

4D Flight Trajectory Prediction Using CURE-Based Clustering of ADS-B Data

Qinghai Zuo^{1*}, Shuai Han², Weijun Pan^{3,4}

¹College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618307, China

²Operation Supervisory Center of CAAC, Beijing 100710, China

³Scientific Research Base, Civil Aviation Flight University of China, Guanghan 618307, China

⁴Key Laboratory of Flight Techniques and Flight Safety of CAAC, Guanghan 618307, China

E-mail: zuoatc@163.com

*Corresponding author

Keywords: flight trajectory, CURE algorithm, cluster analysis, ADS-B

Received: March 14, 2025

Aiming at the problems of insufficient flight trajectory prediction accuracy and low early warning efficiency of air traffic conflicts under the background of the sharp increase in air traffic volume, a 4D trajectory prediction method based on the clustering using representatives (CURE) algorithm is studied and proposed. The method utilizes historical track data provided by the Automatic Dependent Surveillance-Broadcast (ADS-B) system to measure track similarity by means of modified Euclidean distance. It is also combined with hierarchical clustering techniques to cluster and analyze the tracks within the terminal area. Experiments showed that the proposed method outperformed the traditional model in many aspects. For example, in track prediction, the CURE-based model had an average time error of only 17.6 seconds, a height error of 31.5 m, and a horizontal root mean square error (RMSE) of 0.55 nautical miles. Furthermore, using conflict detection based on 4D planes and geometric methods successfully reduced the false alarm rate to 4.30% and controlled the missed alarm rate within 0.08%. Together with the experimental data, these results verify the effectiveness and reliability of this method in complex aviation scenarios. This indicates that the CURE algorithm can improve track prediction accuracy and provide stronger technical support for aviation traffic management.

Povzetek: Napovedovanje 4D avionskih trajektorij je izvedeno z URE-grozdjenjem iz ADS-B; predlagane spremembe povečajo rigoroznost in uporabnost.

1 Introduction

In recent years, with the rapid growth of global air traffic, air traffic management is facing unprecedented pressure and challenges. Effective trajectory prediction technology can not only enhance the utilization of airspace resources and reduce the workload of ground controllers, but also significantly improve the safety of flight operation. Therefore, 4D trajectory prediction method is of great significance in modern aviation management. 4D trajectory prediction technology accurately predicts an aircraft's flight trajectory by combining information from four dimensions: time, space, speed, and altitude. This technology provides effective support for air traffic management. Currently, based on the data from the GPS automatic dependent surveillance-broadcast (ADS-B) system, it is possible to obtain information such as the position, speed, and heading of the aircraft in real time, which provides a reliable data base for trajectory prediction. Traditional trajectory prediction methods mainly rely on ground radar and flight plans. However, these methods have certain limitations in terms of prediction accuracy and real-time performance [1-2]. With the popularization and application of ADS-B system, the trajectory prediction technology based on ADS-B data has gradually become a research hotspot. The 4D trajectory

prediction technology has been widely studied at home and abroad. Many researchers try to improve the accuracy and reliability of trajectory prediction by different methods. Aiming at the problem of low accuracy of short-term prediction of flight trajectory, Yang et al. proposed a bidirectional long-short memory network prediction method based on broadcast auto-correlation surveillance historical data. Experimental results indicated that the method was proved to improve aviation safety in busy airspace [3]. Wang et al. proposed a generalized hybrid recurrent prediction model for flight trajectory prediction. The results showed that the generalized deep learning method not only improved the accuracy of trajectory prediction, but also allowed contextualization by exploring a large amount of data [4]. Han et al. proposed a trajectory prediction method combining a density-based spatial clustering algorithm with noise and a gated loop cell for trajectory prediction. The results indicated that this method could effectively utilize the trajectory data in the terminal area, and the model developed could perform trajectory prediction for multiple flights and improve the accuracy of trajectory prediction [5]. Dai et al. proposed that a deep neural network model based on Kalman filter algorithm unfolding can be used for aircraft trajectory prediction. The results of simulation experiments

indicated that the developed model provided better accuracy and effectiveness for aircraft trajectory prediction than other network models [6]. Rizvi et al. proposed a meta-learning approach to predict short and medium-term aircraft trajectories using historical real flight data collected from multiple genetic algorithm aircraft. The study used random forest regression and long and short-term memory networks to extract aircraft trajectory features and k-nearest neighbors were used to complete the final prediction. The model was shown to have good prediction results in the experimental results

[7]. Chakrabarti et al. used a hidden Markov model to identify and extract heading changes within aircraft trajectories, followed by comparing and clustering the trajectory strings using an edit distance metric combined with the K-medoids clustering algorithm. An application to a set of historical trajectories at Washington National Airport demonstrated the success of the proposed framework in overcoming the shortcomings associated with traditional clustering techniques [8]. The literature review is specifically shown in Table 1.

Table 1: Literature review table.

References	Method	Advantages	Disadvantages
Reference [3]	Bi-LSTM	Be capable of capturing long sequence dependencies. It has a remarkable effect on short-term prediction	The training time is relatively long. Sensitive to hyperparameter Settings
Reference [4]	Generalized mixed cycle model	The accuracy of trajectory prediction has been improved. It has strong generalization ability and can adapt to various data situations	The model is complex and parameter adjustment is difficult. It has high requirements for the volume of data
Reference [5]	DBSCAN-GRU	It has a strong ability to handle high-density data and noise. Adaptive clustering	Parameter selection is sensitive. The clustering effect is uneven for regions with different densities
Reference [6]	Deep neural network	Contextualization can be achieved by exploring a large amount of data	The training is complex and the computational overhead is high. A large amount of labeled data is required for training
Reference [7]	Meta-learning method	Applicable to various situations. It is capable of efficiently utilizing historical data for medium and short-term trajectory prediction	The adaptability to specific problems remains to be verified. High complexity
Reference [8]	Hidden Markov model	The model is simple and easy to understand. It is relatively effective in sequential data	It may be impossible to capture long dependencies. Be sensitive to the initial Settings

Although existing track prediction models, such as Bi-LSTM and generalized mixed loop models, perform well in capturing the short-term dependencies of time series data, they often have difficulty fully handling the complexity of high-dimensional data. This is especially true when it comes to effectively identifying potential patterns in aviation trajectories and performing cluster analysis. Meanwhile, methods such as DBSCAN-GRU and Hidden Markov models are relatively vulnerable to data noise during the clustering process, making it difficult to accurately distinguish the trajectories of different categories. The clustering using representatives (CURE) algorithm is a clustering method that aims to accurately represent the shape and distribution of clusters. It does so by selecting multiple representative points. This allows it to overcome the limitations of traditional clustering methods when dealing with high-dimensional data and complex shapes. By dynamically correcting the similarity calculation and hierarchical clustering methods, the CURE method overcomes the influence of data noise on the clustering results. This improves the accuracy and reliability of flight trajectory prediction. These improvements allow the CURE method to more effectively capture the complex features and potential patterns of aviation trajectories, providing better technical support for flight scheduling and air traffic management. This study aims to explore the potential of the CURE algorithm to improve the accuracy of 4D trajectory prediction in the terminal area. Specific questions include: Can the CURE algorithm significantly improve prediction

accuracy in complex aviation trajectory data environments? Meanwhile, the research will conduct an in-depth analysis of how trajectory similarity estimation using the modified Euclidean distance influences the accuracy of flight conflict prediction. This analysis will help evaluate the effectiveness of the CURE algorithm in air traffic management, providing a theoretical basis and practical guidance for developing future air track prediction methods.

2 Methods and materials

2.1 4D trajectory prediction method and surveillance technology

Currently there are more prediction methods for trajectories. The 4D trajectory prediction method is of far-reaching significance in air traffic management, so the study uses the 4D trajectory prediction technique to predict the flight trajectory. 4D trajectory mainly refers to the main four pieces of information in the airplane trajectory. By recording the time, space, speed and altitude information of aircraft in the air, real-time monitoring and management of aircraft can be achieved. These 4D trajectory data can be used to analyze aircraft flight performance, route planning, aviation safety and other aspects. The airplane flight mainly consists of five directions: take off, departure, cruise, arrival, approach and landing. The whole process of flight is shown in Figure 1.

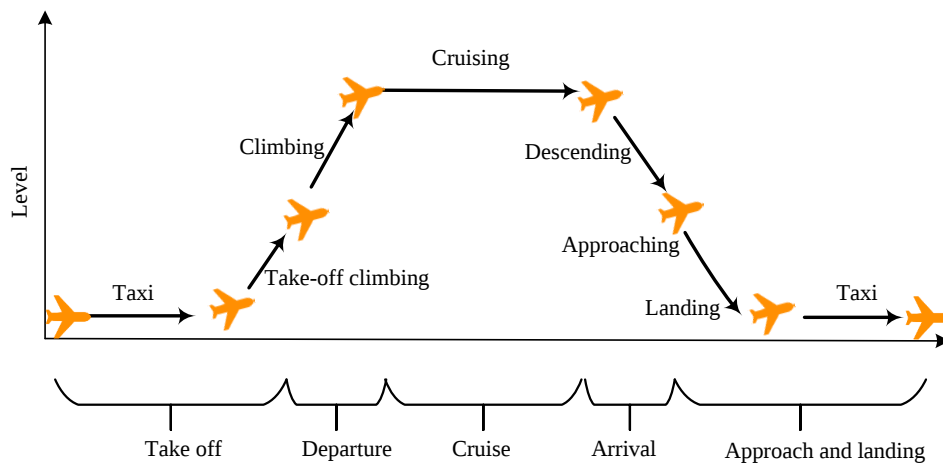


Figure 1: Flight process flow chart.

