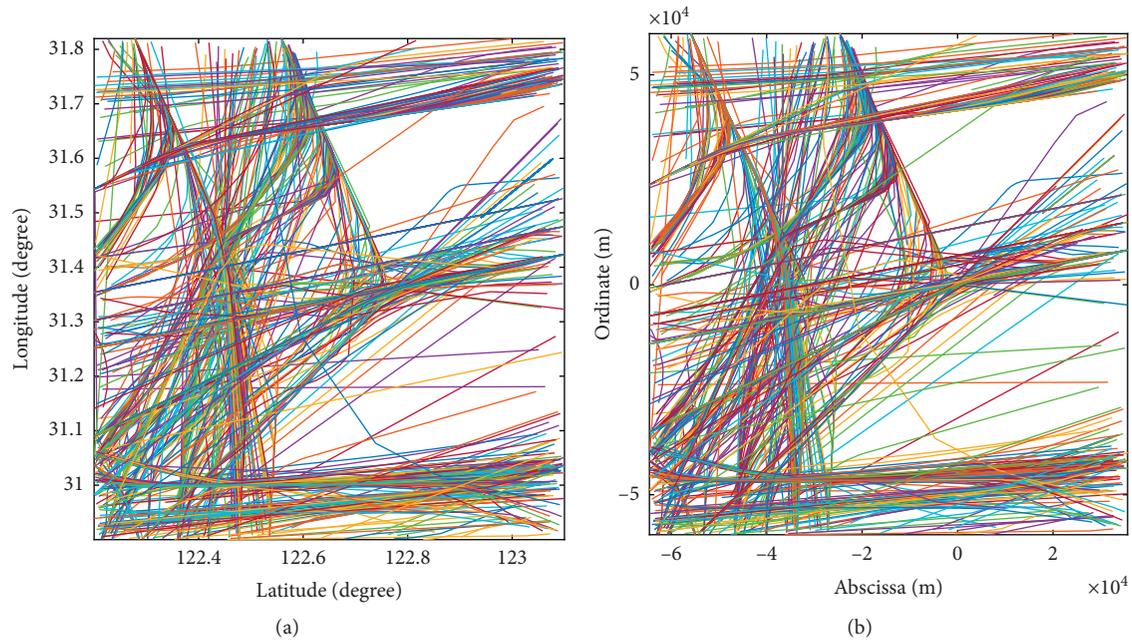


FIGURE 4: December 14, 2019, 122.2°–123.1° east longitude, 30.9°–31.82° north latitude trajectory data. (a) Before processing wrong. (b) After processing wrong.

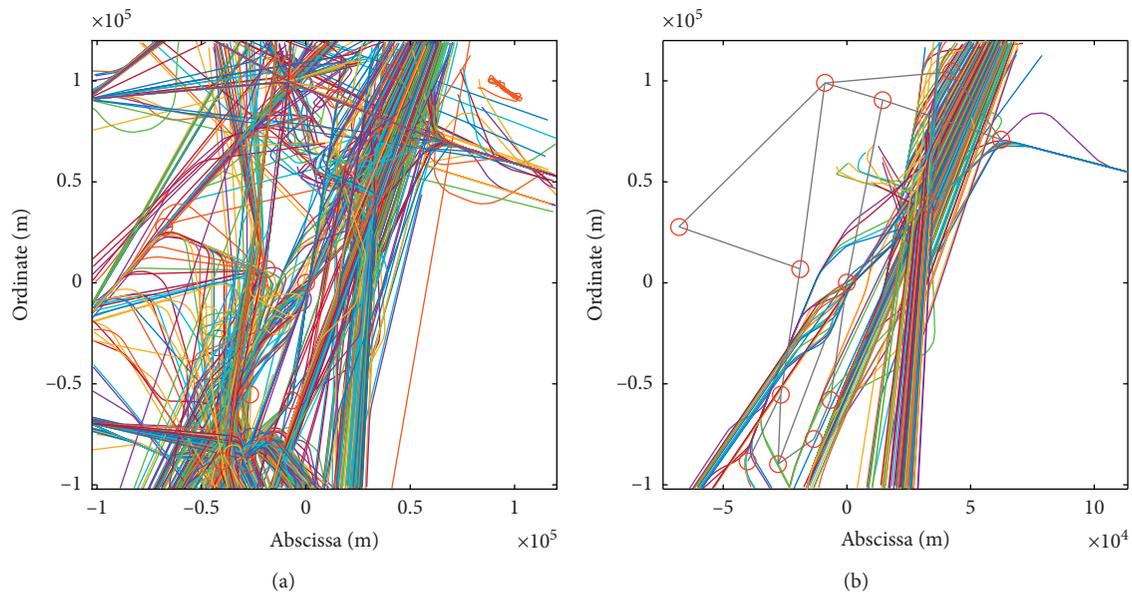


FIGURE 5: Traffic flow through P449 extracted on December 14, 2019. (a) All traffic flows over Laiyang on December 14, 2019. (b) Traffic flow through P449 extracted on December 14, 2019.

$$X_n = [t_{fn}, t_{f(n-t)}, t_{f(n-t)}, \dots, t_{f[n-(m-1)t]}], \quad (1)$$

$$n \in [(m-1)\tau + 1, (m-1)\tau + 2, \dots, N]. \quad (2)$$

Takens' theorem, which was put forward for infinitely long time series with the absence of noise, provided the foundation for the selection of the embedding dimension, and the time delay  $\tau$  may take any value [18]. However,

the system dynamics dimension  $d$  is unknown, and actual time series have a finite length and noise interference; therefore, the selection of the appropriate time delay  $\tau$  and embedding dimension  $m$  are the key to calculation. There are currently two main views regarding the selection of these values. One view holds that the two are mutually independent, for instance, in the time delay autocorrelation method [19], the average mutual

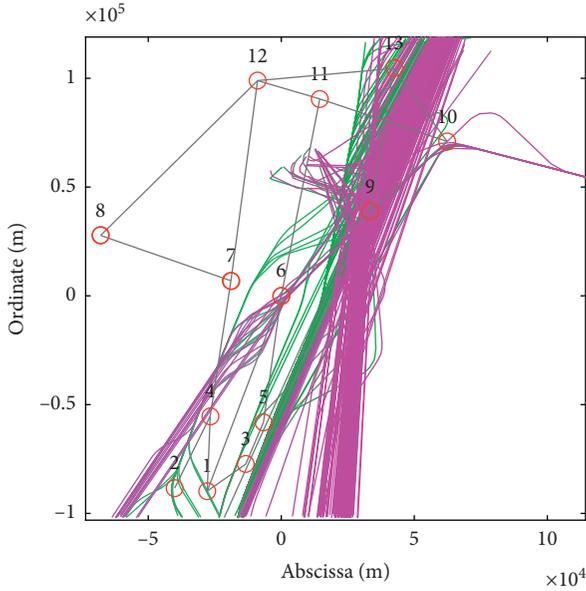


FIGURE 6: Traffic flow through P449 extracted on December 14, 2019.

information method [20], the embedded dimension G-P algorithm [21], the FNN method (false proximity method) [22], and the Cao method (the improvement of the false nearest proximity method) [23]. The other view holds that the two are related, for instance, in the embedded window method [24], C-C method [25], and improved C-C method [26].

In this paper, an improved C-C method is used to calculate these two parameters. The traditional methods have a certain advantage in the reconstructed phase space, while the improved method C-C method better conforms to the actual situation to achieve the estimates of the optimal time delay  $\tau$  and the optimal delay time window  $\tau_\omega$ . The improved method uses the following two processes to modify the associated parameters.

