Transportation Research Part C 124 (2021) 102970

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Lateral collision risk assessment of parallel routes in ocean area based on space-based ADS-B

Fei Lu [a,*], Zichen Chen [a], Huiyu Chen [b]

[a] College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China
[b] Research Institute of Civil Aviation Safety, Civil Aviation University of China, Tianjin 300300, China

ARTICLE INFO

Keywords:
Collision risk
Space-based ADS-B
Segmented Kalman filter
Track fitting
Machine learning
Bayesian network

ABSTRACT

In order to ensure the safe operation of parallel routes in the ocean area, it is necessary to assess the risk of lateral collision of routes. The data used in the assessment often come from the real flights, while the traditional land-based ADS-B equipment has limited coverage and cannot be used for research. The space-based ADS-B system can receive and send the flight position of the aircraft over the ocean area. In this paper, the ocean tracks in space-based ADS-B data are first extracted. According to the different route structures, the data of fixed route and flexible route are preprocessed by segmented Kalman filter and cubic polynomial fitting method respectively. Then the processed track position error is calculated, and Q-Q (Quantile-Quantile) median method is used to test the normal distribution of the error, and k-means clustering is used to obtain the clustering center and occurrence frequency of positioning error and flight error. Finally, a collision risk model based on Bayesian network is established, and the error distribution parameters and occurrence frequency are brought into the model to evaluate the collision risk of parallel routes in the ocean area. The results show that the lateral collision risk based on space-based ADS-B ocean flight data is lower than 0.5×10^{-8} times/hour specified by ICAO when the fixed route separation is 14n miles and the flexible route separation is 49n miles. And the actual lateral interval of the Pacific flexible route is 50n miles (Jiang, 2010), which meets the safety target level.

1. Introduction

In order to ensure the flight safety of aircraft, the International Civil Aviation Organization (ICAO) has made clear regulations on the separation of aircraft in all directions on ocean parallel routes. The separation standards of most countries in the world are specified in accordance with documents issued by the ICAO. However, with the development of the aviation industry, the airspace environment is becoming more and more crowded. At the same time, the real-time performance and accuracy of communication navigation surveillant (CNS) equipment have been greatly improved, which lays a foundation for further reducing the minimum separation standard of ocean parallel routes.

The determination process of whether the minimum separation can be reduced is as follows: 1) Calculate the collision risk of the aircraft under this separation; 2) Compare the collision risk with the target level of safety (TLS), if it is far less than, it means that the separation standard can be reduced theoretically. As the first step in the determination process, the calculation of collision risk is essential and especially critical.

* Corresponding author.
E-mail address: flu@cauc.edu.cn (F. Lu).

https://doi.org/10.1016/j.trc.2021.102970
Received 2 August 2020; Received in revised form 7 November 2020; Accepted 5 January 2021
Available online 18 January 2021

The first scholar who has made constructive achievements in the research of aircraft collision risk is Reich, P.G. of Britain (Reich, 1966a,b,c). The classical Reich model proposed by him regards an aircraft as a cuboid collision slab, with a layer of proximity shell nested outside the collision slab, and the other aircraft as a particle. When the particle enters the proximity shell, it is regarded as a dangerous approach, and when the particle contacts the collision slab, it is regarded as a collision, thus simplifying the calculation of collision risks on parallel routes.

Reich model is widely used. Brooker reported on the application of the Reich model in the design of the North Atlantic route (Brooker, 1983). The study also proposed that not only the capabilities of navigation equipment, but also the way of use is also important for navigation performance. The position error data used in the study to assess risk comes from data within the radar coverage on the coastline, not the entire route. After 1990 s, some scholars improved the model by deleting the restrictive assumptions in the Reich model to make it suitable for actual air traffic conditions (Blom et al., 1998; 2003; Shortle et al., 2004; Davies and Sharpe, 1993; Bakker and Blom, 1993; Bakker et al., 2000). Wang and Xu proposed probability model for the collision risk of intersecting air routes, provided research methods for the formulation of lateral separation (Wang and Xu, 2001). Cox (Cox et al., 1991), Harrison (Harrison and Moek, 1992) and Moek (Moek et al., 1993) collected the radar data of continental type traffic from Hannover and Bretigny, and data of North Atlantic traffic from Aberporth. They calculate the risk of collision by counting the vertical flight errors on the route, and provide support for reducing the vertical separation in the airspace above FL290. The research results show that reducing the vertical separation is theoretically feasible, but the operational and technical performance cannot meet the requirement. Actually, the data of North Atlantic traffic from Aberporth used in the research is also limited by the radar range, and the risk assessment results cannot be applied to the entire transoceanic route. Brooker analyzed the existing safety assessment model and collision risk model, and analyzed the safety separation from an accident perspective (Brooker, 2002a; 2002b). A year later, he proposed 'Post-Reich' (i.e., Event model), which provided a more direct calculation method than the Reich model (Brooker, 2003). When calculating the collision risk, the main parameters of the model are clear and the process is simple. Brooker calculated the lateral collision risk and the longitudinal collision risk separately by using the Event model (Brooker, 2003; 2006). Zhang used VOR navigation data to evaluate the lateral collision risk of parallel routes in non-ocean areas (Zhang et al., 2007). Netjasov proposed a conflict risk evaluation model for airspace strategic planning with conflict probability and number of conflicts giving the minimum flight safety separation (Netjasov, 2012a; 2012b; Netjasov and Babić, 2013)). J.A. Pérez-Castán et.al. made improvements on the basis of the Netjasov collision risk model, and carried out airspace planning based on the risk assessment results (Pérez-Castán et al., 2019).

Existing flight separation studies usually focus on continental airspace (because the data used in the safety assessment is continental radar data), while the data (radar data near the coastline) used in the separation studies in oceanic airspace is also not available enough, because it is not the complete data of the flight in the ocean area. To verify whether the lateral flight interval implemented on transoceanic routes meets the safety target level, or whether the lateral separation of ocean areas can be reduced on the premise of meeting the safety target level, so as to improve the utilization rate of ocean airspace, we will calculate the lateral collisions risk for transoceanic flights, and obtain transoceanic lateral flight separation that meet the safety target level, by collecting space-based ADS-B data for *trans*-Pacific flights.

The reason for using novel space-based ADS-B data in this paper is that it can record the complete operation of each flight on the ocean route, rather than radar data near the coastline. Traditional radar surveillance and land-based ADS-B technology failed to cover the ocean and desert areas. In order to solve this problem, researchers built ADS-B receivers on satellites, thus realizing global coverage of air traffic control. The working principle of space-based ADS-B is shown in Fig. 1. Satellites provide navigation for aircraft in ocean areas. The aircraft transmits ADS-B data to the satellite, and the satellite transmits the data back to the ground station, which is finally received by the control unit.

In this paper, the collected satellite-based ADS-B data will be divided into two categories, one is 'Fixed route', the other is 'Flexible route', and then the two types of data are preprocessed and position error statistics are performed, and the distribution and generation disciplines of errors are analyzed by using parameter estimation and machine learning methods. And lateral flight separations in two types ocean route are carried out after assessment of collision risk by establish a Bayesian network collision risk model. The research process of this paper is shown in Fig. 2.

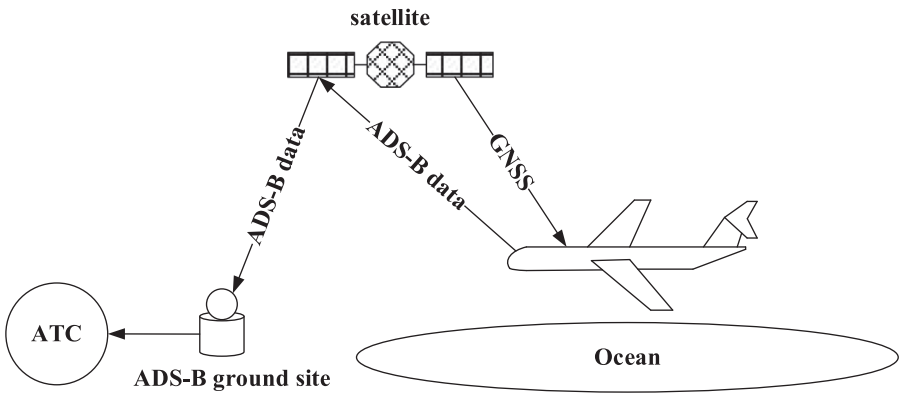

Fig. 1. Space-based ADS-B working principle.