Figure 1 shows the whole process of an airplane reaching the cruising level through the take-off and climbing process, descending and landing after a period of cruise process. In terms of trajectory prediction, if only a single influencing factor is considered, it will result in a lack of accuracy and comprehensiveness of the actual trajectory of the airplane. Therefore, a comprehensive analysis and prediction should be made by considering a variety of factors, including geographic location, wind speed, temperature and other information. Equiangular routing is the process of route planning according to a certain latitude or longitude angular distance equidistant on a map. In aviation, this usually refers to the fact that routes are set up with fixed points at certain latitude or longitude intervals to facilitate pilots' navigation and localization in flight. This makes each route segment relatively equal in length on the map. Usually, equiangular routes fly longer distances than great circle routes [9]. The starting point $B(\Phi_B, \lambda_B)$ and the starting point $A(\Phi_A, \lambda_A)$ are defined. The angle between the equiangular route and the meridian is a . The distance between the starting points is $AB = H$. At this point, according to the trigonometric function can be obtained

route angle a and route distance H , as shown in Equation (1).

$$\begin{cases} \tan a = \frac{\lambda_A - \lambda_B}{\ln \tan(\frac{\pi}{4} + \frac{\phi_A}{2}) - \ln \tan(\frac{\pi}{4} + \frac{\phi_B}{2})} \\ H = (\phi_A - \phi_B) \sec a \quad \text{or} \quad (\lambda_A - \lambda_B) \cos \bar{\phi} \sec a \end{cases} \quad (1)$$

In Equation (1), $\bar{\phi}$ represents the average latitude value between A and B . π represents PI, which is approximately equal to 3.14159. It is a mathematical constant. In aviation navigation and trajectory prediction, " π " is used to calculate angles, distances, and other parameters related to circular navigation trajectories. When the difference in latitude is large, $\tan a$ in Equation (1) is used. When the difference in latitude is small, H is used, and the unit of H needs to be converted from radians to nautical miles. In the process of calculation, based on the calculated a angle to take the corresponding absolute value. When $a < 90^\circ$, based on the position of the course position at this time, the data of the course angle can be calculated, as shown in Figure 2.

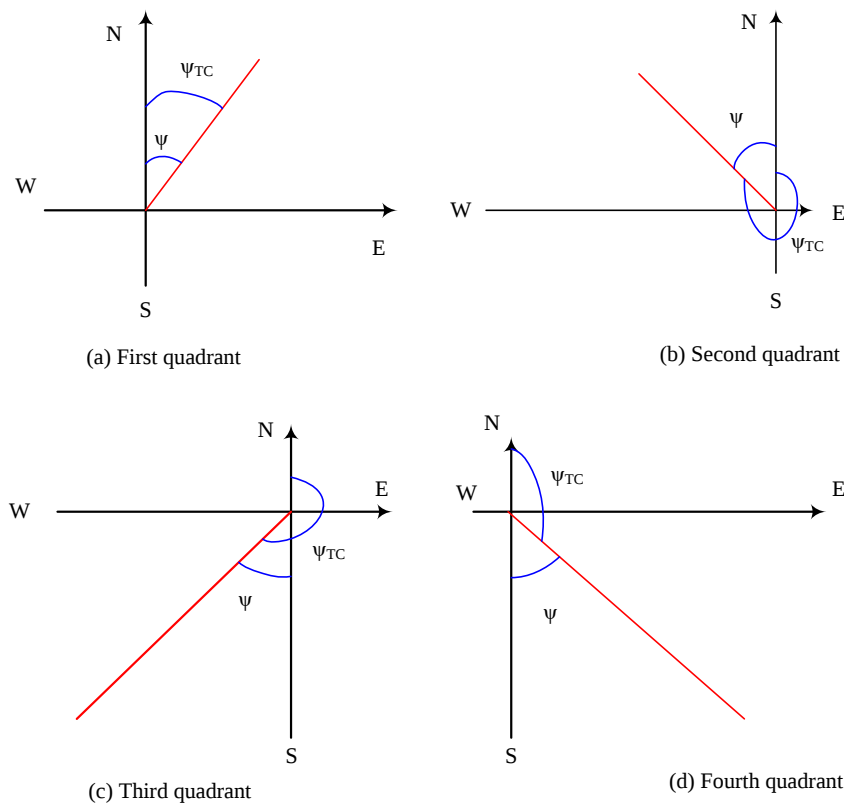


Figure 2: Correlation between route angle and quadrant.

In Figure 2, ψ represents the heading Angle of the equal-angle route, and ψ_{TC} represents the corrected actual heading Angle. When in the first quadrant, ψ_{TC} is identical to ψ . When in the second and fourth quadrants, ψ_{TC} and ψ are circumferential as well as complementary to each other. When in the third quadrant, ψ_{TC} is $\psi + \frac{\pi}{2}$. When $B(\Phi_B, \lambda_B)$, the starting point $A(\Phi_A, \lambda_A)$, the corresponding a and H can be obtained by referring to the trigonometric functions, as shown in Equation (2).

$$\begin{cases} \cot a = \frac{\sin \phi_A \tan \phi_B - \sin \phi_A \cos(\lambda_B - \lambda_A)}{\sin(\lambda_B - \lambda_A)} \\ \cos H = \sin \phi_A \sin \phi_B + \cos \phi_A \cos \phi_B \cos(\lambda_B - \lambda_A) \end{cases} \quad (2)$$

Equation (2) defines the calculation methods for the heading angle and route distance of an equal-angle route. Its core role is to provide a mathematical model for an equal-angle route. It converts the difference between longitude and latitude into actual navigation parameters using triangular geometric relations. This solves the symbol correction problem of heading angles in different quadrants, such as northeast and southeast. The study takes the northeast hemisphere as the first quadrant, the southeast hemisphere as the fourth quadrant, the northwest hemisphere as the second quadrant, and the southwest hemisphere as the third quadrant. When the route is in the first and fourth quadrants, then $a_{TC} = 2\pi + a$. When the route is in the second and third quadrants, then

$a_{TC} = \pi + a$. Finally, H is converted to nautical miles by the corresponding transformation [10]. Trajectory data has shortcomings such as spatio-temporal correlation, multi-dimensionality, high frequency, and large data volume. By analyzing historical trajectory data, the patterns and laws of aircraft flight are mined to predict future trajectories [11]. As the 4D trajectory prediction technique of data mining is characterized by simple operation, high accuracy and simpler principle, it makes it widely used in the field of trajectory prediction [12]. The study adopts this method for flight trajectory prediction.

2.2 Cluster analysis prediction model based on CURE algorithm

The 4D trajectory prediction method is based on the historical data of ADS-B, and ADS-B reaches the automatic surveillance through the GPS system, the ground-to-air system, and the air-to-air three-dimensional data. ADS-B uses GPS to provide real-time position information of aircraft and broadcasts this information to other aircraft and ground control centers via ground-to-air and air-to-air communication systems. Therefore, GPS is an integral part of the ADS-B system, serving as the basis for providing accurate three-dimensional data, such as position, altitude, and speed. Compared with other surveillance radars, ADS-B is able to provide real-time aircraft position and status data at a frequency of multiple updates per second. It can provide more accurate aircraft position and speed data. It is not limited by terrain or geography and can cover a wider area. Based on ADS-B,

data such as latitude, longitude, heading and speed of the aircraft are preprocessed. Duplicate data points are eliminated after categorizing the flight number. Subsequently after screening and deleting data bars under a certain flight altitude, data bars outside a certain range of the receiver and flight numbers with insufficient data bars are also screened and deleted. "The receiver" refers to the device or system that collects and processes the data transmitted from the ADS-B system. Finally, the corresponding trajectory sequence is regenerated. The preprocessed data are categorized and identified, and the cumulative trajectory sequence is obtained. The basic principle of determining trajectory points by the correlation trajectory method is to first determine the overall route of the aircraft, including the departure point, intermediate waypoints and destination. Second, according to the flight plan and flight performance data, the trajectory points are calculated by using the correlation trajectory method, i.e., the trajectory points that need to be passed during the flight. Finally, during the flight, the pilot will correct and adjust the trajectory according to the actual situation. The aerodynamics on the flight path should satisfy Equation (3).

$$\begin{cases} v_{\min} \leq \frac{|x_i - X|}{|t_i - t_x|} \leq v_{\max} \\ v_{\min} \leq \frac{|x_i - X|}{|t_i - t_x|} \leq v_{\max 1} \end{cases} \quad (3)$$

In Equation (3), x_i denotes the ADS-B target position. X denotes the trajectory. M denotes information. The horizontal velocity of the trajectory should be greater than or equal to $v_{\min}$ and less than or

equal to the maximum value of the trajectory $v_{\max}$ of X . The vertical speed of the associated trajectory is the same as the horizontal speed of the trajectory. In addition, the course angle a needs to satisfy certain conditions, as shown in Equation (4).

$$\begin{cases} |a| \leq a_0 \\ a = \arccos\left[\frac{(x_{i+1} - x_i)(x_i - x_{i-1})}{|x_{i+1} - x_i| |x_i - x_{i-1}|}\right] \end{cases} \quad (4)$$

In Equation (4), vector a is the angle between $x_{i+1} - x_i$ and $x_i - x_{i-1}$. In general, to improve the matching probability of the trajectory, a_0 should adopt a larger value. The research focuses on the standardized velocity and route angle constraints in formulas (3) and (4). It presents a table showing the actual value range and source basis, as shown in Table 2.