- (1) The time series standard deviation  $\sigma$ , which depends on the search radius  $r$ , is expanded to a certain extent according to the variation degree of the time series to reduce the interference caused by numerical oscillation.  $\sigma = \text{std}(x)(1 + cv/3)$  is utilized, where  $cv$  represents the coefficient of variation for discrete sequences and  $cv = \text{std}(x)/\text{mean}(x)$ .
- (2) The fixed value of  $r$  is changed to a function related to the embedded dimension  $m$ :  $r(m) = r \ln(m + 1)$ . With the change of the refactoring dimension, the search scope is appropriately expanded to reduce interference, especially high-dimensional data oscillations.

The remaining statistics are calculated according to the rules defined by the traditional C-C method. When analyzing and calculating the results, the first local minimum peak  $\tau_\omega$  of the cyclical nature of  $S_{2_{\text{cor}}}(t)$  is searched as the optimal delay time window. For nonobvious cyclical qualitative results, the value of  $S_{2_{\text{cor}}}(t)$  of the global minimum point is selected to determine the optimal delay time window

$S_{2_{\text{cor}}}(t)$ . Moreover, the value of  $\Delta \bar{S}_2(t)$  of the first local minimum point is used as the optimal time delay  $\tau$ .

The Lyapunov index is an important index to measure the dynamic characteristics of a system. It represents the average exponential rate of convergence or divergence between adjacent orbits in the phase space. The basic characteristic of chaotic motion is that the motion state is highly sensitive to the initial conditions. The orbitals generated by two very close initial values are separated exponentially over time, and the Lyapunov index is a quantitative description of this phenomenon.

If the Lyapunov index is greater than 0, the system is in a chaotic state. This judgment method is simple and definite and has little error. There are many methods to calculate the maximum Lyapunov index, including the definition method, Wolf method, small data volume method, Jacobian method, p-norm algorithm, and singular value decomposition method. For the study of time series, the Wolf method and the small data volume method are more applicable. Both methods require the calculation of the average period  $P$ , which can be obtained by using the Fourier transform to draw the power spectrum curve. In the case of a limited data volume, the small data volume method can better meet the computing needs [27].

#### 4.2. Chaotic Characteristics at Different Time Scales.

Based on the previously mentioned methods, the time series  $T_f$  of air traffic flow is reconstructed by the phase space  $X_n = [t_{f_n}, t_{f(n-\tau)}, t_{f(n-2\tau)}, \dots, t_{f[n-(m-1)\tau]}]$ . The maximum Lyapunov indices of the time sequence under different time scales presented in Section 3 are calculated based on the small data volume method, and the results are shown in Figure 8.

The maximum Lyapunov indices of the traffic flow time series in the sample at different time scales are all greater than zero ( $\lambda_{\text{max}} > 0$ ), indicating that there are chaotic characteristics in the air traffic flow. Via further comparison, it can be concluded that, with the increase of the time scale, the maximum Lyapunov index increases, indicating that the degree of chaos of the system increases. Consistent with previous research results [6], the length of the time scale has an important impact on the mining of traffic flow characteristics. The variation of the time scale will directly affect the morphological characteristics of traffic flow time series and will then affect the expression of system characteristics. If the time scale is too small, it is difficult to reflect the fluctuation of the traffic flow; if the time scale is too large, the overall rule of traffic flow variation throughout the day can be easily reflected, but it will be difficult to reveal the detailed characteristics of the traffic flow variation [6, 13].

## 5. Chaotic Characteristics of Air Traffic Flow Based on the Complex Network Theory

**5.1. Visibility Graph and Horizontal Visibility Graph of Time Series.** A complex network is a system composed of different individuals interacting with each other. The nodes correspond to units in the network, while an edge refers to the

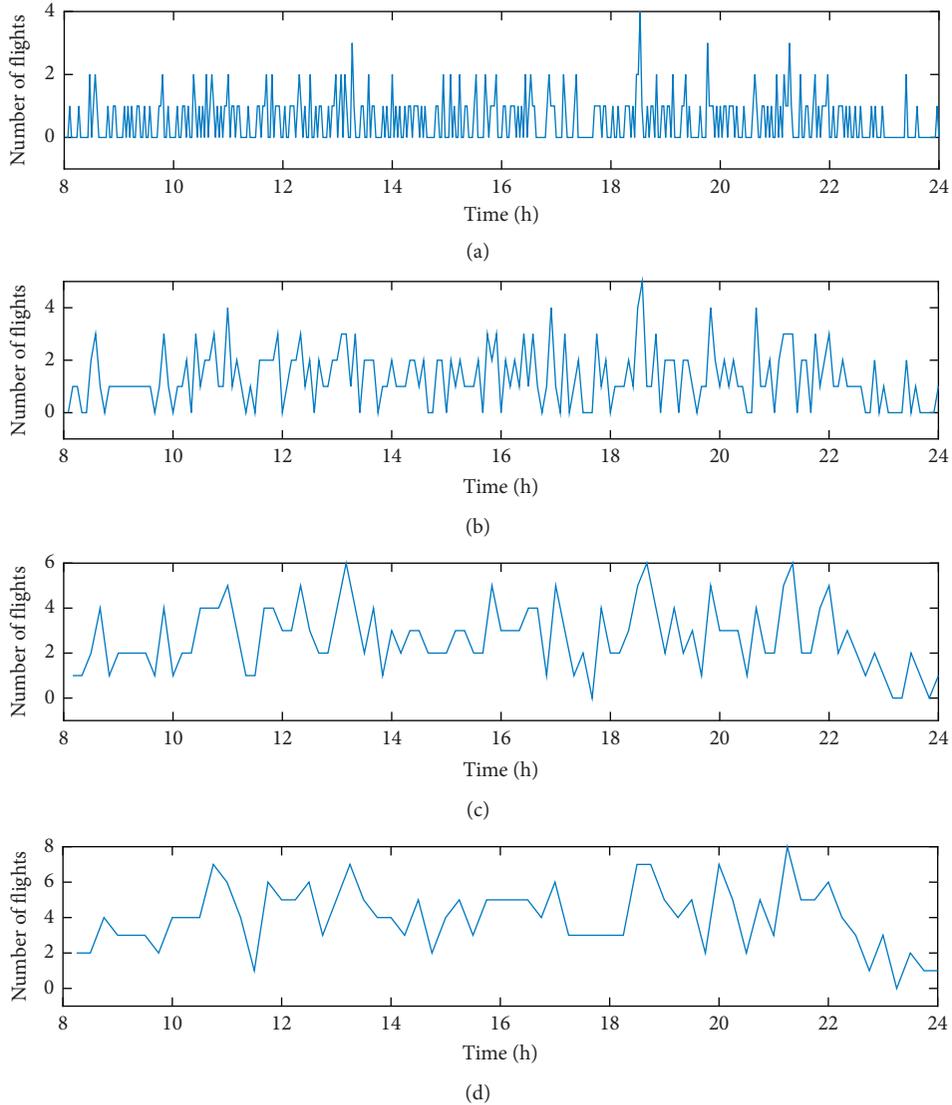


FIGURE 7: Different time scales of the time series. (a)  $\Delta t = 2$  min. (b)  $\Delta t = 5$  min. (c)  $\Delta t = 10$  min. (d)  $\Delta t = 15$  min.

connection between two nodes and represents the interaction between them. There are many statistical characteristics of complex networks, such as the degree distribution, accumulation coefficient, and community characteristics that are used to analyze the features contained in the original time series.

At present, complex network theory has become a new research direction for the use of complex networks as an analysis tool for the dynamic characteristics beneficial to the traffic flow theory, but also provide for time series analysis. To date, there has been little application of the complex network theory to the analysis of traffic flow time series. In this paper, complex networks are applied to the exploration of the characteristics of air traffic flow to expand support for actual traffic management and control.

In 2008, Lacasa et al. [28] proposed a new algorithm, namely, the visibility graph (VG) method, for the mapping of time series into complex networks. The principle of VG is

that each node in the network corresponds to each time point in the discrete time series data. The edge of the network is the connection of two points in these nodes that meet the visibility rules specified in advance. The basic idea of the algorithm is that each node cannot be connected to itself, and each line of sight cannot pass through other straight bars.