2. Processing of track data

Since this study is aimed at the collision risk of parallel routes over the ocean area, it is necessary to screen the track position data, and select the research data based on the screening conditions of sufficient *trans*-ocean routes and position reporting points. In addition, satellite positioning often has errors, so the position data need to be processed.

2.1. Ocean area data selecting

The space-based ADS-B data used in the research is 24-hour operation data on a certain day, with a total of 6,498 flights. Some flight paths are shown in Fig. 3. Color bar represents height (ft).

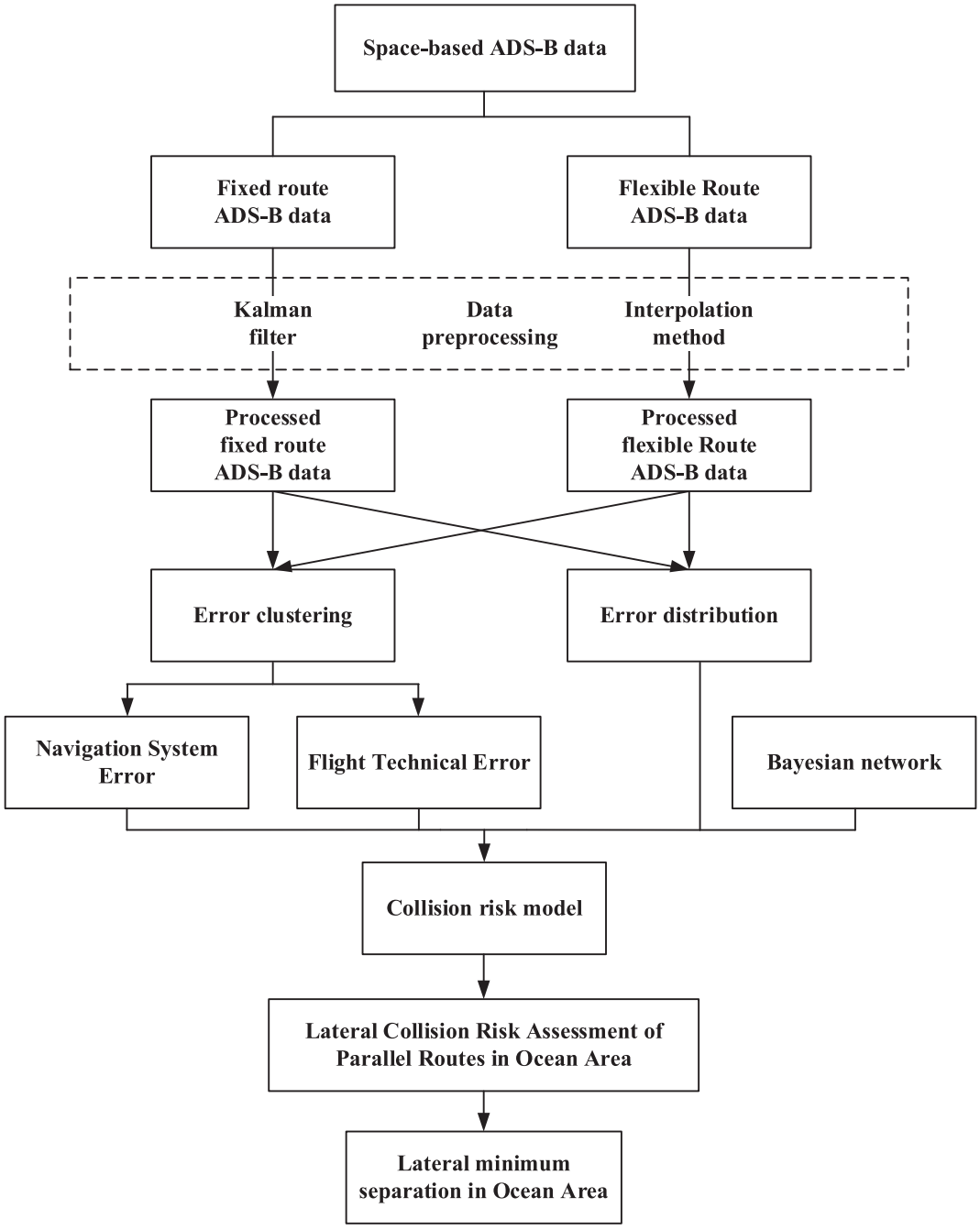

Fig. 2. Research process.

As can be seen from Fig. 1, most *trans*-ocean routes occur between E116°and W71° in our database. Based on the geographical location coordinate range and combined with flight identification number information, 1357 flights have been selected.

Considering the integrity and continuity of the data, the number of location reporting points for the same flight should be sufficient to meet the research needs. Suppose the sufficient number of position reporting points for flight A isλ_A,

$$\lambda_A = \frac{\sqrt{X^2 + Y^2}}{R} \tag{1}$$

wherein X is the difference between the last position reporting point coordinate longitude x_n and the first position reporting point coordinate longitude x_1 of the flight A, and

$$X = \begin{cases} x_n - x_1, & \text{Flight did not cross the Prime Meridian} \\ 360 - (x_n - x_1), & \text{Flight over the Prime Meridian} \end{cases} \tag{2}$$

Y is the difference between the last position report point coordinate latitude y_n and the first position report point coordinate latitude y_1 for flight A, $Y = |y_n - y_1|$, and R is the unit distance of the report point.

The time separation of ADS-B position data sent back by satellite is about 1 min in the cases of the dataset, so R is the distance that the flight flies in one minute. It is taken from a flight B with complete one-hour continuous data. The calculation method is as follows:

$$R = \frac{1}{60} \sum_{i=1}^{60} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \tag{3}$$

x_{i+1} and y_{i+1} are the longitude and latitude of the $i+1$ position reporting point of flight B; x_i and y_i are the longitude and latitude of the i position reporting point of the standard flight. Through research and calculation, the final value of R is 0.17.

Considering the long satellite transmission distance, which is easy to lead to data packet loss (Wang, 2018), when the number of data at the real location reporting point of flight A is not less than 1/2 of λ_A, the data of flight A can be considered to be available.

2.2. Data preprocessing

Space-based ADS-B receiver can effectively track flight, receive and return a large amount of position data. However, due to the problems of long transmission distance, high dynamics and dense signals in high density areas, the returned position data often have leakage points and anomalies, as shown in Fig. 4. The three lines are position data track of airline CCA165, CSN321, CSN381 respectively. The leakage points and anomalies mentioned here refer to the loss of position information and positioning errors caused by the navigation system. Therefore, the space-based ADS-B data does not represent intended routes flown and it is essential to preprocess the position data returned by space-based ADS-B.

We can see from Fig. 1 that in the received space-based ADS-B data, *trans*-ocean flights can be divided into two types. One is the 'Asia-Oceania' flight across the South Pacific. Many islands are distributed on the ocean area over which such routes fly. Aircraft can fly according to the guidance of navigation consoles on the islands, and the routes used on the daily routes are basically fixed. The other is the 'Asia-North America' route across the North Pacific Ocean. There are no navigation consoles on the ocean areas that these routes fly over. The route used is called the " Pacificorganized track system (PACOTS)".

The PACOTS is different from other fixed routes in that it can be considered as a "flexible route", which is set up to optimize route flow, utilize relevant airspace and save flight time. Taking into account the influence of volcanic weather and other factors, the issuing

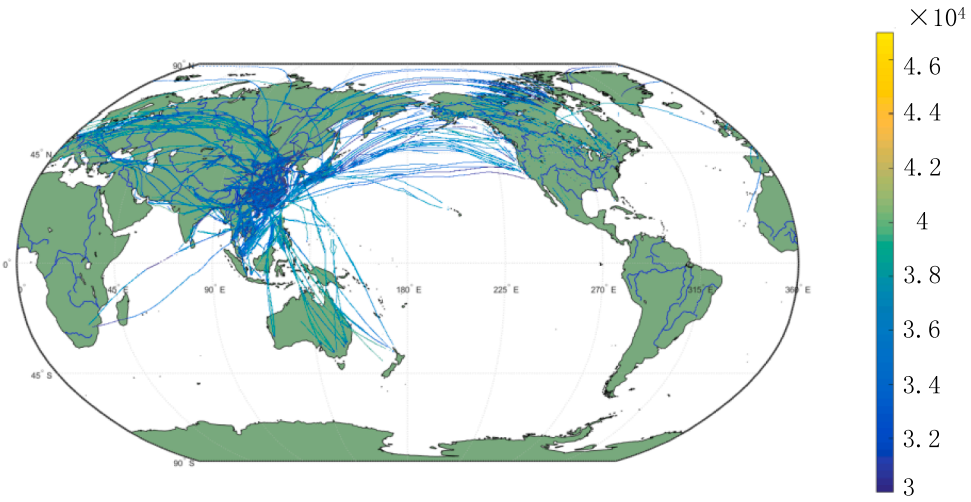

Fig. 3. Partial Space-based ADS-B Data.