After normalizing the time [13], the sampling period is shown in Equation (5) based on the historical flight record.

$$T' = \frac{T_p}{t_i} T \quad (5)$$

T' is the sampling time and T_p is the prediction time. T_i is the total flight time on day i . T is the original sampling period of 4s. The trajectory clustering results can further reflect the correlation between the data samples. Through clustering analysis, the logical relationship between each dataset, the classification pattern and the trajectory data can be obtained. The steps of the clustering algorithm are shown in Figure 3.

Table 2: Aviation performance constraints and data sources.

Parameter	Value range	Source	Applicable phase
Minimum horizontal velocity	130 kt (67 m/s)	ICAO Doc 8168 Vol I	Approach/Wait (H ≤ 10,000 ft)
Maximum horizontal speed	320 kt (165 m/s)	ADS-B Measured Data Statistics (Inbound Flights at Pudong Airport)	Cruise (H ≥ 24,000 ft)
Vertical velocity constraint	[-3, +3] m/s	FAA AC 120-29A	Climb/descend
Maximum deviation of the flight angle	±15°	ICAO Annex 11	Flight path
Minimum turning radius	3 nmi (5.56 km)	EUROCONTROL Base of Aircraft Data	Terminal area mobility
Trajectory matching probability threshold	≥0.85	Flight Procedure Verification Specifications of the Civil Aviation Administration of China	All stages

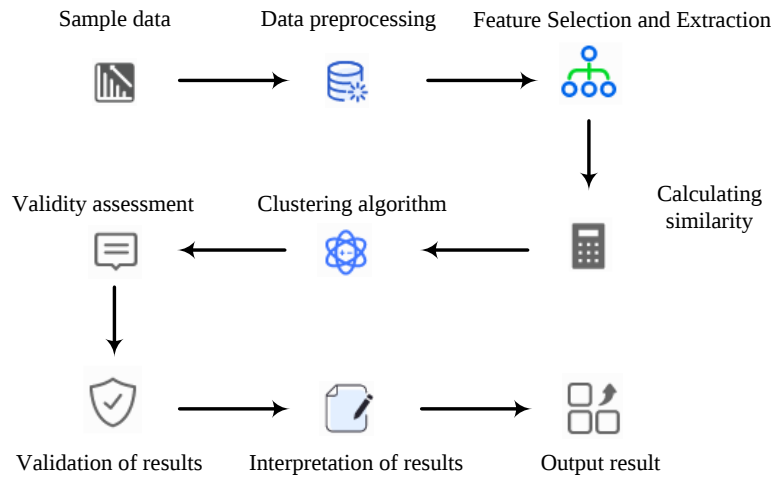


Figure 3: Clustering algorithm process steps.

In Figure 3, based on the obtained sample data is selected through preprocessing, feature extraction. The final result is then output after the relevant data is evaluated, verified, and interpreted through the calculation of similarity and combination with a clustering algorithm. The research selects the hierarchical clustering method in the clustering method. This method is hierarchical, interpretable and does not require pre-specification of the

number of clusters [14-15]. Hierarchical clustering method has a good ability to analyze and deal with large amount of data as well as complex changes in the data. At this stage, it has been shown to be used for analyzing hierarchical clustering of flight trajectory data [16]. In the similarity calculation, the study uses Euclidean distance for calculation. Its similarity is shown in Equation (6).

$$\begin{cases} \text{similarity}(A_1, A_2) = \sqrt{\varphi_1(x_1 - y_1)^2 + \varphi_2(x_2 - y_2)^2 + \dots + \varphi_n(x_n - y_n)^2} \\ \sum_{k=1}^n \varphi_k = 1 \end{cases} \quad (6)$$

In Equation (6), A denotes the aerial trajectory points. These trajectory points are for precise feature extraction and subsequent analysis of the trajectory points to ensure an accurate description of the aircraft's position. n denotes the number of dimensions. φ_k denotes the weight of different dimensions. The weight factor is used to adjust the importance of each dimension in the similarity calculation. It reflecting the extent to which different parameters affect the trajectory characteristics. $\text{similarity}(A_1, A_2)$ denotes the dimension attribute difference, which means the difference between two trajectory points in different dimensions, that is, the specific change in position, velocity, height, etc. Since the Euclidean distance is greatly affected by the trajectory data as well as noise, further corrections are needed when using the Euclidean distance for trajectory calculation. The study uses the point-by-point method to calculate the route distance, and then adjusts the Euclidean distance similarity calculation using the mean value method. Finally, the trajectory is corrected according to the calculated distance variance to obtain the final trajectory similarity. The corrected trajectory calculation is shown in Equation (7).

$$t_i^* = \begin{bmatrix} G_1^{(i)} \\ G_2^{(i)} \\ \vdots \\ G_k^{(i)} \\ \vdots \\ G_M^{(i)} \end{bmatrix} \quad t_j^* = \begin{bmatrix} G_1^{(j)} \\ G_2^{(j)} \\ \vdots \\ G_k^{(j)} \\ \vdots \\ G_M^{(j)} \end{bmatrix} \quad (7)$$

In Equation (7), t_i^* denotes the flight path trajectory. M is the number of trajectory points contained in trajectories t_i^* and t_j^* . $G_k^{(i)}$ and $G_k^{(j)}$ represent a point pair. Therefore, the equirectangular air trajectory between the point pairs can be expressed as $x_k = d(G_k^{(i)}, G_k^{(j)})$. At this time, the set of distances of each point pair is $X = \{x_1, x_2, x_3, \dots, x_M\}$. The similarity is calculated as shown in Equation (8).

$$s_{ij} = \frac{\sum_{k=1}^p x_k^k}{M} + D(x) + \Delta \quad (8)$$

In Equation (8), $D(X)$ represents the structural similarity measure and Δ represents the dynamic correction factor. p represents the number of trajectory points involved in the similarity calculation. Equation (6) has two main flaws in the calculation of aviation trajectory similarity: noise sensitivity and density difference.

After correcting these defects in the research, Equation (8) is obtained. The main derivation of Equation (8) is as follows: assuming that the trajectory contains M trajectory points, the normalization factor $M = \max(m, \bar{m})$ is defined. Among them, $\bar{m}$ is the average number of trajectory points in the dataset. Then the density normalization term EEE is introduced. The density normalization term scales the trajectory point coordinates by $1/M$, which eliminates cumulative error caused by differences in sampling frequencies. The trajectory structure similarity measurement extracts geometric features of trajectories based on principal component analysis. Equation (9) shows the covariance matrix factorization of the trajectory point set.

$$\sum X = \frac{1}{m} (X - \mu_x)^T (X - \mu_x) \quad (9)$$

The research assumes that the eigenvalues λ_1 , λ_2 , and λ_3 represent the energy distribution of the trajectory in the main direction in three-dimensional space. The structural similarity measure is defined as shown in Equation (10).

$$D(X) = \frac{\lambda_2 + \lambda_3}{\lambda_1} \quad (10)$$

In Equation (10), when $D(X)$ approaches 0, it indicates that the trajectory has the characteristic of linear motion. When $D(X)$ approaches 1, it indicates a drastic change in the trajectory direction. Finally, combined with the characteristics of the typing stage, the dynamic correction factor is defined as shown in Equation (11).

$$\Delta = \alpha \cdot |\gamma_x - \gamma_y| + \beta \cdot |\omega_x - \omega_y| \quad (11)$$

In Equation (11), Δ represents the dynamic correction factor. γ represents the climb rate. ω represents the turning angular velocity. α represents the weight coefficient of the climb rate. β represents the turning angular velocity weight. Based on the above correction items, Equation (8) for trajectory similarity calculation is obtained. The similarity matrix S can be obtained after the operation based on Equation (7), as shown in Equation (12).

$$S = \begin{bmatrix} 0 & s_{12} & \cdots & s_{1j} & \cdots & s_{1n} \\ s_{21} & 0 & \cdots & s_{2j} & \cdots & s_{2n} \\ \vdots & & \ddots & & & \vdots \\ s_{i1} & & & 0 & & s_{in} \\ \vdots & & & & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nj} & \cdots & 0 \end{bmatrix} \quad (12)$$

In Equation (12), since S is obtained based on the symmetric Euclidean distance operation, the matrix is also called symmetric matrix. Since the number of discrete points of each aerial trajectory is different, and it is difficult to have a corresponding situation in the data, the study uses the CURE algorithm in order to realize the clustering prediction of the aerial trajectories [17]. The CURE algorithm applied in the study is described below. Firstly the input defines the set of trajectories

$T^* = \{t_1^*, t_2^*, t_3^*, \dots, t_n^*\}$, representing points Q , and the number of clusters k . The second output is the clustering result of the flight trajectories $C = \{C_1, C_2, C_3, \dots, C_k\}$, with clustering centers $AVE = \{ave_1, ave_2, ave_3, \dots, ave_k\}$. Each category represents the aerial trajectory $Q_e = \{Q_{e1}, Q_{e2}, Q_{e3}, \dots, Q_{ek}\}$ of the point. An algorithm based on the modified Euclidean distance aerial trajectory similarity is used to calculate the similarity between 2 aerial trajectories in the trajectory set T^* . The trajectory similarity matrix is then constructed. The trajectory similarity matrix Q_T , as shown in Equation (13), is also constructed.