The visibility criterion is as follows. If any two data points  $(t_a, y_a)$  and  $(t_b, y_b)$  in the time series are visible to each other, then for any point  $(t_c, y_c)$ , where  $t_a < t_c < t_b$ , the following is satisfied:

$$y_c > y_a + \frac{(y_b - y_a)(t_c - t_a)}{t_b - t_a}. \quad (3)$$

With the development of VG research, Luque et al. [29] proposed a geometrically simpler and less statistical algorithm based on the VG algorithm in 2009, namely, the horizontal visibility graph (HVG) method.

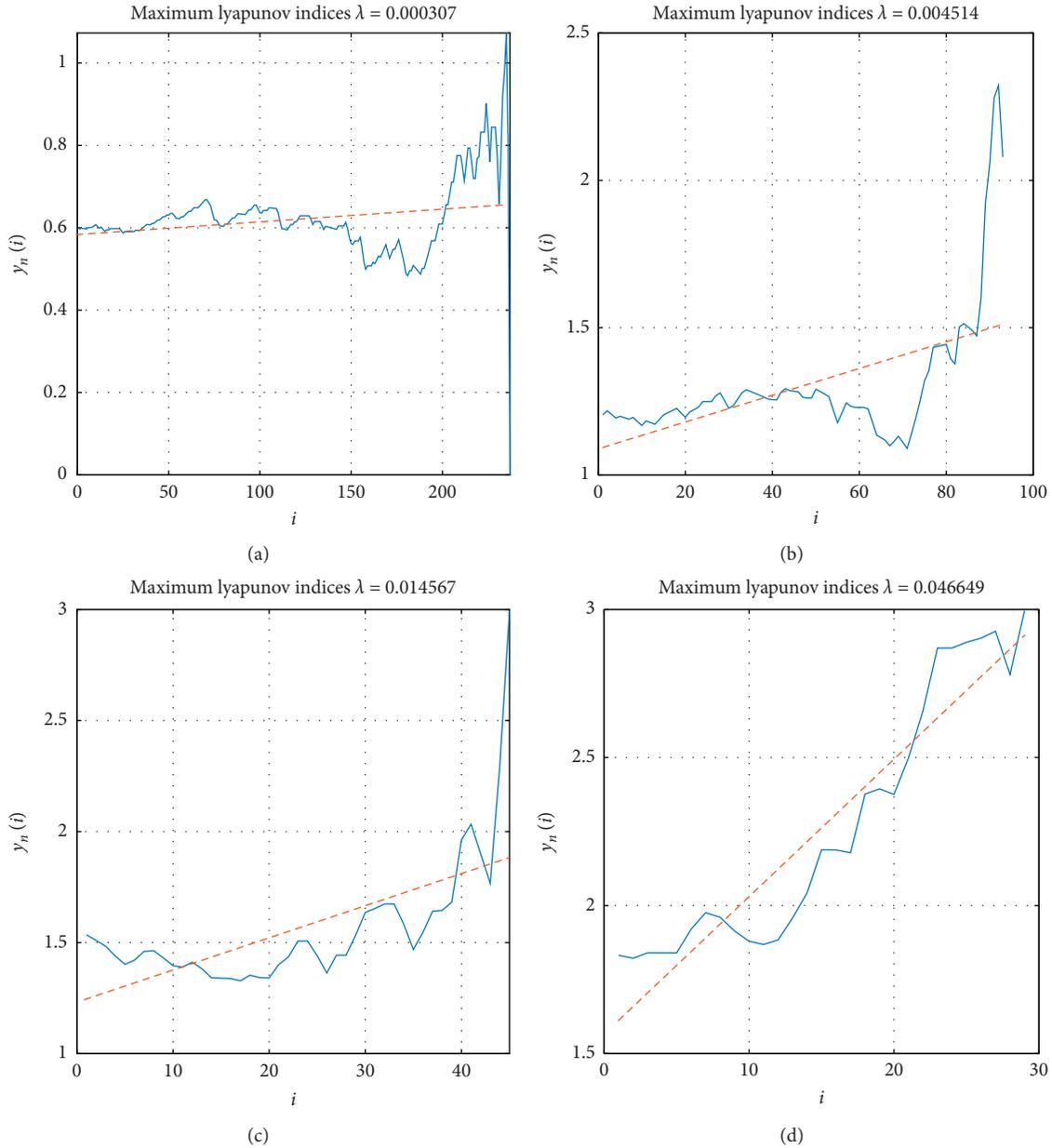


FIGURE 8: Analysis chart of the largest Lyapunov index of the sample. (a)  $\Delta t = 2$  min. (b)  $\Delta t = 5$  min. (c)  $\Delta t = 10$  min. (d)  $\Delta t = 15$  min.

The principle of HVG is expressed as follows. If a horizontal line of sight can be drawn between two straight bars without crossing other straight bars, the two points are considered to be connected in the network. In mathematical terms, the two data points are connected if the height of the corresponding bar is greater than that of the other bars located between the two points.

The visibility criterion is as follows. If any two data points in the time series  $X_n$ , such as  $(t_a, y_a)$  and  $(t_b, y_b)$ , are visible to each other, then for any  $(t_i, y_i)$ , where  $t_a < t_i < t_b$ , the following is satisfied:

$$\begin{aligned} y_i &< y_a, \\ y_i &< y_b. \end{aligned} \tag{4}$$

Taking the time series corresponding to the four time scales presented in Section 3 as an example, the time series are, respectively, mapped into visibility and horizontal visibility graphs, and the results are presented in Figure 9.

### 5.2. Analysis of the Degree Distributions of Complex Networks.

The degree distribution  $p(k)$  is defined as a probability of the random selection of nodes with degree  $k$ . In reality, the degree distributions of many complex networks follow a power-law distribution:  $p(k) = k^{-\lambda}$ . Lacasa and Toral [30] demonstrated that random and chaotic time series of degree distributions of complex networks under the HVG algorithm present exponential distribution forms:  $p(k) \sim e^{-\gamma k}$ .