Fig. 4. Flight tracks in the oceans.

agency issued the route that is most suitable for flying on the same day or at a specified time period. The information regions across the North PACOTS include Japan's Fukuoka Information Region, Anchorage Information Region and Auckland Information Region. Daily information on east–west marshalling routes in the Pacific Ocean area is released by Fukuoka Information Region RJJJ and Auckland Information Region KZAK for airlines to choose.

When the aircraft flies on a fixed route, the navigation console is used as a reference, but the flexible route does not. Therefore, different calculation methods are adopted to process the flight data on the two routes. The preprocessed data will be used in subsequent error analysis

2.2.1. Fixed route

Fixed route is composed of multiple route points connected. Aiming at the problem of abnormal data of location reporting points, segmented Kalman filter is adopted to supplement the missing location point information, reduce fluctuation and smooth track at the same time.

Kalman filter is widely used in the processing of various navigation data, like GPS (Shen et al., 2011; Choi et al., 2010; Li et al.,

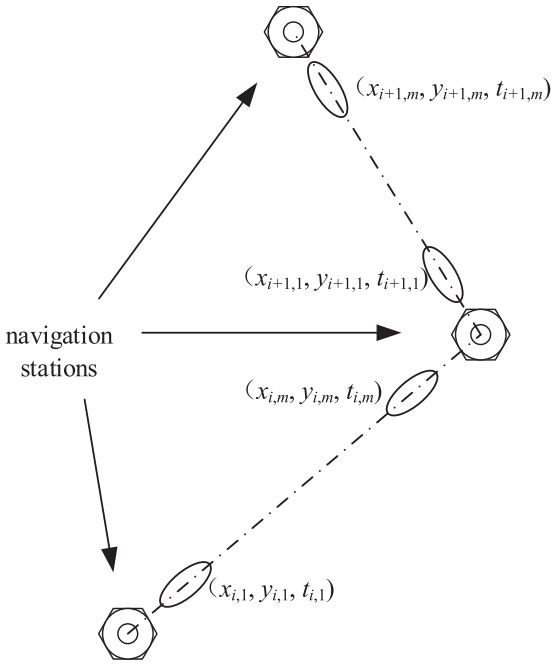

Fig. 5. The motion state of the aircraft.

2013; Chen et al., 2013) and AVL (Cathey and Dailey, 2003). However, for *trans*-ocean routes, aircraft must pass through navigation consoles on the route in turn, and the heading and speed change regularly during flight. Therefore, a segmented Kalman filter is proposed to process space-based ADS-B data.

Kalman filter can optimally estimate the state of the system. It uses linear state equation to describe the state of the system and uses recursive method to reduce the operation storage. The core of Kalman filter is divided into two equations, state equation and observation equation. The motion state of the aircraft flying in the ocean area shows in Fig. 5, and it can be described as:

$P_{i,(k+1)} = P_{i,(k)} + TP_i^{'}$, $i = 1, \cdots, n$, $k = 1, \cdots, m-1$where T is the receiving time separation of the space-based ADS-B receiver

$$P_{i,(k)} = [x_{i,k}, y_{i,k}]^{\mathrm{T}} \tag{4}$$

$$P_i^{'} = \left[\frac{\Delta x_i}{\Delta t_i}, \frac{\Delta y_i}{\Delta t_i}\right]^{\mathrm{T}} \tag{5}$$

And

$$\Delta x_i = x_{i,m} - x_{i,1},\ \Delta y_i = y_{i,m} - y_{i,1},\ \Delta t_i = t_{i,m} - t_{i,1} \tag{6}$$

$x_{i,k}$ is the longitude coordinate of the k-th coordinate point on the i-section route, $y_{i,k}$ is the latitude coordinate of the k-th coordinate point on the i-section route, and $t_{i,1}$ is the first moment when the aircraft enters the i-section route and $t_{i,m}$ is the last moment when the aircraft enters the i-section route.

The equation of state of the aircraft is:

$$X_{i,(k+1)} = AX_{i,(k)} + BU_{i,(k)} + W_{i,(k)},\ k = 1, \cdots, m \tag{7}$$

where A is the state transition matrix of the aircraft; $X_{i,(k)}$ represents the target state vector at the k-th position coordinate point of the i-th route; B is the relationship matrix between input and state; $U_{i,(k)}$ is a known input signal; $W_{i,(k)}$ represents a white Gaussian process noise sequence with zero mean and no correlation, covariance is Q, and Q is a symmetric matrix.

The observation equation of the aircraft is:

$$Z_{i,(k+1)} = HX_{i,(k)} + V_{i,(k)},\ k = 1, \cdots, m \tag{8}$$

Among them, $Z_{i,(k+1)}$ represents the observation vector of the aircraft at the k-th position of the i-th route; H is an observation matrix; $V_{i,(k)}$ represents a white Gaussian observation noise sequence with zero mean and no correlation, covariance is R, and R is a symmetric matrix.

When the state vector $X_{i,(k)} = \left[x_{i,k}, \frac{\Delta x_i}{\Delta t_i}, y_{i,k}, \frac{\Delta y_i}{\Delta t_i}\right]^{\mathrm{T}}$ is used, the state transition matrix A of the aircraft is:

$$A = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

When observation vector $Z_{i,(k)} = \left[x_{i,k}, y_{i,k}\right]^{\mathrm{T}}$, the aircraft's observation matrix H is:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The recursive process of Kalman filter is divided into the following five steps:

(1) Calculate the predictive value of the state step

$$X_{i,(k+1|k)} = AX_{i,(k|k)} + BU_{i,(k|k)} \tag{9}$$

(2) Calculate the mean square error matrix of step prediction

$$P_{i,(k+1|k)} = AP_{i,(k|k)}A^{\mathrm{T}} + Q \tag{10}$$

(3) Calculate the filter gain matrix

$$K_{i,(k+1)} = P_{i,(k+1|k)}H^{\mathrm{T}}/[HP_{i,(k+1|k)}H^{\mathrm{T}} + R] \tag{11}$$

(4) Calculate the state filter value

$$X_{i,(k+1|k+1)} = X_{i,(k+1|k)} + K_{i,(k+1)}[Z_{i,(k+1)} - HX_{i,(k+1|k)}] \tag{12}$$

(5) Calculate the filter mean square error matrix

$$P_{i,(k+1|k+1)} = [I - K_{i,(k+1)}H]P_{i,(k+1|k)} \tag{13}$$

Cycle through the above 5 steps until all position states are calculated.

Taking a section of route between two navigation points passed by a certain flight as an example, the position data of the section of route are preprocessed, and the initial filtering mean square error matrix P_0 is set as follows:

$$P_0 = 0.0015 \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The covariance matrices of process noise and observation noise are as follows:

$$Q = 10^{-6} \times \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} 0.0005 & 0 \\ 0 & 0.0005 \end{bmatrix}$$

The processing results are shown in Fig. 6 below:

The original track data errors and the Kalman filter track data errors are shown in Table 1 below

As can be seen from Fig. 3 and Table 1, Kalman filter can smooth the track information well, supplement and perfect the missing position information, and restore the flight state.

2.2.2. *Flexible route*

Since there is no fixed route and navigation console, using Kalman filter to process track information on flexible routes will produce huge computation. Therefore, interpolation method, the most commonly used and simplest method in data processing, is used. For flight data flying on flexible routes, linear interpolation, nearest interpolation, spherical interpolation, and cubic polynomial interpolation methods are respectively used to process the data.