$$Q_T = \begin{bmatrix} 0 & Q_{12} & \cdots & Q_{1j} & \cdots & Q_{1n} \\ Q_{21} & 0 & \cdots & Q_{2j} & \cdots & Q_{2n} \\ \vdots & & \ddots & & & \vdots \\ Q_{i1} & & & 0 & & Q_{in} \\ \vdots & & & & \ddots & \vdots \\ Q_{n1} & Q_{n2} & \cdots & Q_{nj} & \cdots & 0 \end{bmatrix} \quad (13)$$

In Equation (13), after obtaining the similarity matrix Q_T , the initialization clustering is carried out. Subsequently, the center trajectories ave_a , ave_b of C_a and C_b are calculated. Meanwhile, the aerial trajectory sets Q_{ea} and Q_{eb} are calculated to determine whether the above two aerial trajectories are the aerial trajectory sets of the current clustering result. If they are different genus classes, merge the classes and define a new genus class C_s , and reduce the number of classes in the dynamic array by 1. Calculate the center flight trajectory ave_s of C_s and the representative trajectory set Q_{es} . Finally, the relevant results are obtained as shown in Equation (14).

$$\begin{cases} C = \{C_1, C_2, C_3, \dots, C_k\} \\ AVE = \{ave_1, ave_2, ave_3, \dots, ave_k\} \\ Q_e = \{Q_{e1}, Q_{e2}, Q_{e3}, \dots, Q_{ek}\} \end{cases} \quad (14)$$

C_k denotes the set of Class k flight trajectories. ave_k denotes the cluster center flight trajectory of C_k . Q_{ek} denotes the representative air trajectory set of C_k . The overall trend of flight operation can be somewhat reflected by the cluster trajectories obtained from the clustering of the CURE algorithm. In addition, the relevant information of the next flight trajectory can be obtained by correcting the historical flight trajectory data, as shown in Equation (15).

$$\begin{cases} C^p(\Gamma + h) = C^c(\Gamma + h) - \varepsilon(\Gamma) \\ \varepsilon(\Gamma) = C^c(\Gamma) - C^R \end{cases} \quad (15)$$

In Equation (15), h denotes the step size of the prediction time. C^p , C^c , and C^R denote the prediction, clustering, and real trajectory, respectively. $\varepsilon(\Gamma)$ denotes the correction value coefficient. Equation (15) uses the cluster center trajectory and historical trajectory data

obtained from the clustering of the CURE algorithm, combines the predicted value and the real trajectory, and makes a correction prediction for the next flight trajectory. The correction coefficients can be used to adjust the degree of contribution of the predicted value and the cluster center trajectory, as well as the degree of contribution of the predicted value and the true trajectory. In this way, a more accurate prediction of the next flight trajectory can be obtained, thus providing a more reliable information reference for the overall trend of flight operation. The modified Euclidean distance similarity measure proposed in the research is a custom algorithm module, and its core mathematical operations are implemented through NumPy in Python. The module is integrated with Scikit-learn's Agglomerative Clustering framework. The similarity calculation function is accessed via the custom affinity parameter. The experimental platform is built using the PyTorch Geometric spatio-temporal data processing library. The specific study of pseudo-code is shown in Figure 4.

3 Results

3.1 4D trajectory prediction based on CURE algorithm data mining

The proposed prediction model is designed based on real-time stream processing architecture and can be seamlessly

integrated into the existing air traffic control (ATC) system. The system accesses the ADS-B data stream in real time through Apache Kafka with an update frequency of 1 Hz. It dynamically extracts trajectory fragments using the sliding time window mechanism with a window length of 300 s and a sliding step length of 10 s. The system updates the track pattern online using the incremental CURE clustering algorithm with an update cycle of 30 s. The real-time prediction engine jointly uses Kalman filtering for state estimation and an LSTM time series prediction module for behavior reasoning. This generates 4D trajectory prediction results for the next 300 s in a pipeline manner and achieves synchronization of prediction states among multiple nodes through a distributed Redis database. The CURE algorithm combined with cluster analysis is used to analyze the actual ADS-B data of the approach flights of an airport in East China as the corresponding samples, and the cluster analysis is performed after the relevant data transformation. First, the study conducts different index analyses on various numbers of clusters to determine the optimal number of clusters. The indicators include the silhouette score (SS), the Davies-Bouldin index (DBI), the sum of squared errors (SSE), and the Calinski-Harabasz index (CHI). The specific results are shown in Table 3.

```

Define CURE_Algorithm(data, k, r):
  Preprocess data:
    Clean, filter, and normalize data

  Initialize clusters with initial representatives

  While number of clusters < k:
    For each trajectory in data:
      Assign trajectory to closest cluster based on modified distance

    Update representative points in each cluster

  Return clusters and representatives

Define Conflict_Detection(clustered_data):
  For each trajectory pair in clustered_data:
    If IsPotentialConflict(trajectory1, trajectory2):
      Add to conflict list

  Update metrics (TP, FP, FN, TN)
  Return Precision, Recall, F1-Score

Main:
  Load flight data
  Set parameters: k, r

  clustered_data = CURE_Algorithm(data, k, r)
  conflict_metrics = Conflict_Detection(clustered_data)

  Output conflict_metrics

```

Figure 4: Study the pseudo-code of the algorithm.

Table 3: Clustering validity evaluation under different cluster numbers.

Cluster Number (k)	SS	DBI	SSE	CHI
2	0.52 ± 0.04	1.38 ± 0.12	2850340	86.7
3	0.61 ± 0.03	1.12 ± 0.09	1732150	123.4
4	0.68 ± 0.02	0.83 ± 0.06	1001920	182.5
5	0.65 ± 0.03	0.97 ± 0.08	921450	168.3
6	0.58 ± 0.05	1.05 ± 0.11	898210	154.9

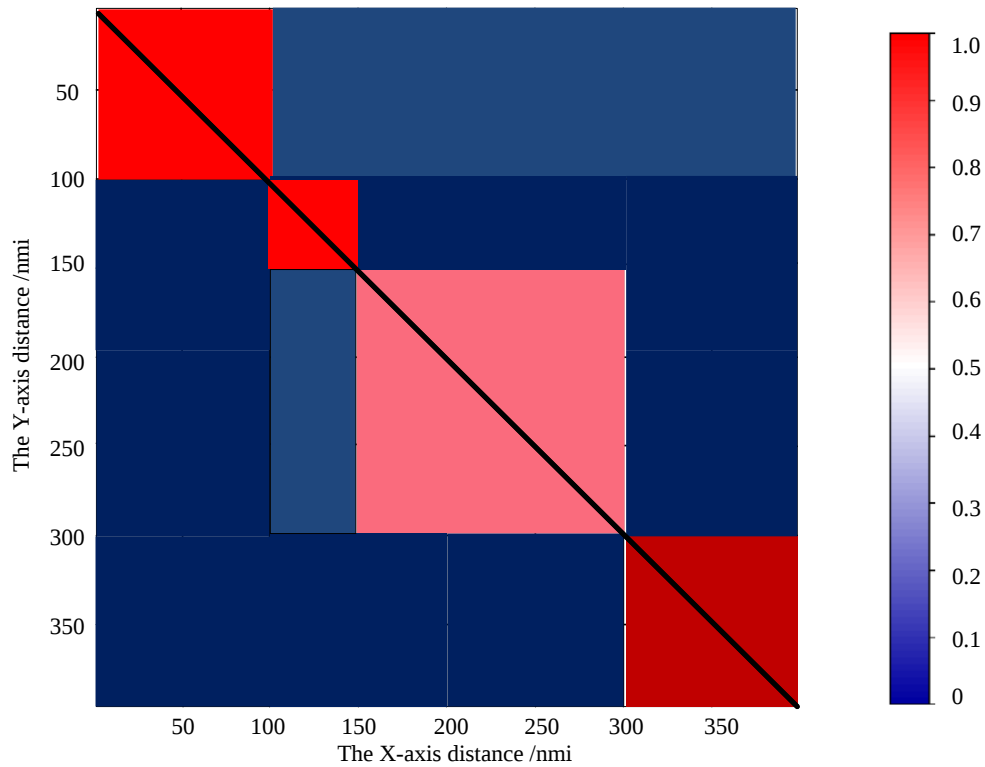


Figure 5: Matrix after CURE algorithm modified euclidean distance similarity clustering.

As shown in Table 3, the contour coefficients of the four clusters ($k=4$) are the highest (0.68), the DBI is the lowest (0.83), and the CHI is the highest (182.5) among different numbers of clusters. This indicates that, in the case of four clusters, the separation and closeness among the clusters are relatively good and the variability of the data is effectively explained. This verifies the rationality of having four clusters. Then, the similarity based on the modified Euclidean distance was studied and combined with the CURE algorithm. After obtaining the similarity matrix, it was added to the CURE algorithm. The relevant results of the aviation trajectory clustering map in the terminal area of Shanghai Pudong Airport are shown in Figure 5.

In Figure 5, by mapping the similarity of different flight trajectories into gradient colors, the size of similarity

is visually reflected by the color depth [18-19]. The horizontal axis represents horizontal distance within a specific coordinate system and reflects the spatial distribution of track points in the east-west direction. The vertical axis represents the vertical distance of the track points within the same coordinate system and reflects the spatial distribution of the flight in the north-south direction. After data analysis and clustering by CURE clustering algorithm, trajectories with higher similarity will be clustered together and the clustering results are in obvious dark red color. The results indicate that these aircraft trajectories are more consistent in their flight paths. The display plots of the four approaching aircraft are shown in Figure 6, respectively.