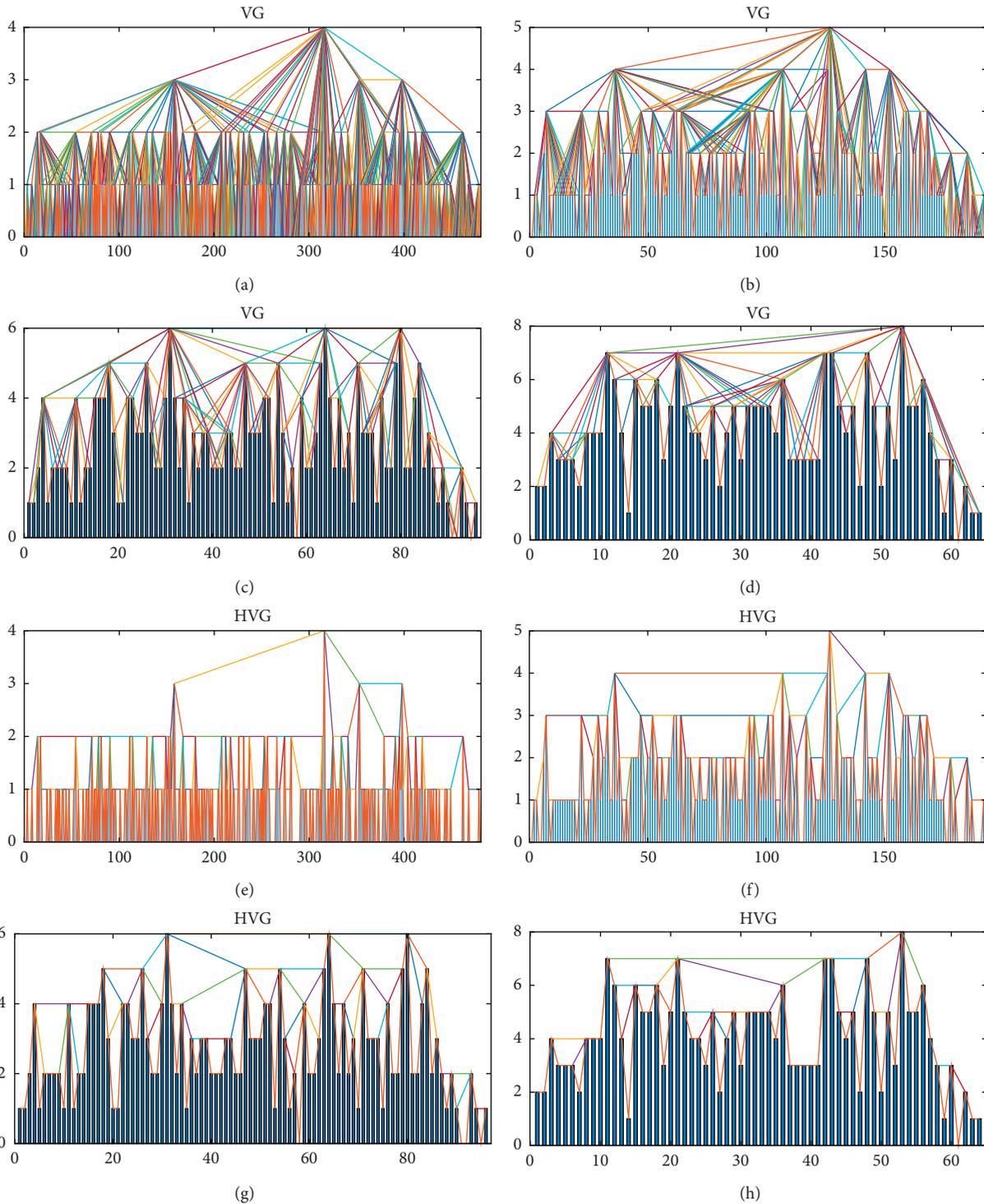


FIGURE 9: Visual graph and horizontal visual graph of the sample. (a)  $\Delta t = 2$  min. (b)  $\Delta t = 5$  min. (c)  $\Delta t = 10$  min. (d)  $\Delta t = 15$  min. (e)  $\Delta t = 2$  min. (f)  $\Delta t = 5$  min. (g)  $\Delta t = 10$  min. (h)  $\Delta t = 15$  min.

The VG and HVG in Figure 9 are used as examples to draw diagrams of the degree distributions, as presented in Figures 10 and 11.

In his research on the visual network analysis of non-linear time series, Zhang [31] mentioned that the effect of the oblique line fitting of degree distributions can be described

via the correlation coefficient  $R^2$ , the residual squared, and the sum of squares due to error (SSE). The smaller the value of SSE, the closer  $R^2$  is to 1, and the higher the reference value of relevant coincidences, the better the fitting. The exponential distribution in a single logarithmic coordinate system is approximately a straight line. According to the

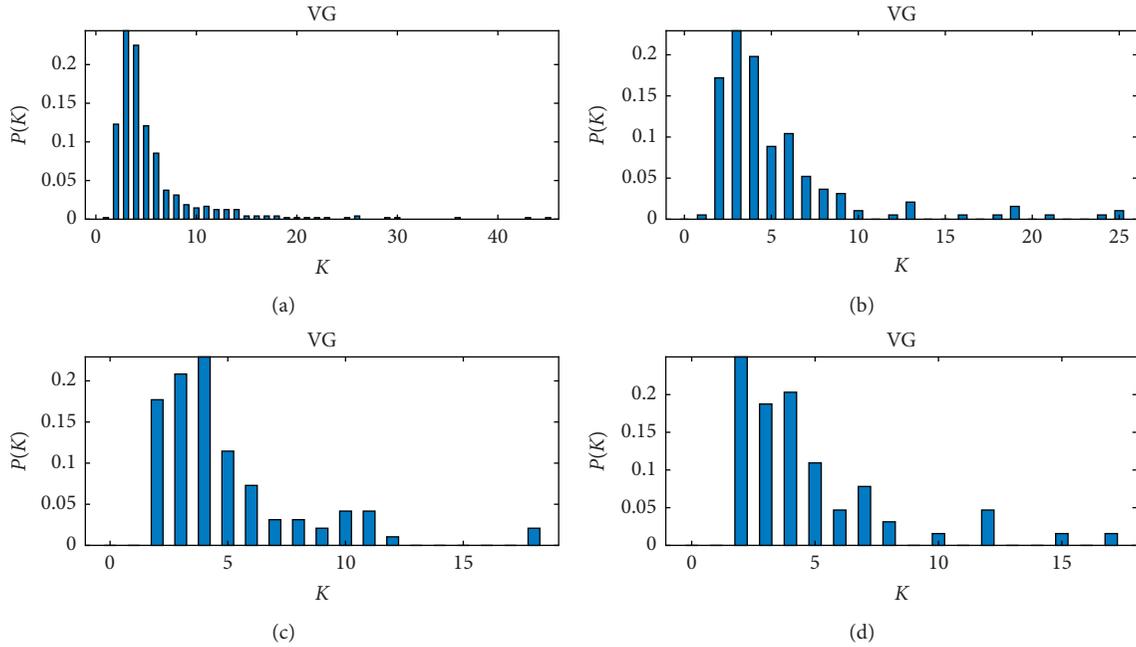


FIGURE 10: Degree distribution of visual graph. (a)  $\Delta t = 2$  min. (b)  $\Delta t = 5$  min. (c)  $\Delta t = 10$  min. (d)  $\Delta t = 15$  min.

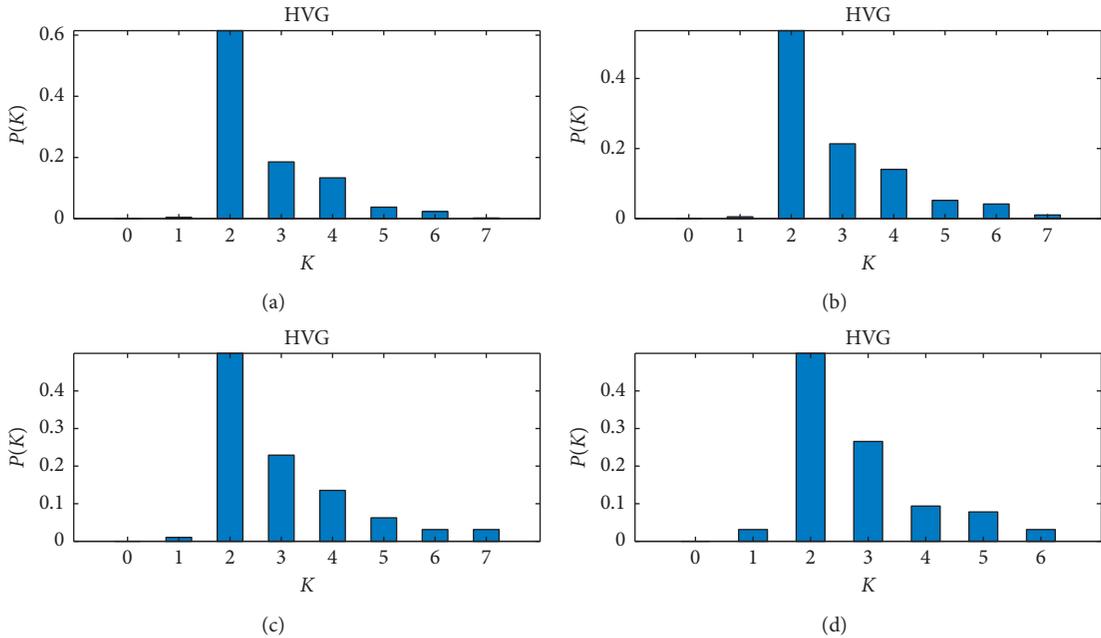


FIGURE 11: Degree distribution of horizontal visual graph. (a)  $\Delta t = 2$  min. (b)  $\Delta t = 5$  min. (c)  $\Delta t = 10$  min. (d)  $\Delta t = 15$  min.

steps outlined previously, the results of the four views in Figure 10 after fitting with the power-law distribution and exponential distribution are, respectively, calculated, as reported in Figure 12 and Table 1.

