Trajectory clustering for arrival aircraft via new trajectory representation

GUI Xuhao, ZHANG Junfeng[*], and PENG Zihan

College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Abstract: Trajectory clustering can identify the flight patterns of the air traffic, which in turn contributes to the airspace planning, air traffic flow management, and flight time estimation. This paper presents a semantic-based trajectory clustering method for arrival aircraft via new proposed trajectory representation. The proposed method consists of four significant steps: representing the trajectories, grouping the trajectories based on the new representation, measuring the similarities between different trajectories through dynamic time warping (DTW) in each group, and clustering the trajectories based on k-means and density-based spatial clustering of applications with noise (DBSCAN). We take the inbound trajectories toward Shanghai Pudong International Airport (ZSPD) to carry out the case studies. The corresponding results indicate that the proposed method could not only distinguish the particular flight patterns, but also improve the performance of flight time estimation.

Keywords: air traffic management, trajectory clustering, trajectory representation, flight pattern.

1. Introduction

Recently, China has witnessed the rapid development of the civil aviation industry. Meanwhile, the rapid growth of air traffic and the limitation of available airspace have led to aviation congestion problems. Therefore, how to modernize and harmonize the air traffic management (ATM) systems is becoming a promising way to tackle the issues of severe flight delays, excessive fuel consumption, and consequent air pollutant emission. The innovative technologies [1] and novel operational concepts [2] are promising ways to carry out the modernization and harmonization of the ATM systems. However, to get a thorough understanding of the current ATM operations

is a prerequisite to address such issues. The large amounts of historical data in the ATM domain provide the possibilities for building and improving our knowledge about the current ATM operations.

The essential type of historical data is aircraft trajectories which are available and accessible through the surveillance facilities or technologies [3]. Since the historical trajectories contain the temporal and spatial information of each aircraft, most of the researchers rely on trajectory clustering to get access to a comprehensive understanding of the real operation [4−7]. Some work paid attention to the en-route phase, in which the researchers tried to model the air traffic flow for enhancing the abilities of en-route traffic flow management [8], strategic planning [9, 10], and dynamic airspace sectorization [11]. Others focused on the terminal area (TMA) and the corresponding studies aimed at identifying prevail flight tracks or common flight patterns. These efforts could help to redesign the procedures or airspace [5], characterize and evaluate the performance in TMA [4] or multi-airport systems [12−14], and improve the precision of short-term estimated time of arrival (ETA) prediction [15,16].

Trajectory clustering is a process of grouping similar trajectories into different clusters to find representative paths or common trends shared by different moving objects [17]. The fundamental components of trajectory clustering are trajectory similarity/distance measurements and trajectory clustering algorithms [18]. Rehm [4] measured the similarity between different trajectories by using Euclidean distance, which was believed as popular similarity measurement. Such measurement required that the number of trajectory points should be equal; thereupon, a fixed sampling rate should be implemented at the beginning, which would lead to the loss of information. Bombelli et al. [9,10] implied Frechet distance to carry out the similarity measurement, which did not need to resample the original trajectories. However, Frechet distance only returned the maximum distance between each pair of tra-

jectories and constructed an enormous similarity matrix ($m \times m$, where m denotes the number of trajectories) for further clustering. Some studies relied on warping based distance to measure similarity with different lengths and uneven sampling rates. Gariel et al. [5] firstly distinguished the turning points of each trajectory, then represented the trajectory with those points, and finally clustered the trajectories by using the longest common subsequence (LCSS). Morris and Trivedi [19] also used LCSS to measure the trajectory similarities and adopted a hierarchical strategy to cluster trajectories. Besides LCSS, Hong and Lee [15] applied dynamic time warping (DTW) to compute the distance and choose the hierarchical clustering method to group the trajectories. Zhao and Shi [20] used DTW to measure the similarity of the compressed trajectories, which were obtained by the Douglas-Peucker compression algorithm, and then clustered the trajectories by the density-based spatial clustering of applications with noise (DBSCAN) algorithm. In spite of measuring the original trajectories, the other studies focused on constructing features from the trajectories and implemented trajectory clustering based on the constructed features, which could make the trajectory clustering more intuitive. Gariel et al. [5], Marzuoli et al. [8], Salaun et al. [21], and Wang et al. [16] constructed new features for the trajectories by the following steps: resampling the trajectories, augmenting the dimensionality, extracting the features based on principal components analysis (PCA), and clustering based on density-based or hierarchical strategies.