Taking a flight data as an example, the processing results of the four methods are shown in Fig. 7 below.

In order to ensure that the interpolation track has a high enough correlation with the real track, it is necessary to check the cor-

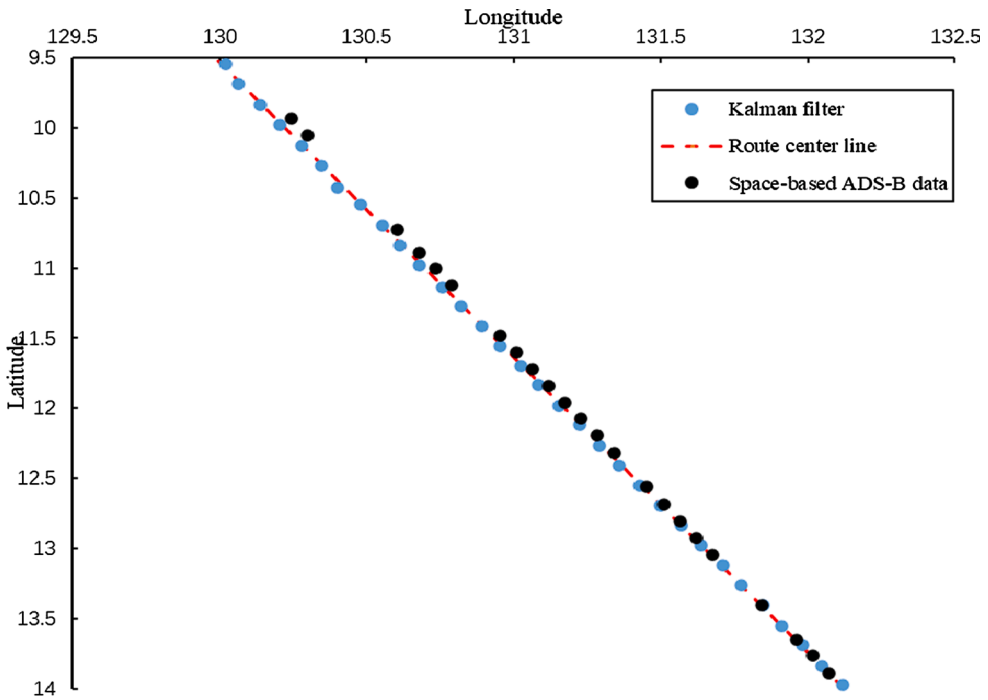

Fig. 6. Segmented Kalman filter Results.

Table 1
Comparison of Track Data before and after Processing.

	Max error/n mile	Mean error/n mile
Original track data	2.747161528	0.931081689
Kalman filter track data	1.057500633	0.259131423

relation of the track data preprocessed by the four methods. Pearson's linear correlation coefficient can accurately reflect the goodness of correlation of the method, and the equation is as follows

$$\gamma_{XY} = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2 \sum(Y_i - \overline{Y})^2}} \tag{14}$$

The correlation coefficient γ_{XY} is expressed in the equation. The closer to 1, the greater the correlation. The correlation of the four interpolation methods is calculated by Eq. (14), and the calculation results are shown in Table 2:

The results of the interpolation test show that all four methods can effectively fit the flight track, but we can see from the Fig. 4 that the results of proximity interpolation obviously do not conform to the normal flight track. The interpolation method with the highest correlation——cubic polynomial interpolation, is selected as the preprocessing method of North Atlantic flight data.

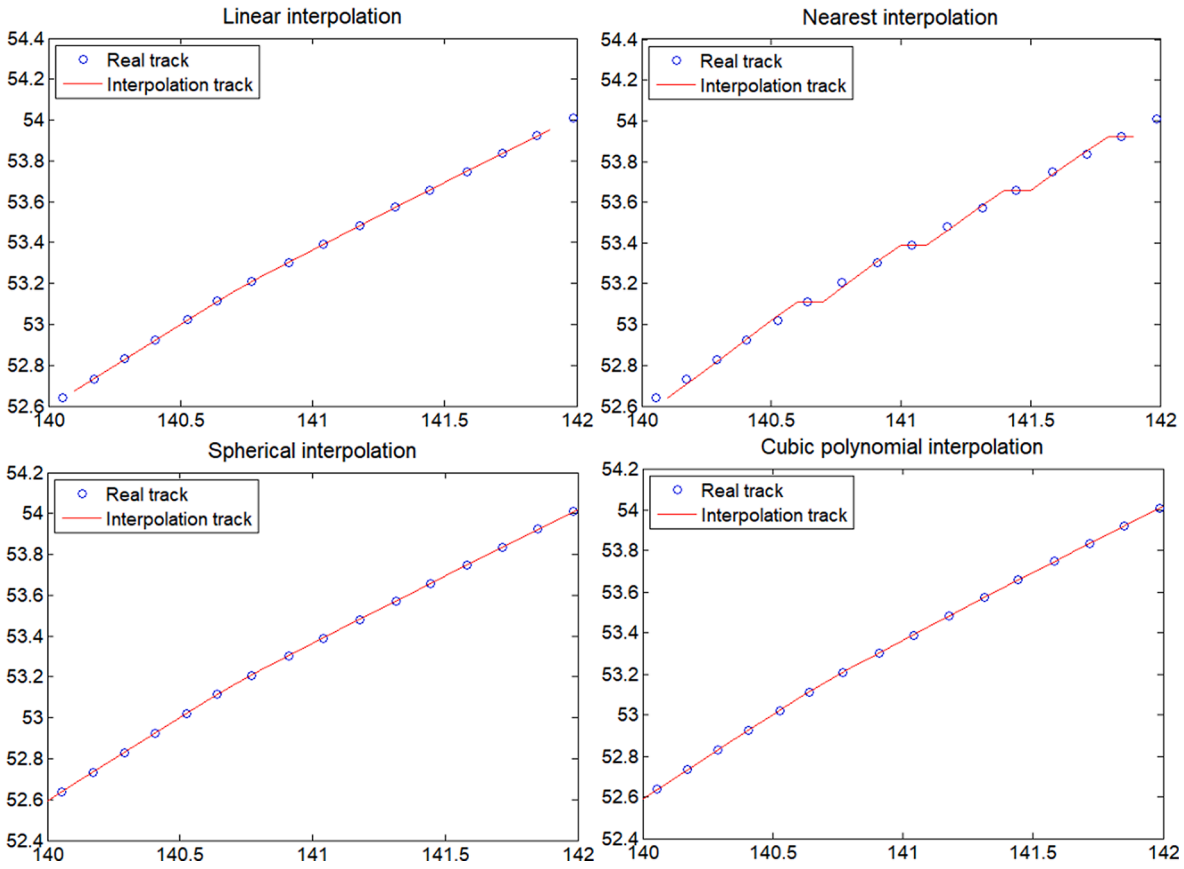

Fig. 7. Interpolation Results.

Table 2
Fitting Degree Results.

interpolation methods	linear interpolation	nearest interpolation	spherical interpolation	cubic polynomial interpolation
correlation	0.999320901	0.999298346	0.999321081	0.99932109

3. Track error analysis

3.1. Error distribution

3.1.1. Sample calculation

Because there are essential differences in structure between fixed routes and flexible routes, the track errors on the two types of routes are calculated separately. When calculating the error, the most important thing is to find the route center line of the flight. There are navigation consoles on the fixed route to guide the course of the aircraft to ensure that the aircraft can fly according to the predetermined route. The connection between adjacent navigation consoles is the route center line of the flight. However, when the plane flies on the flexible route, there is no navigation console as a reference, and the available routes released every day are also different. Therefore, Eq. (15) is used to calculate the centerline of the flexible route.

$$\overline{Y_i} = \frac{\sum_{j=1}^{n} Y_i^{(j)}}{n},\ i = 1, \ldots, m \tag{15}$$

where n is the number of flights flying on the same flexible route in a day, m is the number of route interpolation points, $Y_i^{(j)}$ is the latitude coordinate of the j-th flight at the i-th interpolation point, and $\overline{Y_i}$ is the i-th calculation result of the route center point.