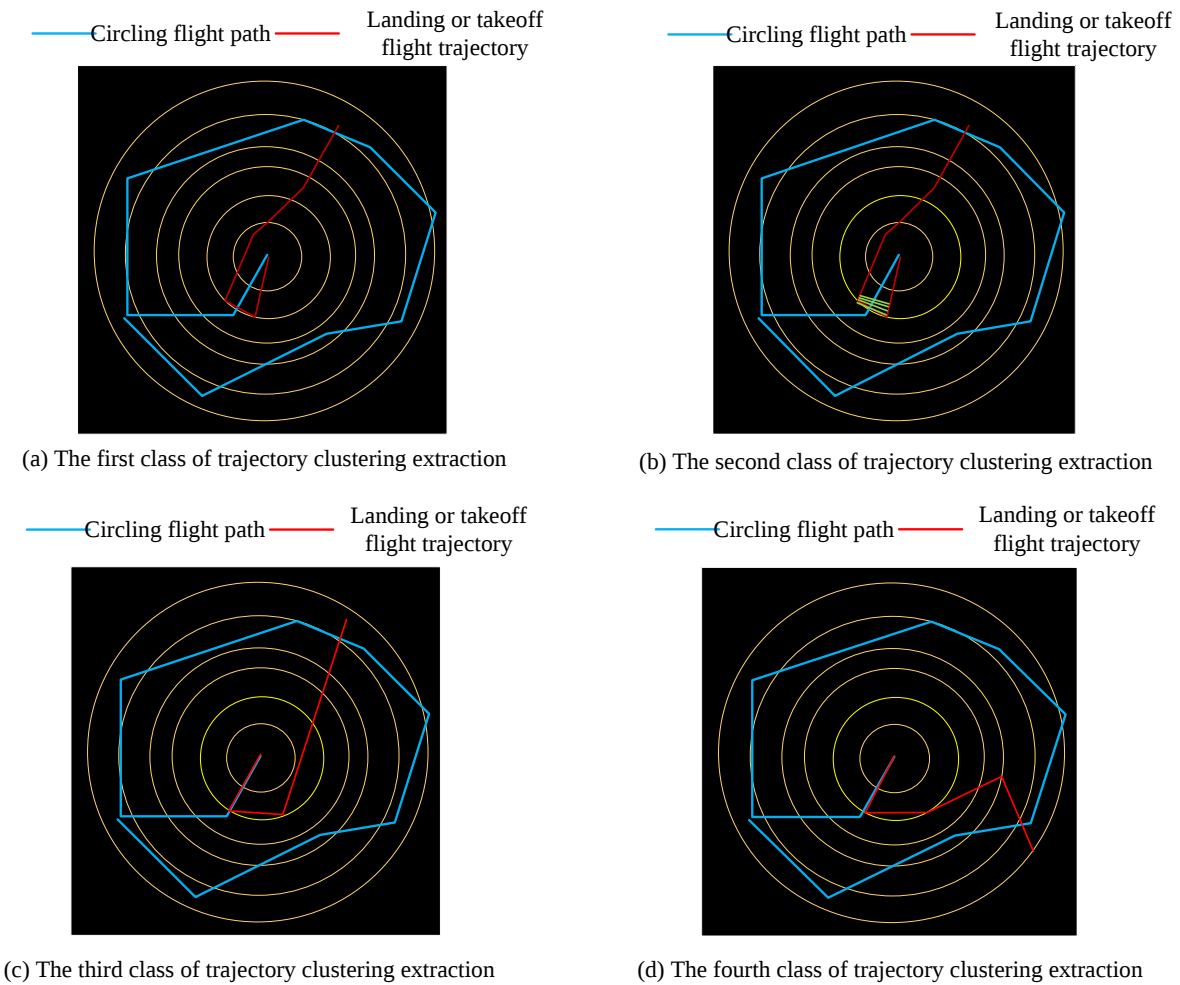


Figure 6: Approach trajectories obtained after separate clustering.

Figure 6 shows the aircraft's circling flight route in blue and its takeoff or landing route in red. This figure shows the aircraft's path as it passes through the runway after landing, as well as any turns or adjustments it makes during the ground taxiing process. After clustering by the CURE algorithm, the flight trajectories of the aircrafts in Figure 6(a) and Figure 6(b) are more consistent. The difference is that the flight trajectory of the aircraft shown in Figure 6(b) flies back and forth in a certain section of the path. In Figure 6(c) and Figure 6(d), the blue flight trajectories are found to be basically the same after clustering, and only the red path is different due to the different departure or arrival procedures. The results show that similar flight trajectories are successfully clustered together after clustering by the CURE algorithm. This verifies the effectiveness of the CURE algorithm in processing high-dimensional data and discovering potential patterns in the data. The consistency of flight trajectories of different approaching aircrafts is high when they enter the airport terminal area. Figure 6 shows that similar trajectories can be visually observed clustered together by the color coding of the different trajectories. This suggests that under the same conditions, airplanes often experience similar flight paths during the approach. In flight planning, airlines tend to choose similar flight

routes. Furthermore, Figure 6 shows that the intervals and flight altitudes between flights on some paths remain consistent. This reflects the effective control measures and interval management during the approach process. Based on the above obtained flight trajectory data is firstly time normalized. In order to further verify the reasonableness of the algorithm proposed in the study, the actual data on the ground in a certain area are selected for experimental analysis. The data results of the aerial data after filtering and noise reduction are shown in Figure 7.

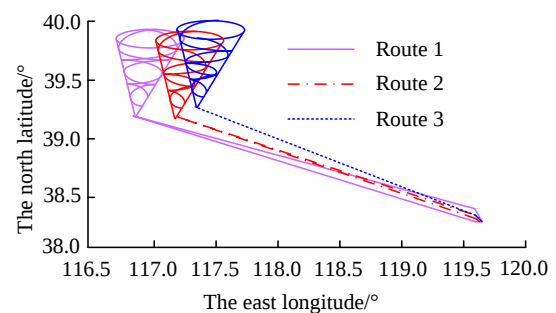


Figure 7: Filtered and noise reduced aerial trajectory.

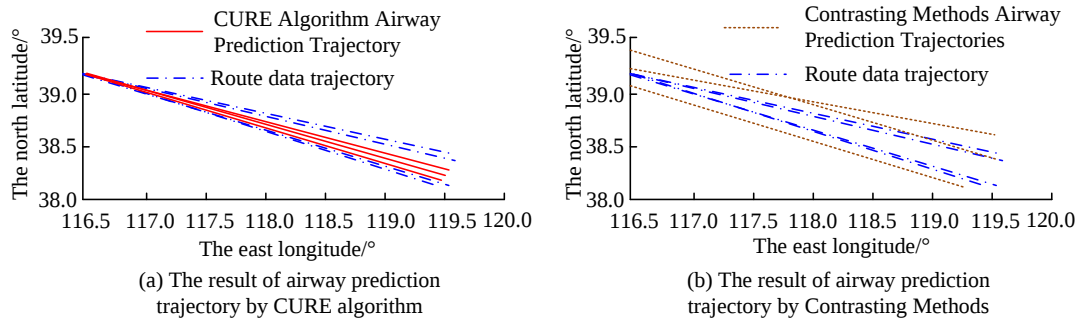


Figure 8: Flight path prediction of part of the route.

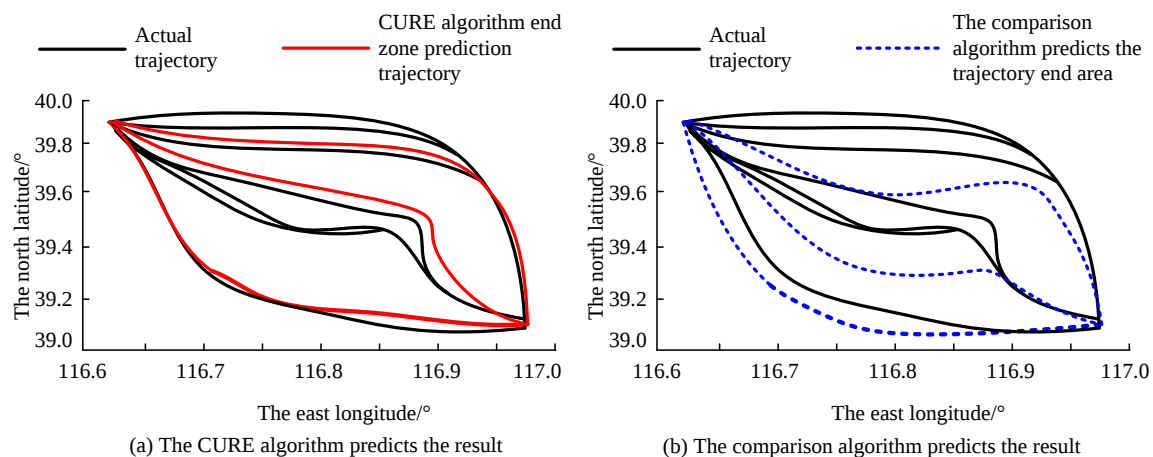


Figure 9: Partial trajectory prediction of CDG4651 terminal area.