To summarize, most of the studies deal with the trajectory clustering problem either from the holistic perspective or based on the partition-and-group framework [22,23]. For the former, each trajectory can be treated as a series of points [4,9,10] or a sample with augmented features [5,8,16,21]. For the latter, each trajectory is represented by several segments [15,19,20]. It is obvious that, no matter what kind of framework is adopted, the similarity measurements and clustering algorithms are the same issues that should be addressed. However, the distinctive characteristic is how to represent the trajectories, like a series of points, the augmented features, or the partitioned segments. The different ways of representing the trajectories tend to need the appropriate similarity measurements and the corresponding clustering algorithms. Therefore, the trajectory presentation plays a significant role in the trajectory clustering, which in turn calls for the problem-oriented background knowledge.

This paper aims at presenting a semantic trajectory clustering approach for arrival aircraft. We focus on this aspect since arrival aircraft is always instructed to deviate from the published routes to ensure safe separation and establish a landing sequence. The proposed semantic trajectory clustering approach is totally based on a new type of trajectory representation, which is closely associated with the real operation in the radar vectoring situation. Such new trajectory representation reflects the clearance delivered by the controllers, which in turn could help us to easily identify the holding pattern, recognize landing direction, determine the landing modes, and distinguish the entry fixes. In return, the distinguished controller intents can bring an advantage to the arrival flight time estimation. By using the DTW measurement and k-means or DBSCAN algorithms, the semantic trajectory clustering approach can fulfill the task of grouping the arrival trajectories. Finally, from the grouped trajectories, we can identify the controller intention and improve the precision of arrival flight time estimation.

Section 2 presents the new trajectory representation method and the corresponding essential characteristics. Trajectory similarity measurement and clustering algorithms are described in Section 3. The results and discussion of real case studies are summarized in Section 4. The concluding remarks are provided in Section 5.

2. Trajectory representation of arrivals

2.1 Previous trajectory representation

Aircraft trajectories, which could be decoded from surveillance radar data or automatic dependent surveillance-broadcast (ADS-B), are often represented by point sequences [1], including the recorded time, aircraft position, speed, heading, and data related to the flight plan. For m aircraft trajectories, we define them as a set $\boldsymbol{TR}$, where $\boldsymbol{TR} = \{\boldsymbol{tr}_1; \boldsymbol{tr}_2; \cdots; \boldsymbol{tr}_i; \cdots; \boldsymbol{tr}_m\}$. Each trajectory $\boldsymbol{tr}_i$ ($i \in \{1, 2, \cdots, m\}$) consists of a series of points: $\boldsymbol{tr}_i = (\boldsymbol{p}_1^i, \boldsymbol{p}_2^i, \cdots, \boldsymbol{p}_{L(i)}^i)$, where $L(i)$ represents the number of points of the ith trajectory. Each point $\boldsymbol{p}_j^i$ is a tuple: $\boldsymbol{p}_j^i = \boldsymbol{t}, \boldsymbol{l}, \{\boldsymbol{a}\}$, which describes the aircraft position ($\boldsymbol{l}$, location) and its corresponding attributes ($\boldsymbol{a}$, attributes) at a specific time t. The position information consists of longitude, latitude, and altitude of such aircraft at the specific time t, while the attributes may include the heading, speed, distance-to-go (DTG) (the remaining distance), and the like.

The original aircraft trajectories are made up of too many points, which will cause computational burden issues during trajectory clustering. We can implement the data compression method, like the Douglas-Peucker algorithm, to simplify the trajectories. After removing those points from a nearly straight line, each trajectory $\boldsymbol{tr}_i$ ($i \in \{1, 2, \cdots, m\}$) could be represented by much fewer points: $\boldsymbol{tr}_i = (\boldsymbol{p}_1^i, \boldsymbol{p}_2^i, \cdots, \boldsymbol{p}_{L'(i)}^i)$, where $L'(i)$ is less than $L(i)$ and affected by the threshold parameter of the Douglas-Peucker algorithm.

Besides representing the aircraft trajectory by actual points, we can also extract features from any given trajectory. For example, we can focus on the attributes of any arrival aircraft when passing a particular fix. Besides, we can utilize the statistical values of the trajectory's attributes, like averages and/or standard deviations of longitude, latitude, altitude, speed, and heading for arrival aircraft. Afterward, the original trajectory is transformed into distinct feature space, which represents complementary characteristics for the trajectory. Furthermore, feature selection and extraction process can be applied for reducing the irrelevant and redundant features.