Similarly, according to the steps outlined previously, the results of the four horizontal visibility graphs in Figure 11 are calculated after fitting with the power-law distribution and exponential distribution, as reported in Figure 13 and Table 2.

In addition to  $R^2$  and SSE, the mean squared error (MSE) is also used to evaluate the fitting effect. MSE is the mean sum of the squares of the corresponding point errors between the predicted data and the original data, i.e.,  $SSE/n$ , and the root-mean-squared error (RMSE) is the square root of the MSE. When the original data increase, the representativeness of SSE will be affected, and the results of MSE and RMSE will be more convincing.

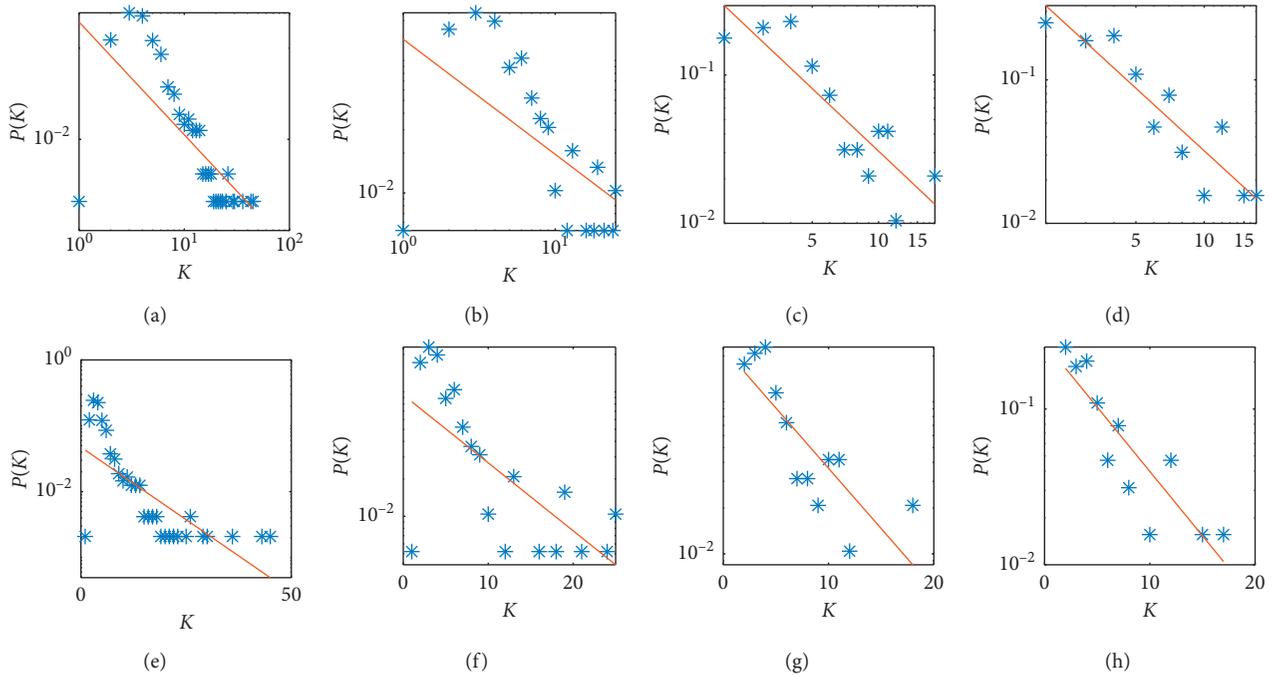


FIGURE 12: Results after power law distribution and exponential distribution fitting. (a)  $\Delta t = 2$  min. (b)  $\Delta t = 5$  min. (c)  $\Delta t = 10$  min. (d)  $\Delta t = 15$  min. (e)  $\Delta t = 2$  min. (f)  $\Delta t = 5$  min. (g)  $\Delta t = 10$  min. (h)  $\Delta t = 15$  min.

TABLE 1: Statistics of goodness of fit parameters.

Figure	Fitting method	$\lambda/\gamma$	SSE	$R^2$	RMSE
(a)	Power law distribution	-1.236059	37.3785	0.4913	1.1554
(b)	Power law distribution	-0.868055	23.3142	0.3079	1.2071
(c)	Power law distribution	-1.403329	2.7372	0.7364	0.5232
(d)	Power law distribution	-1.446997	1.7388	0.8448	0.4395
(e)	Exponential distribution	-0.101365	33.7598	0.5406	1.098
(f)	Exponential distribution	-0.125702	18.0168	0.4651	1.0612
(g)	Exponential distribution	-0.182581	3.8681	0.6556	0.6219
(h)	Exponential distribution	-0.190391	2.4722	0.7795	0.5241

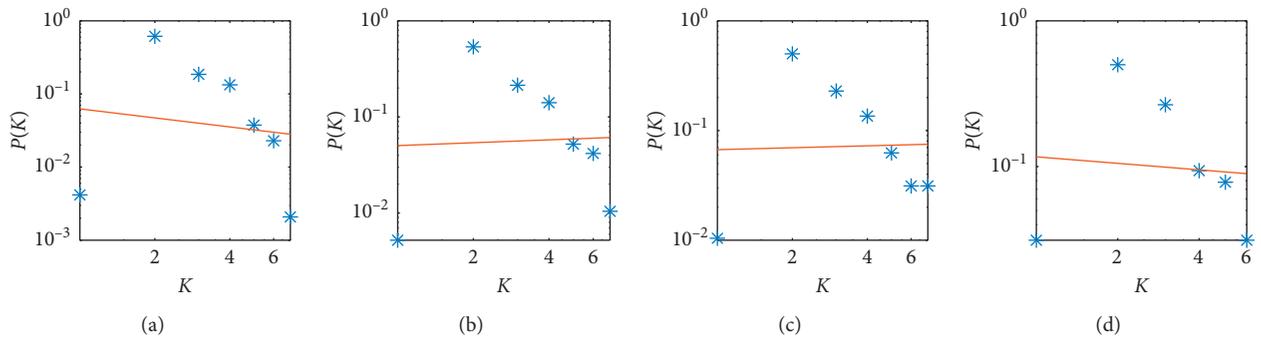


FIGURE 13: Continued.

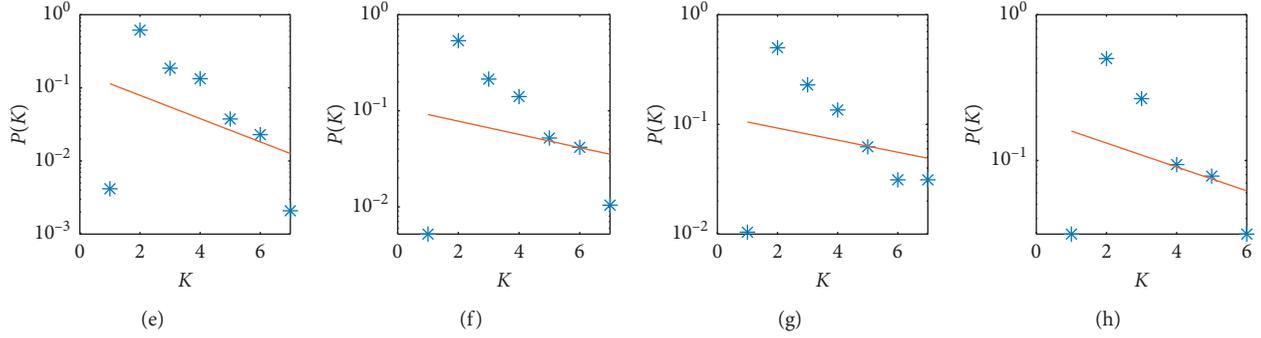


FIGURE 13: Results after power law distribution and exponential distribution fitting. (a)  $\Delta t = 2$  min. (b)  $\Delta t = 5$  min. (c)  $\Delta t = 10$  min. (d)  $\Delta t = 15$  min. (e)  $\Delta t = 2$  min. (f)  $\Delta t = 5$  min. (g)  $\Delta t = 10$  min. (h)  $\Delta t = 15$  min.