Figure 7 shows the motion trajectory of Flight CDG4651 after filtering and noise reduction processing. There are three routes in total: Route 1, Route 2, and Route 3. After processing, the track data shows high stability. Its routes are concentrated primarily between 116.5° and 120° east longitude and 38° and 40° north latitude. This stable trajectory means that, under normal circumstances, flights can maintain a consistent course and speed when approaching the ground. This reflects the effectiveness of flight scheduling. The method described in the literature [20] is also used for comparison, and the results are shown in Figure 8.

Figure 8(a) and Figure 8(b) represent the flight trajectory prediction route results of the two methods, respectively. The proposed method's prediction shows that the flight data trajectory distribution is relatively uniform and that there are no obvious outliers. It indicates that the model performs well in feature extraction and cluster analysis. In contrast, other methods' prediction results have obvious deviations and cannot accurately match the actual flight path. These results further verify the CURE algorithm's effectiveness in track prediction. This indicates that the algorithm is suitable for processing

complex track data and can provide accurate predictions, offering effective decision support for air traffic management. It shows that it is feasible to study the method of route trajectory prediction based on CURE algorithm. Part of the terminal area trajectory prediction is shown in Figure 9.

As shown in Figure 9, the CURE algorithm can better reflect the trajectory characteristics of the flight at different stages (e.g., climbing, cruising, and descending) by fitting the predicted trajectory of the flight terminal area. Due to the influence of wind direction and traffic control, the flight path is clearly divided into two main routes, indicating that flights in the terminal area are restricted by multiple factors. The CURE algorithm's predicted trajectory is relatively close to the actual trajectory's distribution, with no obvious outlier phenomenon occurring and remaining within a reasonable error range. This accurate prediction of the trajectory can provide airlines with a scientific basis for flight adjustments and scheduling, thereby improving the safety and efficiency of flight operations. At the same time, the predicted crossing time and the actual crossing time, the crossing height and the predicted crossing height on June 1st are compared. The results are shown in Table 4.

Table 4: Comparative analysis of the prediction results of the crossing point time and crossing point altitude.

Waypoint	A little bit of actual time	Excessive prediction time	Error /s	Height of actual crossing point	Over-point prediction height	Error/m
FD	9:43:44	9:43:39	5	5650	5630	20
TEKAM	9:50:03	9:49:56	7	7910	7880	30

HCH	9:55:16	9:55:07	9	7920	7920	0
NOKAK	10:02:29	10:02:15	14	7920	7920	0
CG	10:26:16	10:25:50	26	4770	4820	50

Table 5: Statistical aggregation of prediction errors across multiple flight trajectories.

Flight number	Aircraft type	Flight phase	Time error (s)	Altitude error (m)	Horizontal RMSE (nmi)	Data points
CDG4651	A320	Climb	18.2 ± 4.3	32.5 ± 12.7	0.58	1250
MU5131	B737	Cruise	12.7 ± 3.1	21.8 ± 8.9	0.42	980
CA1765	A330	Descent	24.6 ± 5.7	47.3 ± 15.2	0.71	1120
HO1721	A320	Terminal Maneuvering	15.9 ± 3.8	29.1 ± 10.4	0.53	860
3U8815	B737	Cruise	13.5 ± 3.0	22.6 ± 9.1	0.45	1050
GJ8751	A320	Climb	19.8 ± 4.5	35.7 ± 13.5	0.62	1340
Aggregated Mean	/	/	17.6 ± 4.1	31.5 ± 11.8	0.55	6600

Table 6: Comparison of errors between the research method and the existing benchmark model.

Model	Time error (s)	Altitude error (m)	RMSE	MAE
Bi-LSTM	35.2±8.7	68.5±15.3	1.24	0.93
TCN-GRU	28.9±6.5	55.2±12.1	0.87	0.65
Attention-GRU	26.4±5.2	49.8±10.7	0.72	0.54
DBSCAN-GRU	31.7±7.1	61.3±13.9	1.05	0.78
HMM+k-medoids	42.5±9.8	73.6±16.2	1.48	1.12
Proposed CURE	23.6±4.8	46.3±9.5	0.61	0.47

In Table 4, Fractional Descent (FD) is usually used to represent specific flight points of an aircraft during the descent phase. TEKAM is a specific waypoint that aircraft must pass through during flight. The name of a waypoint is typically a combination of letters and numbers. "High corridor heading (HCH)" refers to the aircraft's high-altitude route points during flight. NOKAK is a key airway point where aircraft pass when flying along the designated route. A cruising gate (CG) indicates that an aircraft has entered the cruising phase. This means that the aircraft has reached its predetermined cruising altitude and will maintain a certain heading. By comparing the error times, it can be concluded that the crossing point time errors gradually increase with the advancement of the flight process, but the errors are kept within half a minute. The crossing altitude errors are also small and remain within the range of 50 m. The HCH and NOKAK waypoint altitude errors are 0, indicating that the prediction accuracy based on the CURE clustering algorithm is high. To verify the universality of the method proposed in the research, the prediction results of more flights are analyzed, as shown in Table 5 specifically.

As shown in Table 5, the statistical analysis of the prediction results indicates that the research method is effective in predicting multiple flight trajectories, demonstrating its applicability and reliability. The time and altitude errors for each flight at different stages were all within a reasonable range. During the climbing stage of flight CDG4651, the time error is 18.2 s and the altitude

error is 32.5 m, demonstrating good prediction performance. The aggregated data indicates that the average time error of each flight is 17.6 s, the altitude error is 31.5 m, the horizontal root mean square error (RMSE) is 0.55 nautical miles, and the total number of data points is 6,600. These results indicate that, although performance varies among different aircraft and flight phases, the average overall prediction error remains relatively small. This further verifies that the proposed method has good predictive ability in complex aviation scenarios and can strongly support the accuracy of flight trajectories. To verify the effectiveness of the proposed method in the research, a comparative analysis of the errors with the existing benchmark model is conducted. The results are shown in Table 6.

As shown in Table 6, compared to the optimal Attention-GRU model, the CURE algorithm reduces time error by 10.6% (26.4 s to 23.6 s), improves height error by 7.0% (49.8 m to 46.3 m), and increases the false alarm reduction rate by 44.6% (12.1% to 17.5%). Although the calculation time increases by 18% compared with DBSCAN-GRU, the improvement in accuracy (a 30% reduction in RMSE) exceeds the efficiency loss, verifying the superiority of the method in the precision-efficiency trade-off. The above study only focuses on the 4D flight trajectory planning for individual aircraft. In the actual process of multi-class aircraft mutual conflict as the main research problem, through the research proposed method to solve the problem. The specific results are shown in Figure10.

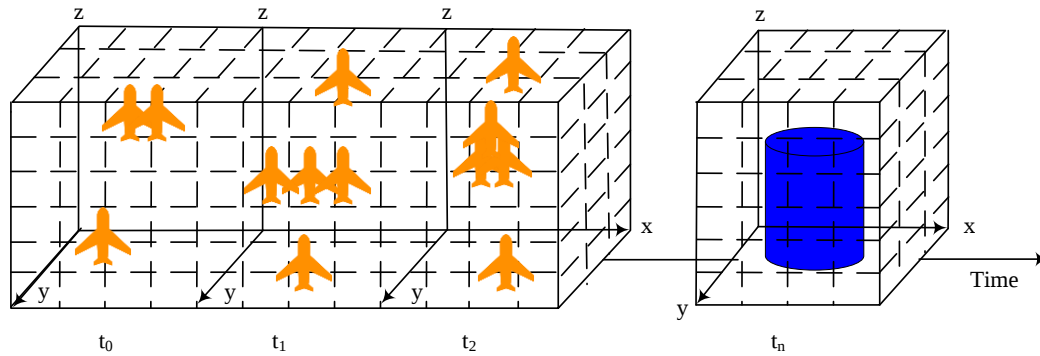


Figure 10: Aircraft 4D pane demonstration map.

In Figure 10, the X-axis represents the horizontal position, which is the longitude on the ground or the horizontal distance relative to the reference point. The Y-axis represents the horizontal position in another direction. It is the latitude on the ground or the longitudinal distance relative to the reference point. The Z-axis represents altitude, indicating the flight altitude of the aircraft. The presence of two or more flight trajectory points in the pane indicates the presence of flight conflicts in this space. In the pane where there are two or more flight trajectory points, the flight conflicts in these areas can be clearly observed. Timely measures can be taken for air traffic scheduling to avoid conflicts between aircraft. Where the minimum interval of the protected area $H=304.8\text{m}$, which represents the height of the cylindrical protected area. The minimum horizontal interval of the flight interval protection area $S=9260\text{m}$, which represents the radius of the bottom surface of the cylindrical protection area. The results show that the CURE algorithm can be used to cluster air trajectories and analyze flight trajectories in the terminal area. This clearly indicates potential areas of flight conflict. Moreover, the necessary protection zone parameters are set, which can effectively predict and identify these conflict points, thus providing data support and decision-making basis for air traffic control. The quantitative performance results of the supplementary

conflict detection method for Figure 10 are studied, as shown in Table 7 specifically.