2.2 The proposed novel trajectory representation

In the civil aviation domain, the aircraft trajectory possesses some unique characteristics. Firstly, during the enroute phase, each aircraft should follow the predesigned air route. Secondly, during the operation within TMA, many departure or arrival aircraft will fly along with the published standard departure or arrival procedures. Thirdly, during the operation within the TMA in rush hour, each aircraft should follow the instructions given by air traffic controllers (ATCOs), while "change heading" is a common instruction. Therefore, not only in the predesigned air routes, published standard instrument departures (SIDs) or standard terminal arrival routes (STARs) but also from the instructions by ATCOs, the headings of aircraft play a significant role.

The Routes, SIDs, and STARs, defined by Navaids or Waypoints, could be represented by distance and course. During the flight, the pilots try to manipulate the aircraft to ensure the track is following the course by changing heading. As a result, the heading is one of the most informative attributes. Meanwhile, during the TMA operation in rush hour, ATCOs tend to prefer heading change instructions for sequence establishment and maintenance, which could help them to accurately monitor the aircraft movements and accordingly improve their situation awareness. Given this, the heading attribute should be paid more attention. More specifically, most of the arrival aircraft finally keep the same direction, i.e., nearly the same heading, due to landing on a single runway or paralleled multiple runways. Moreover, the heading could also be described as the fluctuations with the different DTGs, which measures the distance which should be traveled by arrival aircraft before landing. Therefore, the trajectory of any arrival aircraft could be represented by fluctuated headings along with the DTGs, as shown in Fig. 1. The headings during the final approaching and landing are located at the far left of Fig. 1(b), where the corresponding DTGs are close to zero. During such a flight phase, the heading is about 350°, since the aircraft is going to land on Runway 35. When the aircraft is passing the entry fix as shown in Fig. 1(a), the heading is around 100°, which is located at the far right of Fig. 1(b).
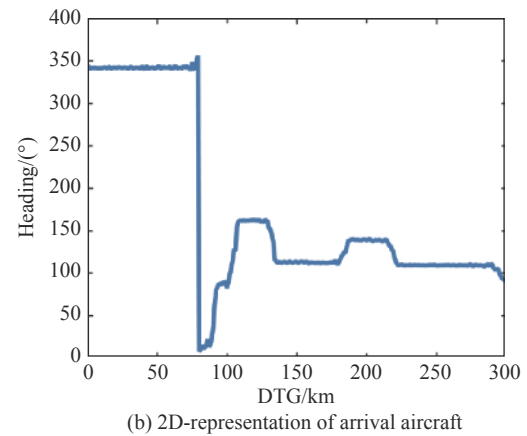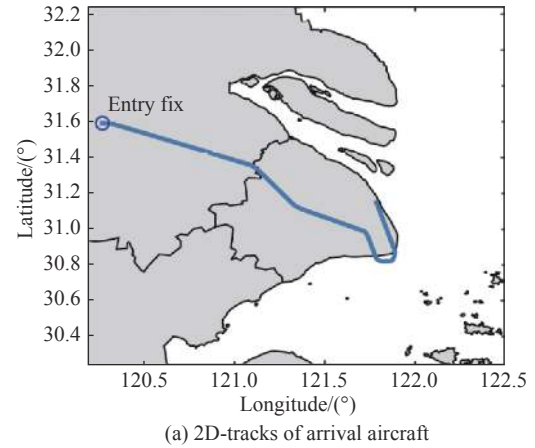

(a) 2D-tracks of arrival aircraft

(b) 2D-representation of arrival aircraft

Fig. 1 Arrival aircraft with different representation methods

As mentioned above, the heading is an important attribute, which can represent the routes, procedures, and instructions. However, the heading is not a monotonous value, and its range is between 0° to 360°. Sometimes, the heading value cannot reflect the actual flight situation. Therefore, the heading needs to be adjusted by considering left- or right-turn, as the following steps.

(i) Find the difference in heading between two consecutive points.

$$\Delta\psi_j^i = \psi_j^i - \psi_{j+1}^i, \ j = 1, \cdots, L(i)-1 \tag{1}$$

where j is the particular point of the ith trajectory and $L(i)$ represents the total number of points of the ith trajectory.

(ii) Modify $\Delta\psi$ according to the left- or right-turn.

$$\begin{cases} \Delta\psi_j^{i'} = \Delta\psi_j^i - 360°, \ \Delta\psi_j^i > 180° \\ \Delta\psi_j^{i'} = \Delta\psi_j^i + 360°, \ \Delta\psi_j^i < -180° \\ \Delta\psi_j^{i'} = \Delta\psi_j^i, \ \text{otherwise} \end{cases} \tag{2}$$

where $\Delta\psi_j^{i'}$ is the modified value of $\Delta\psi_j^i$.