Taking the average error value of a flight as a sample, the sample space consists of 35 samples. It is set that the track position point falls on the left side of the route center line as negative and the right side as positive. The error calculation results of fixed route and flexible route are shown in Tables 3 and 4.

3.1.2. Normality test

The Central Limit Theorem states, that the distribution of means of independent and identically distributed samples will have a normal distribution. We assumed that the position error of each flight follows the same position error distribution and that each flight in Tables 3 and 4 represents a sample randomly drawn from the position error population.

The Q-Q chart in the normal probability paper test method is used to test the sample distribution of error data. The Q-Q chart uses the relation curve between the quantile of sample data distribution and the quantile of the designated distribution to test the distribution of data. As shown in Fig. 8, Tables 3 and 4 can be considered as samples from normal populations$N(\mu,\sigma^2)$, μ and σ^2 is population mean and population variance, respectively.

If the sample mean $\overline{X}$ is an unbiased estimator of the population mean μ, then given the significance level $\alpha(\ 0 < \alpha < 1)$, the confidence interval $1-\alpha$ of the confidence level of μ and σ^2 are calculated as Eqs. (16) and (17) respectively:

$$\left[\overline{X} - t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}, \overline{X} + t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right] \tag{16}$$

Table 3
Samples of Fixed Route.

Sample number	Sample value/m	Sample number	Sample value/m
Flight1	5986.26	Flight17	339.32
Flight2	−912.24	Flight18	236.26
Flight3	177.24	Flight19	−3407.11
Flight4	−4277.61	Flight20	−495.06
Flight5	−2626.94	Flight21	−1738.96
Flight6	1532.90	Flight22	3424.22
Flight7	1103.24	Flight23	2959.67
Flight8	101.47	Flight24	972.86
Flight9	15.35	Flight25	795.90
Flight10	−131.57	Flight26	3548.05
Flight11	−1112.76	Flight27	−99.92
Flight12	−3937.89	Flight28	2383.00
Flight13	−625.95	Flight29	−1594.24
Flight14	−7458.39	Flight30	−762.75
Flight15	−3112.37	Flight31	−47.11
Flight16	4474.32	Flight32	−5807.89

Table 4
Flexible Route Samples.

Sample number	Sample value/m	Sample number	Sample value/m
Flight1	−11622.94	Flight19	5533.14
Flight2	−4520.48	Flight20	−851.75
Flight3	3025.91	Flight21	−3256.81
Flight4	−2906.29	Flight22	1090.89
Flight5	553.31	Flight23	8025.91
Flight6	−34326.59	Flight24	405.60
Flight7	−558.72	Flight25	13450.16
Flight8	993.18	Flight26	11749.44
Flight9	−13481.91	Flight27	−1918.83
Flight10	−8.68	Flight28	887.26
Flight11	−3708.03	Flight29	6993.18
Flight12	21890.89	Flight30	14810.52
Flight13	26253.63	Flight31	−21869.88
Flight14	−18677.60	Flight32	−6811.66
Flight15	1767.89	Flight33	4887.26
Flight16	−9351.75	Flight34	4450.16
Flight17	−6723.80	Flight35	−19437.18
Flight18	−5906.29	–	–

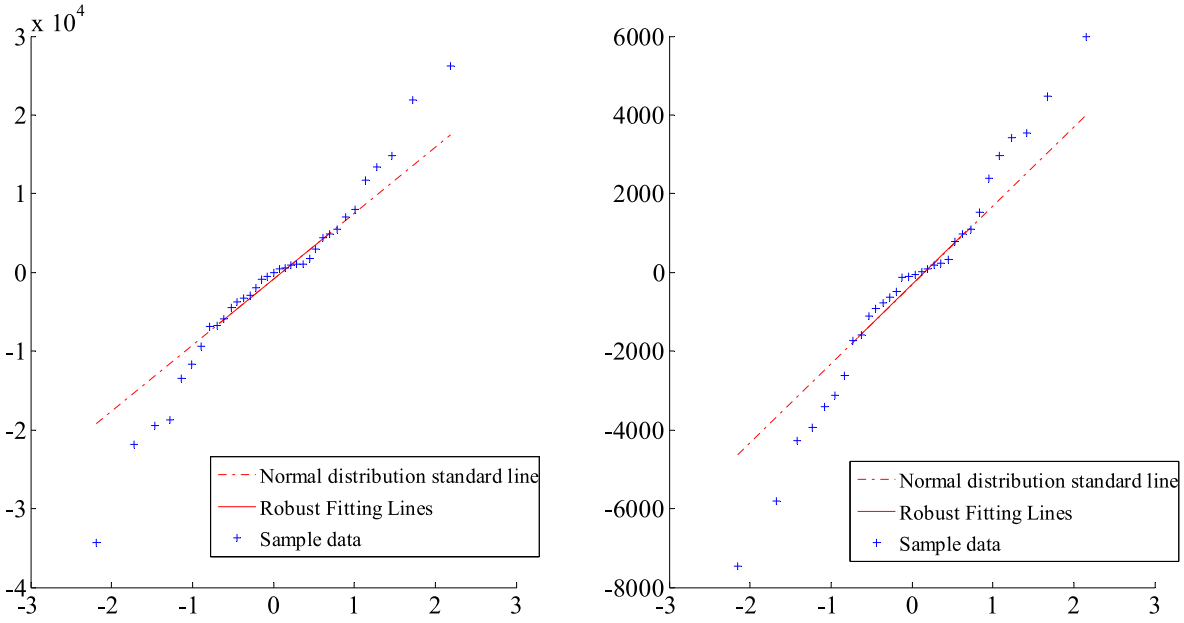

Fig. 8. Q-Q chart.

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right] \tag{17}$$

When the confidence level is 0.95, the confidence interval of the sample population mean μ and population variance σ^2 in Table 3 are (-1297.53, 666.3556) and (3318018.85, 25092890.56), respectively. The position error distribution of the aircraft in the ocean area can be considered to obey the normal distribution with the mean value of −315.585 and the standard deviation of 2834.079. In Table 4, the confidence interval of the population mean value μ and the population variance σ^2 of the samples are (-5014.34, 2776.00) and (59175744.81, 407359255.77), respectively. The distribution of the position error of the aircraft flying in the ocean area can be considered to follow the normal distribution with the mean value of −1119.17 and the standard deviation of 11757.42.

3.2. Error clustering

When flying in the ocean area, the aircraft is not monitored by the controller, and completes the flight mission completely according to the predetermined route and guidance of navigation equipment. Therefore, there are two sources of errors caused by ocean

flight, one is the Navigation System Error (NSE), and the other is the Flight Technical Error (FTE). NSE is caused by navigation accuracy, and the error caused by it is very small, while the FTE often makes the aircraft deviate from the course, resulting in large error.

It is difficult to distinguish these two kinds of errors artificially in real flight data, so we use machine learning methods, k-means clustering algorithm, to cluster and analyze the position errors calculated based on space-based ADS-B data. The algorithm can divide the database into two clusters, the larger value is defined as FTE.

In 1967, MacQueen proposed k-means clustering method, and gave the detailed algorithm steps. K-means clustering method is still widely used more than 50 years later due to its simple algorithm, easy implementation and high efficiency.

For aircraft position error data, set the data set be , k-means clustering method divides the data in the set into K classes, that is, the number of clustered subsets generated is K, the set of subsets is , and each subset has a class center . The similarity of the data in the subset is judged by calculating the Euclidean distance between the center and other data in the subset by Eq. (18)

$$C_j = \sum_{x_i \in a_j} \left\| x_i - \lambda_j \right\|^2 \tag{18}$$

The objective function C of clustering is

$$C = min \quad \sum_{j=1}^{K} C_j \tag{19}$$

The flow of K-means clustering algorithm is shown in Fig. 9 below:

The absolute value of the error results calculated in fixed route and flexible route is taken and clustered. Because the initial clustering center of k-means clustering method is randomly selected, the clustering center results will deviate every time. Therefore, we repeated the clustering experiment for 30 times and took the average value as the final clustering center result. The error clustering results of fixed route and flexible route are shown in Table 5.

It can be seen from the table that the FTE generated by aircraft in flexible routes are relatively large, while the FTE in fixed routes occur more frequently. It can also be said that the errors generated in fixed routes are more concentrated and smaller and the flexible routes are more extreme. Combined with the samples in 2.1, we found that the reason for this phenomenon is the existence of navigation consoles in fixed routes.