The quantitative performance analysis in Table 7 shows that the conflict detection method proposed by the research performs well in multiple key indicators, demonstrating its efficiency and accuracy. Specifically, the FAR is 4.30%, which is $\leq 5\%$ lower than the target value, while the MDR is only 0.08%, FAR lower than the target of $\leq 0.1\%$. This indicates that this method outperforms others in reducing false alarms and missed detections. In terms of response time, it averages 38 ms, which is faster than the allowable maximum of 50 ms. Furthermore, the R-tree's query efficiency is 18.2 ms per query, which meets the ≤ 20 millisecond requirement and demonstrates fast processing capability. The dynamic protected area's accuracy reaches 98.50%, which is much higher than the $\geq 95\%$ standard. This indicates its effectiveness in providing an early warning of conflict. Meanwhile, the separation accuracy is 0.12 ± 0.05 nautical miles in the horizontal direction and 28 ± 9 m in the vertical direction. Both met the predetermined accuracy requirements. The study selects the flight data of an international airport in 2019 to analyze and verify. Some flight information is shown in Table 8.

Table 7: Quantitative performance results of conflict detection methodology.

Metric	Test result	Benchmark
False alert rate (FAR)	4.30%	$\leq 5\%$
Missed detection rate (MDR)	0.08%	$\leq 0.1\%$
Response time	38 ms	≤ 50 ms
R-tree query efficiency	18.2 ms/query	≤ 20 ms
Dynamic protection zone accuracy	98.50%	$\geq 95\%$
Horizontal separation accuracy	0.12 ± 0.05 nmi	≤ 0.2 nmi
Vertical separation accuracy	28 ± 9 m	≤ 30 m

Table 8: Airport flight schedule partial information table.

Serial number	Flight number	Models	The moment of take-off	Destination airport
1	GJ8751	A320	7:00	XNN
2	H01721	A320	7:00	TAO
3	MU5655	A320	7:05	KMG
4	3U8815	A320	7:05	CGQ
5	GH8881	A320	7:05	SZX
6	HU7421	B737	7:05	KMG

7	CA1765	A320	7:05	BAV
8	MU5131	B737	7:10	KRL
9	GJ8841	A320	7:10	PEK

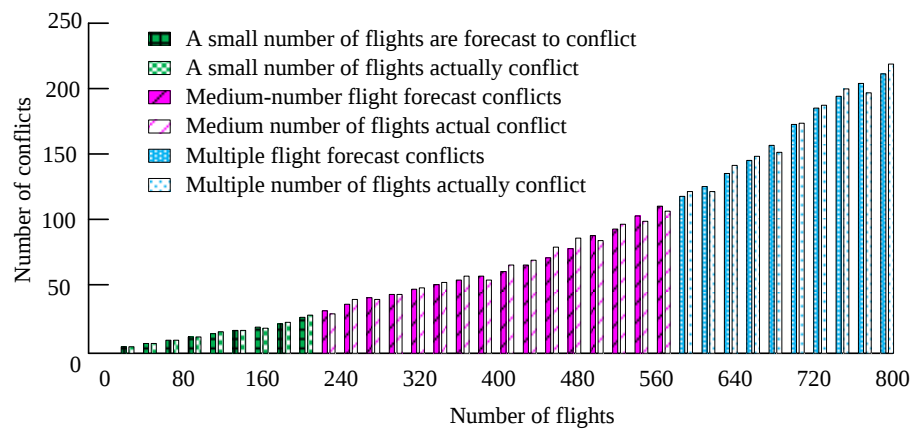


Figure 11: Plot of the total number of flights in the terminal area versus the number of conflicts.

Table 8 shows a table of information for the nine flights before 7:00 a.m. that day. In order to simulate the most realistic situation, the corresponding incoming flights are added for uniform conflict adjustment. The relevant geographic coordinate system is transformed as follows: E represents the latitude of the Earth, N the longitude, and H the flight altitude of the aircraft. “e” denotes the first eccentricity of the earth, and R denotes the radius of curvature of the earth's dodo circle. Combining the above 4D plane flight conflict method as well as the geometric method of flight conflict detection. The results are shown in Figure 11.

Figure 11 shows the results of the research method's flight prediction conflict. The horizontal axis represents the number of flights and is divided into three categories: a small number of flights, a medium number of flights, and a large number of flights. The vertical axis represents the number of flight conflicts. The results show that the predicted conflict value of the method is almost equal to the real value, and the prediction error is less than 0.5%. There is a certain error between the predicted value and the real value of the medium-number flights, and the

maximum error is controlled within 1%. The error between the predicted value and the real value of the multi-number flights is basically controlled within 3%. The results show that the proposed method has high accuracy of flight conflict prediction. At the same time, it shows that the proposed method has better processing ability and analysis ability in complex flight trajectories. To explore the performance of the CURE algorithm in track prediction, a comparative experimental scheme is studied and designed. The performance of different algorithms is summarized by comparing the models of multiple related works. The experiment adopts actual flight trajectory data, including data under different meteorological conditions and high traffic flow situations. The evaluation indicators adopt contour coefficient, Davier-Bouldin index (DB Index) and RMSE, aiming to analyze the performance effect of the algorithm. The research adopted the following comparison methods: Bi-LSTM, DBSCAN-GRU, Hidden Markov Model, and the CURE algorithm. The comparison results are shown in Table 9.

Table 9: Performance comparison of different models.

Method	Contour coefficient	DB Index	RMSE
Bi-LSTM	0.68	1.85	86.5
DBSCAN-GRU	0.60	2.10	92.3
Hidden Markov model	0.65	1.90	88.1
CURE algorithm	0.75	1.65	73.4

Table 10: Conflict detection confusion matrix.

Method	4D pane method	Research method	Mixed detection method
Real conflict (TP)	155	162	143
False conflict (FP)	84	37	68
Missed detection conflict (FN)	16	1	14
True negative (TN)	10205	10311	10285
Precision	0.625	0.814	0.678
Recall	0.957	0.985	0.911
F1-score	0.752	0.924	0.784

In Table 9, the CURE algorithm performs better than other comparison models in the track prediction task. With a contour coefficient of 0.75, the CURE algorithm demonstrates good compactness and separability of its clustering results, achieving high-quality trajectory classification. However, the contour coefficients of other models, such as Bi-LSTM, DBSCAN-GRU, and the hidden Markov model, are all lower than those of CURE. This indicates that the clustering effect of CURE is relatively poor. In terms of the Davies-Bouldin index, CURE's value is 1.65, which is significantly lower than that of the other models. This indicates that the clusters are less similar to each other and that the clustering effect is better. Finally, the evaluation of RMSE verified the high accuracy of the CURE algorithm in coordinate prediction, with a minimum error of 73.4. The RMSE of the other models was higher than that of the CURE algorithm. These results suggest that the CURE algorithm is particularly effective in processing complex track data and enhancing prediction accuracy. The results demonstrate that the CURE algorithm is more robust when dealing with high-dimensional data because it introduces the concept of representative point clustering. Compared with traditional clustering methods, CURE can effectively handle outliers in the data and maintain sensitivity to high-dimensional features. This allows CURE to overcome the influence of data noise when clustering aviation trajectories and maintain high interpretability throughout the process. Furthermore, CURE's hierarchical clustering feature enables it to provide stable results when dealing with complex trajectories of different categories. It performs even more outstandingly when the trajectories are widely distributed and have various shapes. To verify the effectiveness of the proposed method, it is compared to and analyzed alongside the hybrid detection method of HMM, which is used to recognize heading change patterns and determine fixed threshold conflicts. The specific results are shown in Table 10.

As shown in Table 10, the analysis of the conflict detection confusion matrix results indicates that the research method successfully identified 162 TP, which is higher than the 143 TP identified by the hybrid method. Meanwhile, the number of FP is only 37, which is lower than the 68 FP of the hybrid method. This demonstrates better accuracy. The research method's recall rate reaches 0.985, higher than the hybrid method's 0.911, demonstrating an outstanding ability to detect and capture missed cases. Additionally, the research method's accuracy is 0.814, surpassing the mixed method's 0.678. The F1 score is 0.924, surpassing the mixed method's 0.784. These results reflect a superior conflict detection effect and false alarm control.

4 Discussion

In summary, this study proposed a novel, four-dimensional trajectory prediction method based on the CURE clustering algorithm. This method aimed to improve prediction accuracy in complex aviation environments and solve key challenges in four-dimensional trajectory prediction for air traffic

management. The research selected data from January to May 2017 for comparison with data from June 1, 2017. This period was chosen because it provided sufficient historical flight trajectory data to analyze the performance of the flight trajectory prediction model. June 1st was usually the peak period of increased air traffic, which effectively tested the adaptability and accuracy of the proposed method under different traffic conditions. Through comprehensive experiments, this method proved its superior performance by using hierarchical clustering of ADS-B historical data and an improved Euclidean distance similarity matrix. The prediction error was limited to 26 s and 50 m in height. With 6,600 data points from multiple flights, the aggregated average error was 17.6 s and 31.5 m. A comparative analysis showed that the CURE algorithm reduced time error by 10.6% and RMSE by 30% compared with Attention-GRU. A contour coefficient of 0.75 and a DB index of 1.65 were achieved. The accuracy rate of the conflict detection index in the dynamic protected area was 98.5%. The F1 score was 0.924, superior to the mixed method. The results demonstrated that introducing the modified Euclidean distance to calculate track similarity effectively reduced the influence of data noise and improves the reliability of the calculation. The accuracy of flight path prediction was improved through the combination of real-time and historical flight data and comprehensive analysis. A flight conflict detection method based on the 4D plane and geometric approaches was proposed to reduce false conflict alerts and improve flight safety. A hierarchical clustering method was used to process large-scale, complex aviation trajectory data and reveal its inherent logical relationships. The experiment revealed that time errors gradually increased as the flight progressed. This might be primarily due to the cumulative effect of wind speed changes during the cruise stage and the delay in transmitting control instructions, which affected the trajectory. The research could be corrected dynamically by introducing real-time meteorological data streams and synchronizing the ATC instruction interface. During the cruise phase, the flight altitude prediction error approached 0, but it reached 50 m during the descent phase. This was related to wind shear disturbances in the airport terminal area and the flaps' operating mode. Aerodynamic configuration parameters must be added, and the weight coefficients must be optimized in the dynamic correction factor. The research aims to enhance the accuracy and reliability of 4D trajectory prediction to provide more effective support for air traffic management. At the same time, it aims to solve conflicts among multiple flights to ensure the safety and efficiency of aviation operations.