(iii) Calculate the adjusted heading.

$$\begin{cases} \theta_j^i = \psi_j^i, \ j = L(i) \\ \theta_j^i = \theta_{j+1}^i + \Delta\psi_j^{i'}, \ j = L(i)-1,\cdots,2,1 \end{cases} \tag{3}$$

(iv) Resample trajectory at equal flight distance intervals.

$$\begin{cases} \hat{\theta}_k^i = \theta_{j-1}^i + \dfrac{\theta_{j-1}^i - \theta_j^i}{D_{j-1}^i - D_j^i}\left(\omega k - D_{j-1}^i\right), \ D_j^i \leqslant \omega k \leqslant D_{j-1}^i \\ \hat{\theta}^i = \left(\hat{\theta}_0^i, \hat{\theta}_1^i, \cdots, \hat{\theta}_{D_1^i/\omega}^i\right) \end{cases} \tag{4}$$

where θ_j^i, D_j^i and $\hat{\theta}_k^i$ represent the ith trajectory's adjusted heading, DTG and resampled adjusted heading, ω is the size of the distance interval, $k = \{0,1,2,\cdots, D_1^i/\omega\}$. Now, $\hat{\theta}^i$ denotes the ith trajectory in the new representation.

(v) Apply a low pass filter for noise reduction, since each trajectory is a bit noisy.

Eventually, Fig. 2 provides the original and adjusted heading according to the corresponding DTGs. The adjustment could not only eliminate the leaping heading but also reflect the pilot's manipulation of aircraft by making a left- or right-turn. The increase of heading means a right-turn and vice versa.
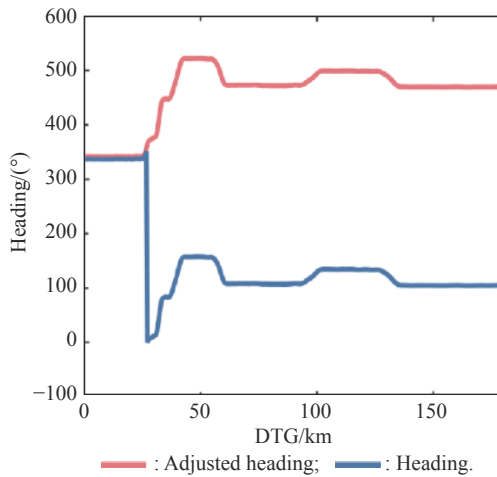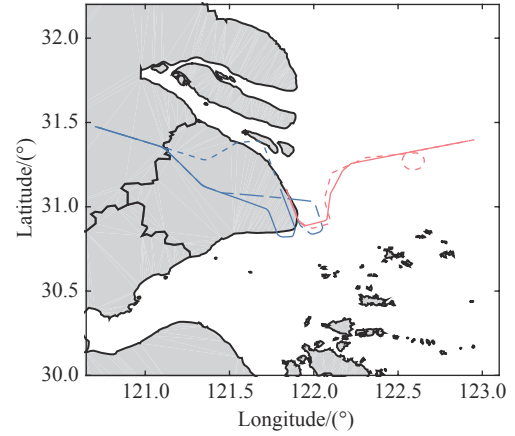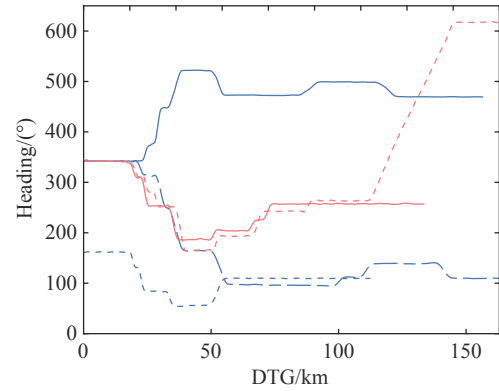

Fig. 2   Illustration of the original and adjusted heading

2.3   Essential characteristics of novel representation

Heading vs. DTG, can represent the routes, procedures, and instructions. Moreover, such novel representation for arrival aircraft possesses several other essential characteristics, as shown in Fig. 3.

(a) 2D-tracks of arrival aircraft

(b) 2D-representation of arrival aircraft

—— : AC#1a;   - - - - : AC#1b;   —— : AC#2a;
—— : AC#2b;   - - - - : AC#2c.

Fig. 3   Previous and proposed representation of arrival aircraft

First of all, the proposed trajectory representation could discriminate the arrival trajectories via different entry fixes. AC#1a and AC#2a come from different entry fixes, and the adjusted headings of these two aircraft are far from each other at the entries.