TABLE 2: Statistics of goodness of fit parameters.

Figure	Fitting method	$\lambda/\gamma$	SSE	$R^2$	RMSE
(a)	Power law distribution	-0.410612	24.9337	0.0187	2.231
(b)	Power law distribution	-0.099414	16.2789	0.0017	1.8044
(c)	Power law distribution	-0.058225	10.6363	0.0009	1.4585
(d)	Power law distribution	-0.148364	6.2668	0.0077	1.2517
(e)	Exponential distribution	-0.366281	21.6516	0.1478	2.0809
(f)	Exponential distribution	-0.158647	15.6019	0.0432	1.7665
(g)	Exponential distribution	-0.126737	10.1961	0.0422	1.428
(h)	Exponential distribution	-0.188867	5.6909	0.0988	1.1928

By comparing the results in Tables 1 and 2, it can be found that, for the VG algorithm, when the time scale is 2 min and 5 min, the degree distributions are closer to the exponential distribution; however, at this time,  $R^2$  is too small, and the fitting results are not reliable. When the time scale is 10 min and 15 min, the distributions of the time degree are closer to the power-law distribution, and  $R^2$  approaches 1; thus, the fitting results are reliable. However, for the HVG algorithm, the distribution of the uniform degree is closer to the exponential distribution. Additionally, with the increase of the time scale, the RMSE is smaller, but the results are unsatisfactory; moreover,  $R^2$  is too close to 0, and the fitting results are very unreliable.

Inspired by the existing VG and HVG algorithms, this paper proposes a new visual graphic network construction method, namely, the equal-height horizontal visibility graph (EHHVG) method. The proposed EHHVG differs from the HVG in that a horizontal visual line can be drawn between two vertical bars and can pass through a vertical bar of the same height. The visibility criterion is as follows. For any two points, such as  $(t_a, y_a)$  and  $(t_b, y_b)$  that are visible in the time series  $X_n$ , for any  $(t_k, y_k)$ , where  $t_a < t_k < t_b$ , the following is satisfied,

$$\begin{aligned} y_k &\leq y_a, \\ y_k &\leq y_b. \end{aligned} \quad (5)$$

According to the new criterion, four EHHVG maps of the sample and their degree distributions are redrawn, and the results are presented in Figures 14 and 15.

The results of the fitting of the degree distributions are reported in Figure 16 and Table 3.

As is evident from Table 3, the fitting results are not ideal for the time scales of 2 min and 5 min, while they are good for the time scales of 10 min and 15 min. Compared with the degree distribution of the HVG presented in Table 2, the fitting results are more reliable. When the time scale is too small, the degree distribution points are very scattered, which is similar to the fitting results of the degree distribution of the VG reported in Table 1.

By comparing the results of the VG, HVG, and EHHVG network construction methods presented in Figures 9 and 14, it is clear that the connectivity between each node in the VG and EHHVG is relatively complex. By comparing Figures 10 and 15, the degree distribution results of the VG and EHHVG are also more dispersed, but there are some differences between the performances of various algorithms and the shape of air traffic flow after time-series feature recognition.

The conclusions of this section can be drawn as follows by comparing the goodness-of-fit parameters.

- (1) In terms of the goodness of distribution fitting of the three complex network construction methods, the time series under the time scales of 10 min and 15 min are better than those under the time scales of 2 min and 5 min.
- (2) The time series of air traffic flow are transformed into more concentrated and simple degree distributions of HVG.

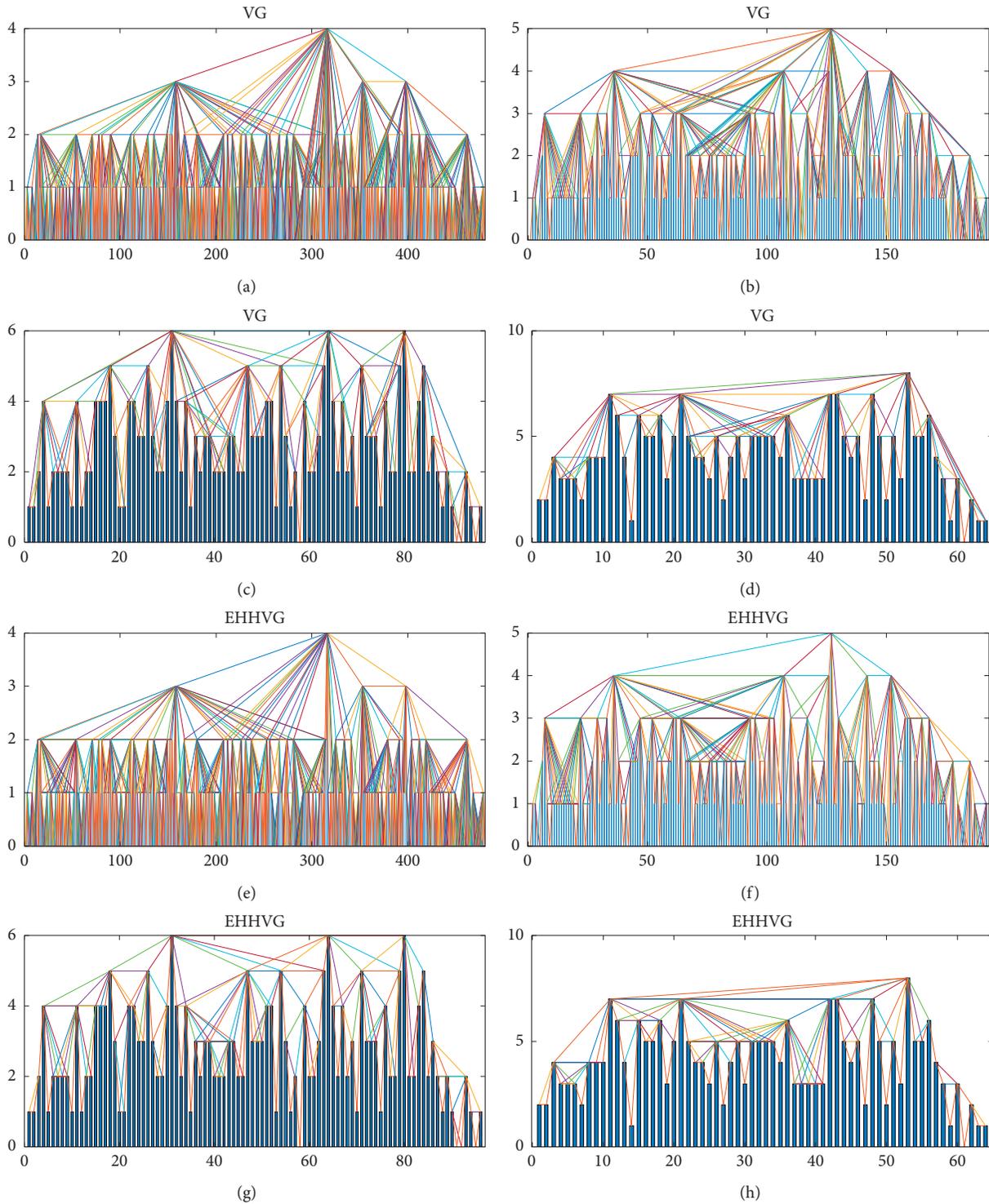


FIGURE 14: Equal height horizontal visibility graph of the sample. (a)  $\Delta t=2$  min. (b)  $\Delta t=5$  min. (c)  $\Delta t=10$  min. (d)  $\Delta t=15$  min. (e)  $\Delta t=2$  min. (f)  $\Delta t=5$  min. (g)  $\Delta t=10$  min. (h)  $\Delta t=15$  min.