When flight errors occur, there is no navigation console to guide aircrafts flying flexible routes and they cannot adjust their course in time. Aircraft flying on a fixed route can adjust its course in time when flight errors occur. However, navigation accuracy increases with the distance of the navigation station, it often passes through the center line of the route and adjusts its course repeatedly.

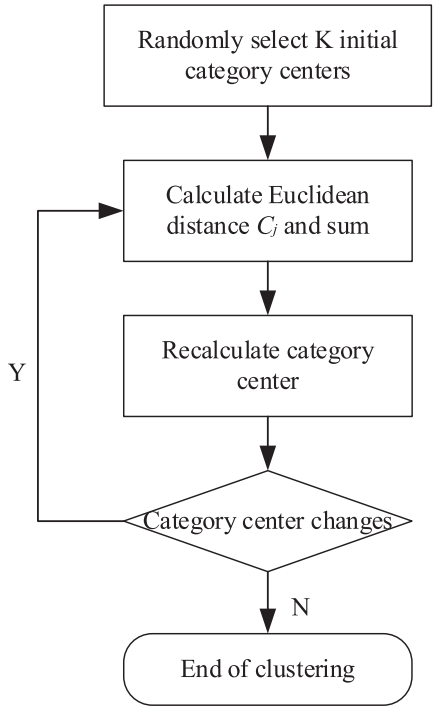

Fig. 9. K-means clustering process.

Table 5
Error clustering results.

	fixed route		flexible route	
	NSE	FTE	NSE	FTE
clustering center/m	120.52	19510.21	39.20	49380.63
frequency	0.45	0.55	0.83	0.17

4. Collision risk model

The collision of parallel routes in the ocean area is essentially caused by the position error of the aircraft. According to experience, the position error consists of NSE and FTE. Although space-based ADS-B data can reflect the size of the error and the probability of generating the error, and are applied to the calculation of collision risk, the internal causes of the error also need to be considered. Therefore, a collision risk model is established based on Bayesian network for parallel routes in the ocean area.

Bayesian network is a directed acyclic probability network model, which is mostly used in system risk assessment (Zhang and Liang, 2014; Jones et al., 2010) and reliability analysis (Yin et al., 2008; Amrin et al., 2018). It can clearly explain the causal relationship between events, and can carry out probabilistic bidirectional reasoning based on Bayesian theory.

In this study, a Bayesian network of collision risks of parallel routes in the ocean area is established from the perspective of the actual flight operation environment in the ocean area, as shown in Fig. 7.

Satellite failure and untimely detection during failure will cause the space-based ADS-B to malfunction. As a surveillance equipment, space-based ADS-B does not affect the risk of flight collision during flight operation. However, during risk assessment, data quality will affect the calculation results. The error caused by space-based ADS-B data, in calculation, is reflected in the flight error. Unfavorable conditions of pilot operations, TCAS and VOR navigation will also cause flight errors. In addition, because the deviation caused by the navigation error is small, so the simple navigation error will not directly lead to the occurrence of the lateral collision, and the flight error will (This is why the upper layer of the lateral overlap in the Bayesian network is the flight error). But we still need to consider the navigation error, so we incorporate the navigation error into the unfavorable state of the satellite, and its impact on the flight error will further affect the risk of collision.

In the Fig. 10, nodes represent random variables, and lines between nodes represent dependencies between variables, which are specifically described by the conditional probability distribution table of each node. Each node event is numbered, as shown in Table 6.

There are two states of event occurrence, and “event failure or occurrence”, i.e. unfavorable state, is recorded as 0; “The event is operating normally or does not occur”, i.e. the favorable state is recorded as 1. When the prior probability of each event is known, the posterior probability of the event can be obtained through probability reasoning.

The prior probabilities of adverse states of events A_1, A_2, A_3, A_4, A_5 and A_8 can be obtained by consulting reference (Zhang and

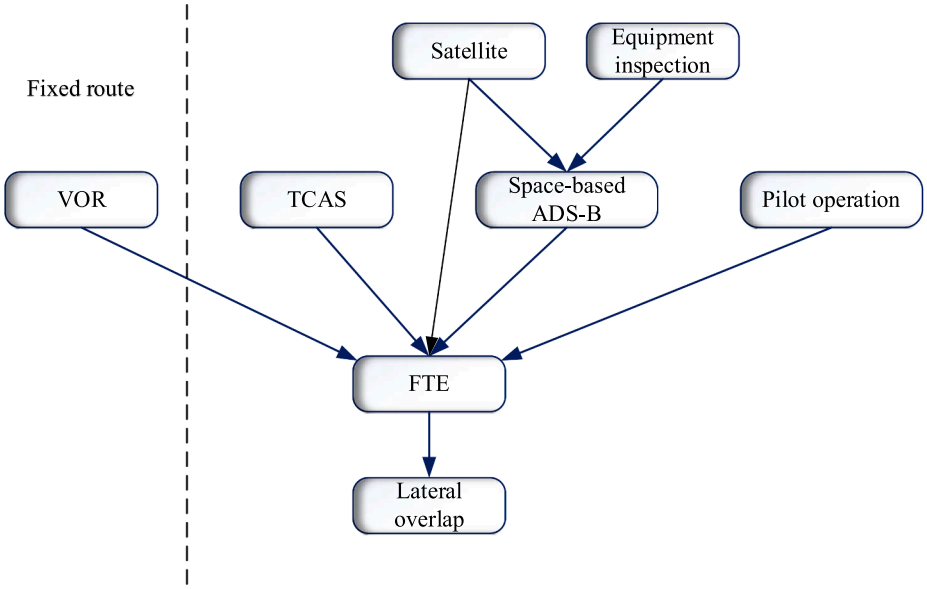

Fig. 10. Bayesian Network for Collision of Parallel Route in Ocean area.

Table 6
Event Number and Name.

Event number	Event	Event number	Event
A_1	TCAS	A_5	Space-based ADS-B
A_2	Satellite	A_6	FTE
A_3	Equipment inspection	A_7	Lateral overlap
A_4	Pilot operation	A_8	VOR

Liang, 2014; EUROCAE; Huang, 2012; Huang, 2013), which are 1.7×10^{-3}, 10^{-4}, 0.001, 0.003, 10^{-6} and 3×10^{-4} respectively. The prior probability of the unfavorable state of event A_6 can be obtained according to the result of error clustering in 2.2. The occurrence frequency is approximately equal to the probability, with 0.55 for fixed routes and 0.17 for flexible routes. The prior probability of the unfavorable state of event A_7 is calculated from the position error. If the lateral separation between the two aircrafts on the parallel route is L, then

$$L = D + x_1 - x_2 \tag{20}$$

$$L \sim N\left(D, 2\sigma^2\right) \tag{21}$$

where D is the distance between the two routes, $x_1 \tilde{} N(\mu_1, \sigma_1^2)$, and $x_2 \tilde{} N(\mu_2, \sigma_2^2)$ are the position error of the two planes, and $\mu_1 = \mu_2 = \mu, \sigma_1 = \sigma_2 = \sigma$. When event A_6 is in an unfavorable state, the lateral overlap probability of the two machines can be calculated by Eq. (22).

$$P\{L \leqslant d | A_6 = 0\} = \int_{-d}^{d} f(x)dx \tag{22}$$

where d is the size of the wingspan of the aircraft, and

$$f(x) = \frac{1}{2\sigma\sqrt{\pi}} e^{-\frac{(x-D)^2}{4\sigma^2}} \tag{23}$$

Through the forward reasoning of Bayesian network, the posterior probability $P\{L \leqslant d\}$ of lateral overlap of two planes on parallel routes can be obtained, then the lateral collision risk is Zhang et al. (2007)

$$R_y = \frac{1}{2S_x} P_z(0) P_y(S_y)(V_x + V_y) + \frac{d}{S_x} P_y(S_y) N_z(0) \tag{24}$$

In the equation, R_y is the risk of lateral collision, S_x is the given separation standard, $P_z(0)$ is the probability of vertical overlap between the two aircraft, V_x and V_y are the relative speeds of the aircraft in the longitudinal and lateral directions respectively, and $N_z(0)$ is the frequency of vertical overlap between the two aircraft, and $P_y(S_y) = P\{L \leqslant d\}$.