Secondly, the proposed trajectory representation could differentiate the landing directions via the same entry fix. It could be seen in Fig. 3 that AC#2a and AC#2c come from the same entry fix while landing on the opposite runways. The adjusted headings of these two aircraft differ approximately 180°, as shown in the far left of Fig. 3(b).

Thirdly, the proposed trajectory representation could distinguish the landing modes of arrival trajectories via the same entry fix and on the same runway. From Fig. 3(a), we could find that AC#2a makes continuous left-turn while AC#2b makes continuous right-turn, during the approaching phase. Correspondingly, in Fig. 3(b), the adjusted heading of AC#2a is decreasing while that of AC#2b is increasing.

Last but not least, the proposed trajectory representation could identify whether the holding pattern is used or not. Taking AC#1a and AC#1b as examples, there is 360°

separated between their adjusted headings at the entry fix, as shown in Fig. 3(b).

3. Trajectory clustering of arrivals

3.1 Trajectory similarity

For those arrival aircraft, via the same entry fix, on the same runway, and using the same landing mode, the proposed representation method could attribute significantly to recognizing the flight patterns from the different trajectories.

We take AC#1 and AC#2 as examples, as shown in Fig. 4. AC#1 generally follows the published STAR, which consists of Downwind Leg, Base Leg, and Final. On the contrary, AC#2 seems to follow the instructions delivered by ATCO, who guides AC#2 direct to the Base Leg and cut into the Final earlier than AC#1. These are basic instructions within TMA and denoted as "short-cut" and "parallel offset." Such instructions are clearly reflected in the corresponding representation of the trajectories, as shown in Fig. 4 (b). The "parallel offset" is displayed as the horizontal shift about 30 km distance from landing, while the "short-cut" is shown as the vertical shift around 50 km distance from landing.
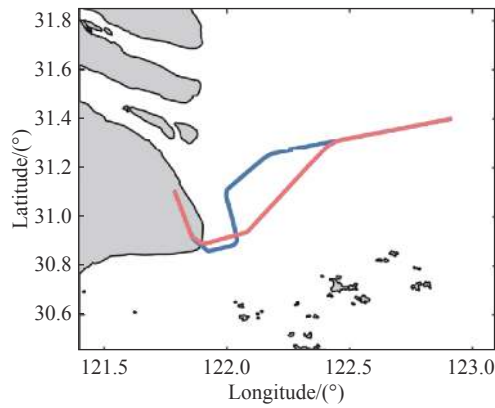
(a) 2D-tracks of arrival aircraft
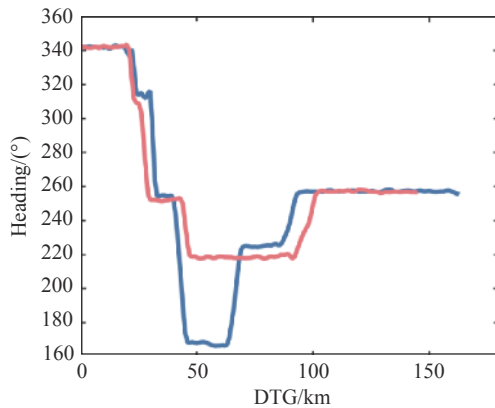
(b) 2D-representation of arrival aircraft

—— : AC#1;   —— : AC#2.

Fig. 4   Illustration of flight patterns for arrival aircraft

To measure the dissimilarity between these two trajectories, DTW is used and the schematic diagram is provided in Fig. 5. Therefore, a similarity matrix could be obtained by calculating each pair of all those trajectories.
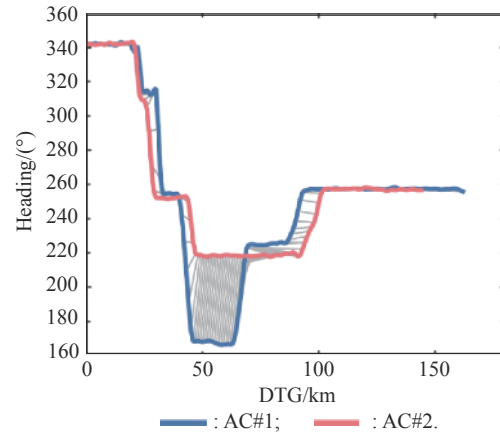

—— : AC#1;   —— : AC#2.