(3) The fitting results of the power-law distribution and exponential distribution of the time series of the EHHVG algorithm are superior to those of the HVG method.

5.3. *Chaotic Characteristics at Different Time Scales.* Zhou et al. [32] found that the degree distribution of a complex network converted by the VG algorithm from chaotic time series presented an irregular multi-peak structure, but this

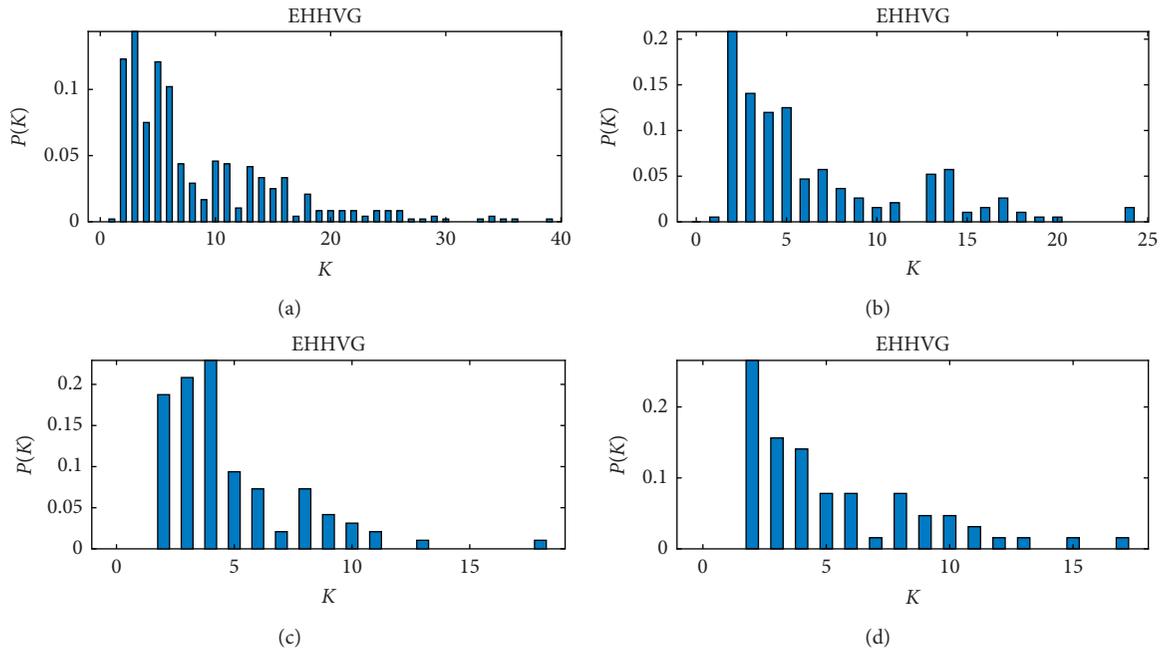


FIGURE 15: Degree distribution of visual graph and horizontal visual graph. (a)  $\Delta t = 2$  min. (b)  $\Delta t = 5$  min. (c)  $\Delta t = 10$  min. (d)  $\Delta t = 15$  min.

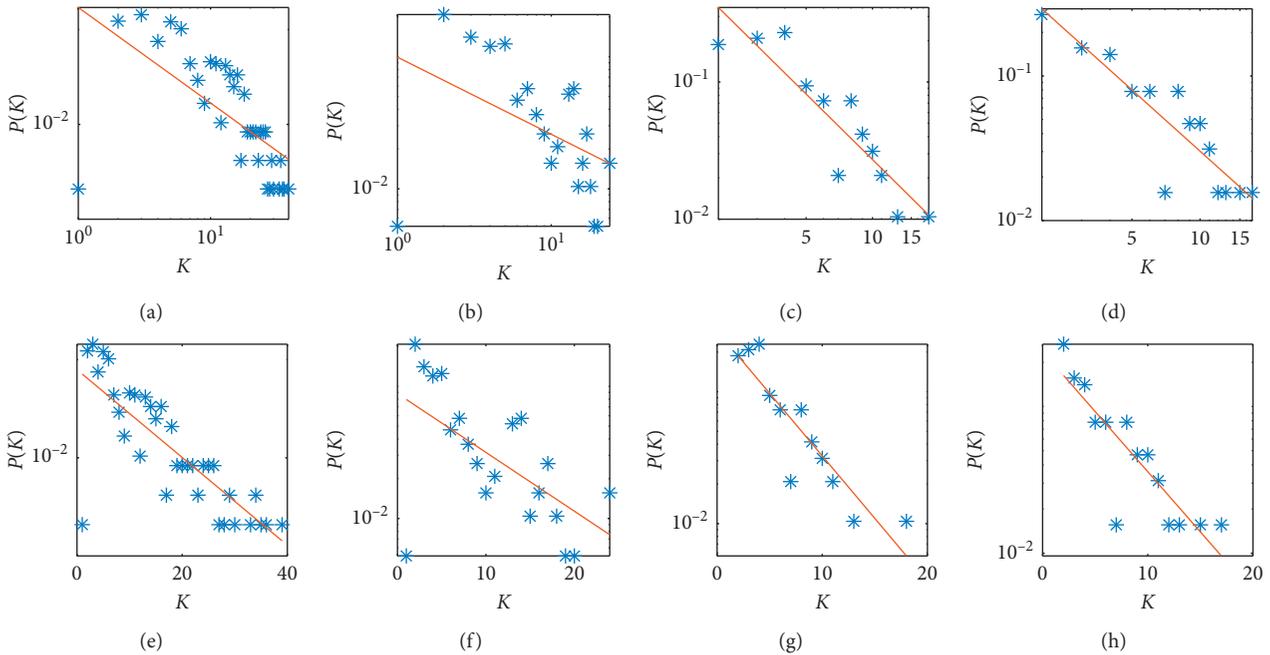


FIGURE 16: Results after power law distribution and exponential distribution fitting. (a)  $\Delta t = 2$  min. (b)  $\Delta t = 5$  min. (c)  $\Delta t = 10$  min. (d)  $\Delta t = 15$  min. (e)  $\Delta t = 2$  min. (f)  $\Delta t = 5$  min. (g)  $\Delta t = 10$  min. (h)  $\Delta t = 15$  min.

structure disappeared due to the weak antinoise capability of the HVG algorithm. As shown in Figure 10, the complex networks converted by the VG algorithm conform to the multipeak structure, proving that the four groups of air traffic flow data at different time scales all exhibit chaotic characteristics. A multipeak structure is not present in Figure 11, indicating that the noise-resistance capability of the HVG

algorithm is weaker than that of the VG algorithm. By analyzing the degree distribution diagram of EHHVG presented in Figure 15, it can be found that a similar multipeak structure is also present. The EHHVG algorithm therefore has better noise resistance than the classical HVG algorithm.

In the research by Lacasa Toral [30], it was proven that the HVG algorithm could accurately distinguish random

TABLE 3: Statistics of goodness of fit parameters.

Figure	Fitting method	$\lambda/\gamma$	SSE	$R^2$	RMSE
(a)	Power law distribution	-1.007224	38.9936	0.4031	1.0870
(b)	Power law distribution	-0.583436	19.6875	0.1894	1.0458
(c)	Power law distribution	-1.585173	2.4013	0.8220	0.4900
(d)	Power law distribution	-1.408122	2.4476	0.8064	0.4516
(e)	Exponential distribution	-0.102947	23.2955	0.6434	0.8402
(f)	Exponential distribution	-0.102463	15.4925	0.3621	0.9277
(g)	Exponential distribution	-0.218255	2.5323	0.8123	0.5032
(h)	Exponential distribution	-0.188666	3.0696	0.7571	0.5058

time series from chaotic time series. After fitting the degree distribution into an exponential distribution form  $p(k) \sim e^{-\gamma k}$ ,  $\gamma = \ln(3/2)$  is the boundary between random and chaotic time series; when  $\gamma > \ln(3/2)$ , the time series is chaotic. As presented in Figure 13, the results under the four time scales all conform to  $\gamma < \ln(3/2)$ ; thus, the time series have chaotic characteristics under the four time scales.