5. Results and discussion

Assuming that both aircraft are Boeing 777, the wingspan d is 64.8 m, the fixed route lateral distance D_1 is 20n miles, and the fixed route lateral distance D_2 is 50n miles. The error standard deviation of the fixed route and the flexible route are $\sigma_1 = 2834.079, \sigma_2 = 11757.42$, respectively in 2.1. Substituting into Eqs. (22) and (23), the prior probability of the fixed route event A_7 is 3.7×10^{-21}, and the prior probability of the flexible route event A_7 is 5.7×10^{-10}.

The prior probabilities for each event are shown in Table 7 below, "Event failure or occurrence" is recorded as 0; "Event is operating normally or does not occur" is recorded as 1.

Through Bayesian network model, the lateral overlap rate of two aircraft in a fixed route $P_{fy}(S_y) = 2.0 \times 10^{-21}$ and the lateral overlap rate in a flexible route $P_{Fy}(S_y) = 9.91 \times 10^{-11}$ are obtained.

According to the technical parameters of the aircraft, the cruising speed of Boeing 777 is 900 km/h. When the yaw angle of two aircrafts are 3°and 5°, V_x is 2.19 km/h and V_y is 125.54 km/h, respectively. Fig. 11 shows the state. Taking $P_z(0)$ as 0.39 (Anderson and Lin, 1996) and $N_z(0)$ as 60 aircraft/h (Ying and Xu, 2002), submitting all the parameters into Eq. (24), the lateral collision risk of fixed route is $R_{fy} = 4.19 \times 10^{-20}$, and the lateral collision risk of flexible route is $R_{Fy} = 1.98 \times 10^{-9}$.

When the lateral distance of the route is changed, the change of lateral collision risk is shown in Fig. 12. As can be seen from the figure, when the lateral separation of the fixed route is reduced to 14n miles and the flexible route is reduced to 49n miles, the lateral collision risk can still meet the TLS of 5×10^{-9} times/hour stipulated by the ICAO.

According to the calculation results, the lateral collisions risk of flexible routes is much greater than fixed routes. This is due to the presence of navigation stations on fixed routes, which can provide accurate course guidance for aircraft. But on the flexible route, the aircraft can only perform flight tasks according to the waypoints in the flight plan, which is easy to deviate from the route and cause lateral flight errors.

In this paper, for the first time, the space-based ADS-B data is applied to the calculation of the lateral collision risk of parallel routes in the ocean area. It could fill the gap that there is no data support when calculating the collision risk of the ocean area. Further

Table 7
Event Prior Probability.

	0	1
A_1	0.0017	0.9983
A_2	0.0001	0.9999
A_3	0.001	0.999
A_4	0.003	0.997
A_5	10^{-6}	$1–10^{-6}$
A_6	0.55/0.17	0.45/0.83
A_7	$3.7 \times 10^{-21}/5.7 \times 10^{-10}$	$1–3.7 \times 10^{-21}/1–5.7 \times 10^{-10}$
A_8	0.0003	0.9997

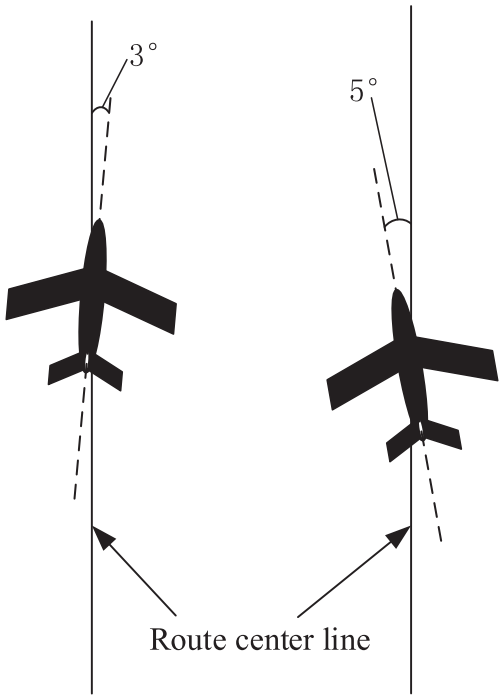

Fig. 11. The state of motion of the two aircraft.

development will focus on the optimization of the collision risk model, and the influence of more factors on the collision risk, such as weather and human factors, should be considered.

6. Conclusion

(1) For the fixed route track information, the segmented Kalman filter can effectively smooth the track and supplement the track information points. For flexible routes, the cubic polynomial interpolation is selected to process the track with the highest correlation without increasing the calculation amount.
(2) The error between track information after processing and route centerline is statistically analyzed. The distribution of track error on both types of routes obeys normal distribution, and the error on fixed routes is relatively average, while the error on flexible routes is polarized.
(3) Introducing the relevant parameters of error analysis into the collision risk model based on Bayesian network can effectively calculate the lateral collision risk of parallel routes in the ocean area, and the calculation results show that the lateral separation standard of parallel routes in the ocean area can meet the specified TLS.
(4) From the calculation results in this paper, the actual lateral interval of the Pacific flexible route is 50n miles (Jiang, 2010), which meets the safety target level.

CRediT authorship contribution statement

Fei Lu: Conceptualization, Methodology, Validation, Resources, Writing - original draft, Writing - review & editing, Visualization,

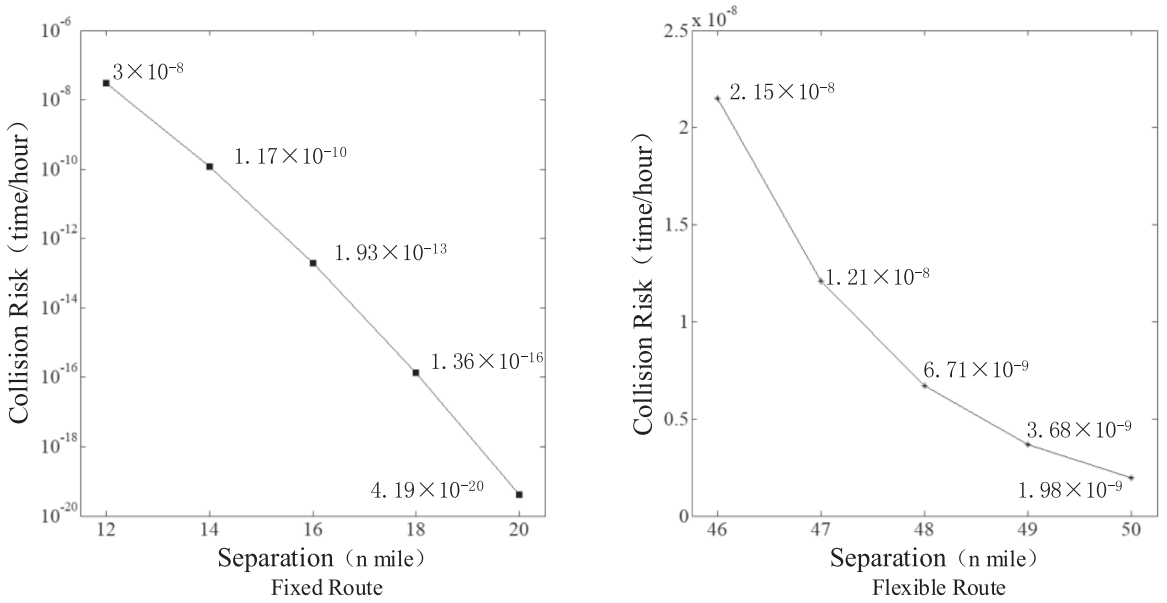

Fig. 12. Lateral collision risk at different separations.

Funding acquisition. **Zichen Chen:** Conceptualization, Software, Validation, Data curation, Data curation, Visualization. **Huiyu Chen:** Software, Validation, Formal analysis, Data curation, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China.

Formatting of funding sources

Funding: This work was supported by the National Natural Science Foundation of China [grant number 71701202];

References

Amrin, A., Zarikas, V., Spitas, C., 2018. Reliability analysis and functional design using Bayesian networks generated automatically by an “Idea Algebra” framework. Reliab. Eng. Syst. Saf. 180, 211–225. https://doi.org/10.1016/j.ress.2018.07.020.