Fig. 5   Schematic diagram of DTW distance between two trajectories

$$
\mathrm{DTW}\left(\tilde{\theta}^{x},\tilde{\theta}^{y}\right)=
\begin{cases}
0, & \tilde{\theta}^{x}\text{ and }\tilde{\theta}^{y}\text{ are empty}\\[2pt]
\infty, & \tilde{\theta}^{x}\text{ or }\tilde{\theta}^{y}\text{ is empty}\\[2pt]
\min\left\{\mathrm{DTW}\left(R\left(\tilde{\theta}^{x}\right),R\left(\tilde{\theta}^{y}\right)\right),\right.\\
\qquad \mathrm{DTW}\left(R\left(\tilde{\theta}^{x}\right),\tilde{\theta}^{y}\right),\\
\left.\qquad \mathrm{DTW}\left(\tilde{\theta}^{x},R\left(\tilde{\theta}^{y}\right)\right)\right\}+\\
\qquad d\left(\tilde{\theta}_0^{x},\tilde{\theta}_0^{y}\right),\text{ otherwise}
\end{cases}
\tag{5}
$$

where $d\left(\tilde{\theta}_0^{x},\tilde{\theta}_0^{y}\right)$ represents the absolute values between trajectory points $\tilde{\theta}_0^{x}$ and $\tilde{\theta}_0^{y}$, $R\left(\tilde{\theta}^{x}\right)$ and $R\left(\tilde{\theta}^{y}\right)$ represent the trajectory segments of $\tilde{\theta}^{x}$ and $\tilde{\theta}^{y}$ after removing their first trajectory points, respectively.

3.2 Trajectory clustering

Each trajectory could be represented as a sample with the same size by the obtained similarity matrix. We could take advantage of all kinds of standard clustering methods for trajectory clustering. In this paper, k-means and DBSCAN algorithms are implemented to cluster the arrival trajectories. The whole processes are illustrated in Fig. 6, which consists of the following steps.

Step 1   Acquire the raw data, radar data or ADS-B data.

Step 2   Obtain the trajectories, including call sign, position, speed, heading, and flight plan-related information.

Step 3   Represent the obtained trajectories by adjusted heading and DTGs by using (1)−(4).

Step 4   Obtain the particular set of trajectories sharing the following characteristics:

(i) Coming via the same entry fix;

(ii) Landing in the same direction;

(iii) Using the same landing mode;

(iv) Eliminating the trajectories of holding aircraft.

Step 5   Calculate the similarity based on DTW distance by using (5).

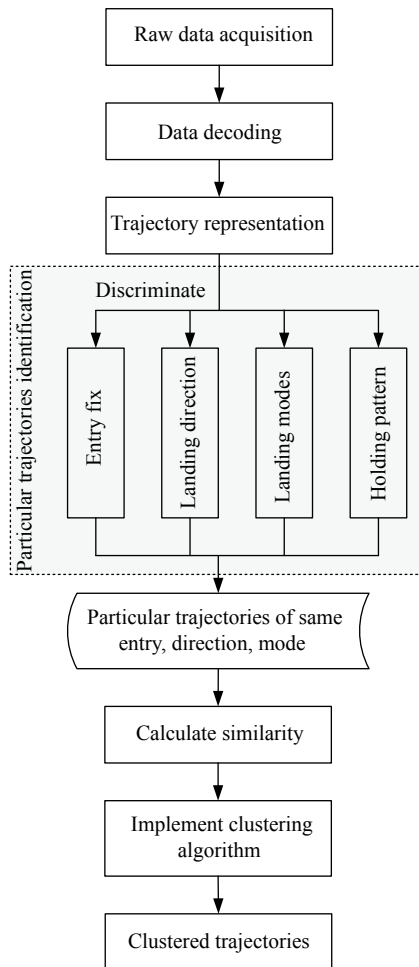**Step 6**   Implement clustering algorithms to obtain the grouped trajectories.

Fig. 6   Processes of trajectory clustering based on the represented trajectories

As for DBSCAN, two parameters should be predefined: ε and *MinPts*, where ε is the size of neighborhood, and *MinPts* is the minimum number of points required to form a cluster. For a given trajectory tr_i, if there are more than *MinPts* trajectories in its neighborhood ($\text{DTW}(\theta^{tr_i}, \cdot) < \varepsilon$), a new cluster is built and all trajectories in its neighborhood are considered to be the same group. When no trajectory could be added to any cluster, the DB-

SCAN terminates.

As for k-means, each trajectory is represented by a vector based on (5). The cluster number k should be predefined, and the algorithm is performed as follows:

Step 1   Choose k trajectories as initial centroids.

Step 2   Compute Euclidean distance of all trajectories to each centroid.

Step 3   Assign trajectories to different groups if such reassignment decreases the indicator of the within-cluster metric.