## 6. Conclusions

The innovations of the present work are as follows:

- (1) This paper simplifies the steps of pattern recognition in air traffic flow extraction and proposes a method to quickly extract the traffic flow passing through a certain waypoint.
- (2) This paper uses an advanced improved C-C method to calculate the reliability parameters. These methods are applied to the analysis of air traffic flow time series, thereby laying a solid foundation for the analysis of the chaotic characteristics of the system.
- (3) This paper proposes a new visibility graph network construction method, analyzes its degree distribution characteristics, and compares it with two existing classic visibility algorithms, thereby enriching the research on complex networks for the determination of the nonlinear characteristics of time series.
- (4) Three complex network analysis methods, namely, the VG algorithm, HVG algorithm, and proposed EHHVG algorithm, are compared in terms of the goodness-of-fit parameters of the degree distributions in different time scales. Regarding the goodness of distribution fitting of the three complex network construction methods, the time series under the time scales of 10 min and 15 min are better than those under the time scales of 2 min and 5 min.

There are many aspects of the characteristics of air traffic flow that are worthy of the in-depth study. Theoretical research on air traffic flow can better guide the management and control of local air traffic flow, which is of great importance for solving the problems of heavy traffic loads and frequent flight delays.

## Data Availability

The ADS-B data provide detailed data for individual flight, such as serial number, registration number, ICAO, flight call

sign, latitude, longitude, altitude, speed, heading, vertical speed, receiver IP address, and flight time. The research object of this paper is air traffic flow, so it is necessary to extract time series of traffic flow from the track data. The data will be made available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This study was supported in part by the National Natural Science Foundation of China (No. 71801215) and in part by the Research Funds for Interdisciplinary Subject, Northwestern Polytechnical University.

## References

- [1] J. E. Disbro and M. Frame, "Traffic flow theory and chaotic behavior," *Transportation Research Record*, vol. 1225, no. 1, pp. 109–115, 2018.
- [2] D. Low and P. Addison, "Chaos in a car-following model with a desired headway time," *International Symposium on Automotive Technology & Automation Att/its Advances for Enhancing Passenger Freight & Intermodal Transportation Systems*, vol. 1225, no. 1, pp. 175–182, 1997.
- [3] Y. S. Tang et al., "Chaos forecast for traffic conflict flow," *Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)*, vol. 35, no. 6, pp. 646–648, 2005.
- [4] Y. Liu and J. Zhang, "Predicting traffic flow in local area networks by the largest Lyapunov exponent," *Entropy*, vol. 18, no. 1, p. 32, 2016.
- [5] G. W. Frank, T. Lookman, M. A. H. Nerenberg, C. Essex, J. Lemieux, and W. Blume, "Chaotic time series analyses of epileptic seizures," *Physica D: Nonlinear Phenomena*, vol. 46, no. 3, pp. 427–438, 1990.
- [6] X. Zheng, "Research on nonlinear characteristics of air traffic flow," Doctoral dissertation, Civil Aviation University of China, Tianjin, China, 2016.
- [7] F. Wang, "Empirical analysis on air traffic flow long phase correlation based on Hurst exponent," *Journal of Civil Aviation University of China*, vol. 37, no. 2, pp. 1–4, 2019.
- [8] S.M. Li, X.H. Xu, and L.H. Meng, "Flight conflict forecasting based on chaotic time series," *Transaction of Nanjing University of Aeronautics & Astronautics*, vol. 29, no. 4, pp. 388–394, 2012.

- [9] W. Cong and M. Hu, "Chaotic characteristics analysis of air traffic system," *Transaction of Nanjing University of Aeronautics & Astronautics*, vol. 31, no. 6, pp. 636–642, 2014.
- [10] Y. Yang, "Research on short term forecasting method of air traffic flow," Doctoral dissertation, Civil Aviation University of China, Tianjin, China, 2017.
- [11] Y. Yang and H. Yang, "Complex network-based time series analysis," *Physica A: Statistical Mechanics and Its Applications*, vol. 387, no. 5-6, pp. 1381–1386, 2008.
- [12] Z. Gao and N. Jin, "Complex network from time series based on phase space reconstruction," *Chaos An Interdisciplinary Journal of Nonlinear Ence*, vol. 19, no. 3, p. 41, 2009.
- [13] R. V. Donner, Y. Zou, and J. F. Donges, "Ambiguities in recurrence-based complex network representations of time series," *Physical Review E*, vol. 81, pp. 1–4, Article ID 015101, 2010.
- [14] J. Pan, H. Jiang, J. Gao, and P. Yang, "Condition diagnosis with complex network-time series analysis," in *Proceedings of the Annual Reliability & Maintainability Symposium*, January 2011.
- [15] J. Tang, F. Liu, W. Zhang, S. Zhang, and Y. Wang, "Exploring dynamic property of traffic flow time series in multi-states based on complex networks: phase space reconstruction versus visibility graph," *Physica A: Statistical Mechanics and Its Applications*, vol. 450, pp. 635–648, 2016.
- [16] C. Wang and M. Zhu, "Characterizing dynamic property of air traffic flow time series based on complex network," *Computer Engineering and Applications*, vol. 35, no. 6, pp. 81–85, 2018.
- [17] K. Holger and S. Thomas, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, Second edition, 2003.
- [18] F. Takens, "Determining strange attractors in turbulence," *Lecture Notes in Math*, vol. 898, pp. 361–381, 1981.
- [19] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Vol. 127, Cambridge University Press, Cambridge, UK, 1997.
- [20] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.
- [21] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena*, vol. 9, no. 1-2, pp. 189–208, 1983.
- [22] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, no. 3403, pp. 34–45, 1992.
- [23] J. Ma, Y. Zhang, and Z.-H. Cao, "Chaos characteristics of space object RCS based on Cao method," *Modern Radar*, vol. 31, no. 5, pp. 61–64, 2009.
- [24] D. Kugiurmtzis, "State space reconstruction parameters in the analysis of chaotic times series-the role of the time window length," *Physica D*, vol. 95, pp. 13–28, 1996.
- [25] H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamics, delay times, and embedding windows," *Physica D: Nonlinear Phenomena*, vol. 127, no. 1-2, pp. 48–60, 1999.
- [26] N. Huang and L. Ma, "The prediction model of metro vault settlement based on developed C-C method phase space reconstruction and LS-SVM," *Mathematics in Practice & Theory*, vol. 44, no. 20, pp. 130–139, 2009.
- [27] L. I. Xinjie, H. U. Tiesong, and D. Xiuming, "Application of Recurrence Plots to Nonlinear Analysis of Runoff Time Series," *Engineering Journal of Wuhan University*, vol. 46, no. 1, pp. 62–66, 2013.
- [28] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, "From time series to complex networks: the visibility graph," *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pp. 4972–4975, 2008.
- [29] B. Luque et al., "Horizontal visibility graphs: exact results for random time series," *Physical Review E*, vol. 80, Article ID 046103, 2009.
- [30] L. Lacasa and R. Toral, "Description of stochastic and chaotic series using visibility graphs," *Physical Review E*, vol. 82, no. 3, pp. 1–26, 2010.
- [31] R. Zhang, "Nonlinear time series analysis by means of visibility graphs," Doctoral dissertation, Civil Aviation University of China, Tianjin, China, 2016.
- [32] T.T. Zhou, N.De Jin, Z.Ke Gao et al., "Limited penetrable visibility graph for establishing complex network from time series," *Acta Physica Sinica*, vol. 61, no. 3, Article ID 30506, 2012.