Anderson, D., Lin, X.G., 1996. Collision Risk Model for a Crossing Track Separation methodology. J. Navig. 49 (3), 337–349. https://doi.org/10.1017/S0373463300013576.

Bakker, G., Blom, H., 1993. Air traffic collision risk modelling. In: Proceedings of 32nd IEEE Conference on Decision and Control, pp. 1464–1469.

Bakker, G., Kramer, H., Blom, H., 2000. Geometric and probabilistic approaches towards conflict prediction. Proceedings of 3rd USA/Europe Air Traffic Management R&D Seminar, Italy.

Blom, H., Bakker, G., Blanker, P., Daams, J., Everdij, M., Klompstra, M., 1998. Accident risk assessment for advanced ATM. Proceedings of 2nd USA/EuropeAir Traffic Management R&D Seminar, USA.

Blom, H., Bakker, G., Everdij, M., van der Park, M., 2003. Collision risk modelling of air traffic. Proceedings of European Control Conference, United Kingdom.

Brooker, P., 1983. Aircraft Separation Assurance: Systems Design. J. Navigation 36 (1), 28–36. https://doi.org/10.1017/S0373463300028496.

Brooker, P., 2002a. Future Air Traffic Management: Quantitative En Route Safety Assessment Part 1 – Review of Present Methods. J. Navigation 55 (2), 197–211. https://doi.org/10.1017/S0373463302001716.

Brooker, P., 2002b. Future Air Traffic Management: Quantitative En Route Safety Assessment Part 2 – New Approaches. J. Navigation 55 (3), 363–379. https://doi.org/10.1017/S037346330200187X.

Brooker, P., 2003. Lateral Collision Risk In Air Traffic Track Systems: A ‘Post-Reich’ Event Model. J. Navigation 56 (3), 399–409. https://doi.org/10.1017/S0373463303002455.

Brooker, P., 2006. Longitudinal Collision Risk for ATC Track Systems: A Hazardous Event Model. J. Navigation 59 (1), 55–70. https://doi.org/10.1017/S0373463305003516.

Cathey, F.W., Dailey, D.J., 2003. A prescription for transit arrival/departure prediction using automatic vehicle location data. Transport. Res. Part C Emerg. Technol. 11 (3-4), 241–264. https://doi.org/10.1016/S0968-090X(03)00023-8.

Chen, X., Shen, C., Zhang, W.-B., Tomizuka, M., Xu, Y., Chiu, K., 2013. Novel hybrid of strong tracking Kalman filter and wavelet neural network for GPS/INS during GPS outages. Measurement 46 (10), 3847–3854. https://doi.org/10.1016/j.measurement.2013.07.016.

Choi, E.-J., Yoon, J.-C., Lee, B.-S., Park, S.-Y., Choi, K.-H., 2010. Onboard orbit determination using GPS observations based on the unscented Kalman filter. Adv. Space Res. 46 (11), 1440–1450.
Cox, M.E., Ten Have, J.M., Forrester, D.A., 1991. European studies to investigate the feasibility of using 1000 ft vertical separation minima above FL 290, parts I. J. Navig. 44 (2), 171–183. https://doi.org/10.1017/S0373463300009917.
Davies, L., Sharpe, A., 1993. Review of the North Atlantic Lateral Collision Risk Model. Air Traffic Control Quart. 1 (3), 225–253. https://doi.org/10.1128/JB.05469-11.
EUROCAE ED 126, Safety, Performance and Interoperability Requirements Document for ADS-B-NRA Application.
Harrison, D., Moek, G., 1992. European studies to investigate the feasibility of using 1000 ft vertical separation minima above FL 290, parts II. J. Navig. 45 (1), 91–106.
Huang, B.J., 2012. Lateral Collision Risk Model on Parallel Routes Based on Communication, Navigation and Surveillance Performances. Journal of Southwest Jiaotong University 47(6), 1075-1080+1091. https://doi.org/10.3969/j.issn.0258-2724.2012.06.026.
Huang, K., 2013. Research on Security Planning of Air Traffic Control Equipment Based on Complex Network Theory. Nanjing University of Aeronautics and Astronautics.
Jiang, Z.F., 2010. Introduction of flexible routings in Pacific area. Air Traffic Management 1, 45–48.
Jones, B., Jenkinson, I., Yang, Z., Wang, J., 2010. The use of Bayesian network modelling for maintenance planning in a manufacturing industry. Reliab. Eng. Syst. Saf. 95 (3), 267–277. https://doi.org/10.1016/j.ress.2009.10.007.
Li, L., Quddus, M., Zhao, L., 2013. High accuracy tightly-coupled integrity monitoring algorithm for map-matching. Transport. Res. Part C Emerg. Technolog. 36, 13–26. https://doi.org/10.1016/j.trc.2013.07.009.
Moek, G., Ten Have, J.M., Harrison, D., et al., 1993. European studies to investigate the feasibility of using 1000 ft vertical separation minima above FL 290, parts III. J. Navig. 46 (2), 245–261.
Netjasov, F., 2012a. Framework for airspace planning and design based on conflict risk assessment. Transport. Res. Part C Emerg. Technolog. 24, 190–212. https://doi.org/10.1016/j.trc.2012.03.002.
Netjasov, F., 2012b. Framework for airspace planning and design based on conflict risk assessment. Transport. Res. Part C Emerg. Technolog. 24, 213–226. https://doi.org/10.1016/j.trc.2012.03.003.
Netjasov, F., Babić, O., 2013. Framework for airspace planning and design based on conflict risk assessment. Transport. Res. Part C Emerg. Technolog. 32, 31–47. https://doi.org/10.1016/j.trc.2013.04.002.
Pérez-Castán, J.A., Gómez Comendador, F., Rodríguez-Sanz, A., Armas Cabrera, I., Torrecilla, J., 2019. RPAS conflict-risk assessment in non-segregated airspace. Saf. Sci. 111, 7–16. https://doi.org/10.1016/j.ssci.2018.08.018.
Reich, P.G., 1966a. Analysis of Long-Range Air Traffic Systems-Separation Standards I. J. Navig. 19 (1), 88–98. https://doi.org/10.1017/S037346330004056X.
Reich, P.G., 1966b. Analysis of Long-Range Air Traffic Systems-Separation Standards II. J. Navig. 19 (2), 169–186. https://doi.org/10.1017/S0373463300047196.
Reich, P.G., 1966c. Analysis of Long-Range Air Traffic Systems-Separation Standards III. J. Navig. 19 (3), 332–347. https://doi.org/10.1017/S0373463300047445.
Shen, Z., Georgy, J., Korenberg, M.J., Noureldin, A., 2011. Low cost two dimension navigation using an augmented Kalman filter/Fast Orthogonal Search module for the integration of reduced inertial sensor system and Global Positioning System. Transport. Res. Part C Emerg. Technolog. 19 (6), 1111–1132. https://doi.org/10.1016/j.trc.2011.01.001.
Shortle, J.F., Xie, Y., Chen, C.H., Donohue, G.L., 2004. Simulating Collision Probabilities of Landing Airplanes at Nontowered Airports. Simulation 80 (1), 21–31.
Wang, X., Xu, X.H., 2001. Study of Lateral Separation in Air Traffic Control. J. Civil Aviation Univ. China 19 (1), 1–5.
Wang, Y.W., 2018. Research of Land-based and Space-based ADS-B System Data Quality. Civil Aviation University of China.
Ying, A.L., Xu, X.H., 2002. Reich Model in Collision Risk Study of Airspace Flight. J. Civil Aviation Univ. China 20 (4), 6–10.
Yin, X.W., Qian, W.X., Xie, L.Y., 2008. A Method for System Reliability Assessment Based on Bayesian Networks. Acta Aeronautica ET Astronautica Sinica 29 (6), 1482–1489. https://doi.org/10.1007/s11123-007-0067-1.
Zhang, Z.N., Liang, Y.W., 2014. Bayesian Network-based Study on Collision Risk in Free Flight. China Safety Sci. J. 24 (9), 40–45. https://doi.org/10.16265/j.cnki.issn1003-3033.2014.09.016.
Zhang, Z.N., Zhang, X.Y., Li, D.B., 2007. Computation Model of Lateral Collision Rate on Parallel Routes Based on VOR Navigation. J. Traffic Transportation Eng. 7 (3), 21–24.