Step 4   Average trajectories in each cluster to obtain k new centroids.

Step 5   Repeat Step 2 to Step 4 until the cluster assignments do not change, or the number of iterations reaches its predefined value.

The clustered trajectories could reflect the prevail tracks flown by the arrival aircraft. The potential advantages of such prevail tracks mining are twofold. One is assisting us to distinguish the arrival flight patterns, which in turn could identify the controller's intent while delivering instructions. The other is helping us to obtain more composite distribution of flight time since different arrival aircraft within each group share some common characteristics. Therefore, in this paper, the evaluation process will not only focus on the flight patterns but also pay attention to the distribution of flight time based on kernel density estimation (KDE).

## 4. Case studies

4.1   Data preparation

We take the trajectories of aircraft which landed on Shanghai Pudong International Airport (ZSPD) as examples in this paper.

The raw data used in this paper is from radar recording, and each record of trajectories contains the following information: time stamp, call sign, heading, latitude, longitude, pressure altitude, etc. Here, each record with the same call sign belongs to a particular arrival aircraft. The value of DTG needs to be calculated based on the latitude/longitude and the universal transverse mercator (UTM) projection.

Fig. 7(a) and Fig. 7(b) present the radar trajectories about four rush hours during two weeks' operations within Shanghai TMA. Fig. 7(a) provides all the trajectories of departure, arrival, and overfly aircraft operated in Shanghai TMA. Fig. 7(b) focuses on the arrival trajectories landing on ZSPD, which also include four entering fixes (SASAN, BK, MATNU, and DUMET). Since those are the operations during rush hours, the ATCOs need to rely on radar vectoring for establishing and maintaining the landing sequence. As a result, the arrival trajectories are disorganized and haphazard.
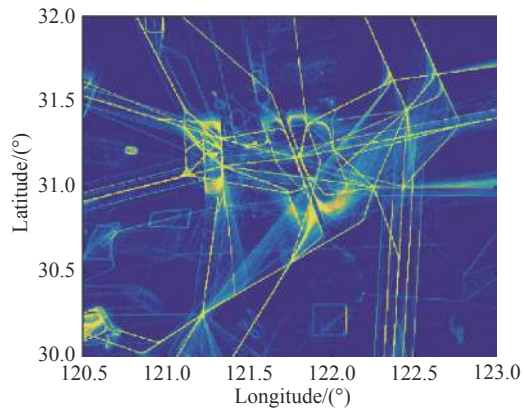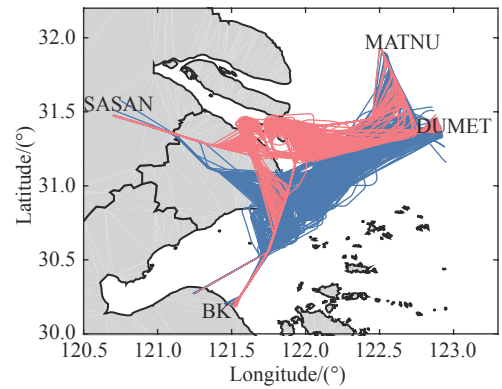
(a) 2D-tracks of aircraft within Shanghai TMA
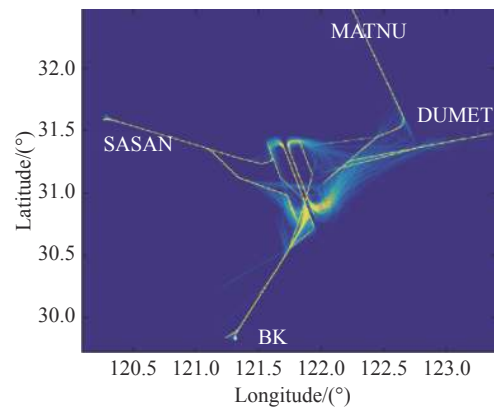
(b) 2D-tracks of arrival aircraft of ZSPD
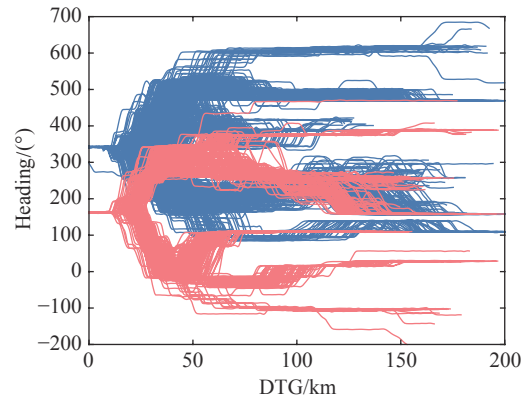
Fig. 7 Illustration of aircraft trajectories