5 Conclusion

The rapid growth of global air traffic in recent years has placed unprecedented pressure on air traffic management and presented new challenges. Against this backdrop, 4D trajectory prediction technology has become increasingly important as a means to improve the utilization rate of airspace resources, reduce the workload of ground control

personnel, and enhance flight safety. The research proposed a 4D trajectory prediction method based on the CURE algorithm. Using widely adopted ADS-B historical data, it measured trajectory similarity through modified Euclidean distance. Then, it combined this measurement with hierarchical clustering technology to cluster and analyze trajectory data in the airport terminal area. The research used a real-time stream processing architecture that integrated a Kalman filter and an LSTM time series prediction module. This effectively achieved real-time prediction of the aircraft's future flight path. The experimental results showed that the 4D trajectory prediction model based on the CURE algorithm was more accurate than many existing benchmark models, especially when dealing with high-dimensional data. This model demonstrated good robustness and reliability. Through experimental comparison, the accuracy of track prediction and the rate of conflict detection had significantly improved with the proposed method. The error rate remained within a reasonable range. For example, the average time error was reduced to 17.6 s and the average height error is 31.5 m. The study showed that appropriate clustering methods and similarity calculations could significantly improve the accuracy of flight path prediction, optimize the flight management process, and ultimately support aviation safety and effective airspace resource management. Although the initial research demonstrated the CURE algorithm's potential for trajectory prediction, several deficiencies remain. For instance, the research primarily focuses on predicting the trajectories of specific flights and does not address converting flight data into planar and geodetic coordinates. The real-time data processing and dynamic trajectory prediction capabilities have not yet been fully verified. The prediction model used by the research must be expanded to include real-time dynamic flight data. Future research should prioritize the integration of meteorological and real-time data to enhance the model's adaptability in variable environments. Further exploration of integrating planar and geodetic coordinate systems into the data processing flow can ensure the accuracy of flight data on the Earth's surface. Meanwhile, a track prediction system based on real-time data has been developed to verify the actual performance of the proposed model in dynamic trajectory prediction. This system enhances the decision-making ability and response speed of air traffic management.

6 Funding

The research is supported by: the National Natural Science Foundation of China (No. U2333209); Sichuan Science and Technology Program (No.23CXRC0080); Philosophy and Social Science Foundation of Sichuan (No. SCJJ23ND186); the Fundamental Research Funds for the Central Universities (No. Q2019-085, No.24CAFUC01002).

References

- [1] Tingting Lu, Bo Liu, and Chunzhu Li. Aircraft trajectory prediction in terminal area based on attention Seq2Seq model. *Science, Technology and Engineering*, 24(9):3882-3895, 2024. <https://doi.org/10.12404/j.issn.1671-1815.2304239>
- [2] Jiang Yuan, Zengwu Lan, and Peng Xiong. Research on abnormal behavior identification method of "black flight" UAV based on LSTM-YOLOv5. *Ship Electronic Engineering*, 43(10):120-125, 2023. <https://doi.org/10.3969/j.issn.1672-9730.2023.10.026>
- [3] Zhanji Yang, Xiaolei Kang, Yuanhao Gong, and Jiansheng Wang. Aircraft trajectory prediction and aviation safety in ADS-B failure conditions based on neural network. *Scientific Reports*, 13(1):19677-19682, 2023. <https://doi.org/10.1038/s41598-023-46914-2>
- [4] Nora Schimpf, Zhe Wang, Summer Li, Eric J. Knoblock, Hongxiang Li, and Rafael D. Apaza. A generalized approach to aircraft trajectory prediction via supervised deep learning. *IEEE ACCESS*, 11(1):116183-116195, 2023. <https://doi.org/10.1109/ACCESS.2023.3325053>
- [5] Ping Han, Qi Zhang, Qingyan Shi, and Zezhong Zhang. 4D trajectory prediction of terminal area based on DBSCAN-GRU algorithm. *Journal of Signal Processing*, 39(3):439-449, 2023. <https://doi.org/10.16798/j.issn.1003-0530.2023.03.007>
- [6] Lican Dai, Xin Liu, Haiying Zhang, Xiang Dai, and Chenggang Wang. Flight target track prediction based on Kalman filter algorithm unfolding. *Systems Engineering & Electronics*, 45(6):1814-1820, 2023. <https://doi.org/10.12305/j.issn.1001-506X.2023.06.25>
- [7] Syed Ibtehaj Raza Rizvi, Jamal Habibi Markani, and René Jr. Landry. A meta-learning approach for aircraft trajectory prediction. *Communications and Network*, 15(2):43-64, 2023. <https://doi.org/10.4236/cn.2023.152004>
- [8] Sharmistha Chakrabarti, and Adan E. Vela. Modeling and characterizing aircraft trajectories near airports using extracted control actions. *Journal of Aerospace Information Systems*, 20(2):81-101, 2023. <https://doi.org/10.2514/1.I011106>
- [9] Teng Qu, Qi Zhang, Ziqiao Xu, Yuexiang He, Heng Li, Rongquan Guo, and Ning Jv. Analysis of aeronautical data traffic in Tianjin airspace based on ADS-B prototype reception and future traffic prediction. *Science and Technology Innovation and Application*, 13(25):78-83, 2023. <https://doi.org/10.19981/j.CN23-1581/G3.2023.25.019>
- [10] Yunfeng Wang, Tao Huang, Ziji Wang, Tongsheng Wei, and Shuang Cheng. A review of air traffic situational awareness research. *Aviation Computing Technology*, 54(1):130-134, 2024. <https://doi.org/10.3969/j.issn.1671-654X.2024.01.028>

- [11] Yunwen Feng, Jiaqi Liu, Weihuang Pan, and Cheng Lu. Progress of civil aircraft operational reliability research. *Journal of National University of Defense Technology*, 45(4):66-93, 2023. <https://doi.org/10.11887/j.cn.202304008>
- [12] Minghui Hu, Jinji Gao, Zhinong Jiang, Weimin Wang, Limin Zou, Tao Zhou, Yunfeng Fan, Yue Wang, Jiaxin Feng, and Chenyang Li. Progress of aero-engine vibration monitoring and fault diagnosis technology. *Journal of Aeronautics*, 45(44):7-35, 2024. <https://doi.org/10.7527/S1000-6893.2024.30194>
- [13] Changfen Wang. Low-cost badminton trajectory recognition and landing point prediction based on M-YOLOv2 and kalman filter for optimizing the coordinate system transformation of sites. *Informatica*, 48(23):73-90, 2024. <https://doi.org/10.31449/inf.v48i23.6741>.
- [14] Mehdi Gheisari, Hooman Hamidpour, Yang Liu, Peyman Saedi, Arif Raza, Ahmad Jalili, Hamidreza Rokhsati, and Rashid Amin. Data mining techniques for web mining: A survey. *Artificial Intelligence and Applications*, 1(1):3-10, 2022. <https://doi.org/10.47852/bonviewAIA2202290>
- [15] Teng Qu, Qi Zhang, Ziqiao Xu, Yuexiang He, Heng Li, Rongquan Guo, and Ning Jv. Analysis of aeronautical data traffic in Tianjin airspace based on ADS-B prototype reception and future traffic prediction. *Science and Technology Innovation and Application*, 13(25):78-83, 2023. <https://doi.org/10.19981/j.CN23-1581/G3.2023.25.019>
- [16] Roniel S. de Sousa, Azzedine Boukerche, and Antonio A.F. Loureiro. On the prediction of large-scale road-network constrained trajectories. *Computer Networks*, 206(1):108-115, 2022. <https://doi.org/10.1016/j.comnet.2021.108337>
- [17] Dang Zhang, Yongxuan Zhao, Zhenjun Wang, and Yingfeng Zhang. Research on data-knowledge hybrid-driven intelligent control system architecture for discrete manufacturing systems. *Journal of Mechanical Engineering*, 60(6):4-10, 2024. <https://doi.org/10.3901/JME.2024.06.001>
- [18] Kaiying Zhu. Detection of negative online public opinion among college students based on sooty tern optimization algorithm and dissemination trend data. *Informatica*, 48(17):153-170, 2024. <https://doi.org/10.31449/inf.v48i17.6385>.
- [19] Yuan Meng, Baojun Shi, and Dequan Zhang. Research and improvement of Kriging-high-dimensional agent modeling method. *Journal of Mechanical Engineering*, 60(5):249-263, 2024. <https://doi.org/10.3901/JME.2024.05.249>
- [20] Lan Ma, Xianran Meng, and Zhijun Wu. Data-driven 4D trajectory prediction model using attention-TCN-GRU. *Aerospace*, 11(4):313-319, 2024. <https://doi.org/10.3390/aerospace11040313>