(a) 2D-tracks of north- and south-bound operation

(b) New representation of north- and south-bound operation
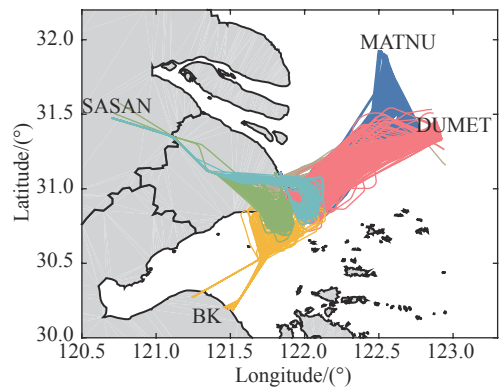
(c) 2D-tracks of north-bound operation
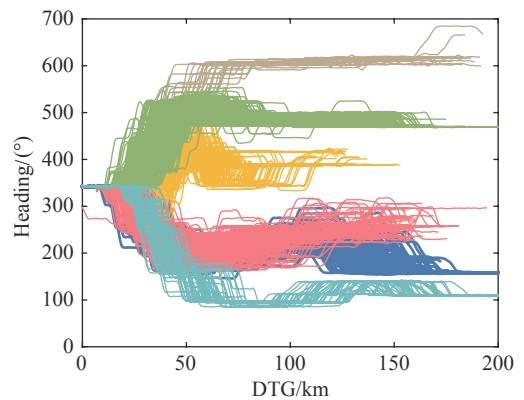
(d) New representation of north-bound operation

Fig. 8 Illustration of trajectories discrimination

4.2 Trajectory clustering

Firstly, we apply (1)−(4) for the heading ψ and DTG D of arrival trajectories and obtain the new representation $\hat{\theta}$, with which the aircraft of implementing the holding pattern are eliminated. The trajectory resampling interval ω is set as 2 km, which can reduce the number of trajectory points while keeping the information of the trajectory as much as possible.

Secondly, we discriminate the particular trajectories, as shown in Fig. 8, with the same landing direction (in Fig. 8(b), there exist two segments when DTG is zero), via the same entry, and by the same landing mode (in Fig. 8(b), there exist several branches at the end of the DTG axis with the same landing direction). In this process, the discriminating task could be easily fulfilled by only using the adjusted headings. From Fig. 8 (d), we could find that the proposed heading adjustment can identify different landing modes. By taking aircraft via SASAN as examples, the adjusted headings of two landing modes (denoted by green and cyan colors) are different.

Thirdly, we chose the arrival trajectories via SASAN and DUMET as candidates for trajectory clustering, since those trajectories contain more flight patterns. We calculate the similarity matrix based on DTW distance, and then implemente k-means and DBSCAN algorithms for clustering.

Finally, the clustered results are shown in Figs. 9−12 based on different clustering algorithms and via particular entry fixes, BK, and SASAN. When implementing the DBSCAN algorithm, the algorithm parameters should be appropriately predefined, including the radius of the neighborhood ε and the minimum number of neighbors $MinPts$. In the case of BK entry, $\varepsilon = 240$ and $MinPts = 6$. In the case of SASAN entry, $\varepsilon = 140$ and $MinPts = 8$. When implementing the k-means algorithm, the number of clusters is predefined as five due to the comparison purpose with the DBSCAN algorithm. Every figure consists of each group of trajectories (a) to (e), all clustered arrival trajectories (f), the whole heading representation (g), and flight time and distance distributions (h) and (i). Moreover, the figures for each group are arranged in order of flight distances.

From Fig. 9 to Fig. 12, several interesting findings are summarized as follows. On the one hand, the proposed methods of trajectory representation and clustering could successfully partition the arrival trajectories into different groups. Each group of trajectories shares the same flight pattern, for examples: "short-cut" pattern in Fig. 9 (a) and Fig. 10 (a); "dog-leg" pattern in Fig. 9 (b) and Fig. 9 (d), Fig. 10 (b) and Fig. 10(d); and "parallel offset" pattern in Fig. 11 (d) and Fig. 12 (d). Furthermore, after partitioning, the flight time and distance distribution are more compact. On the other hand, DBSCAN is a competitive candidate algorithm in the clustering domain since it could identify the outliers. However, in our attempts, the DBSCAN algorithm sometimes treat the potential flight pattern into outliers. Fig. 10 (e) is a case in point, in which the right "dog-leg" pattern is wrongly classified as outliers.
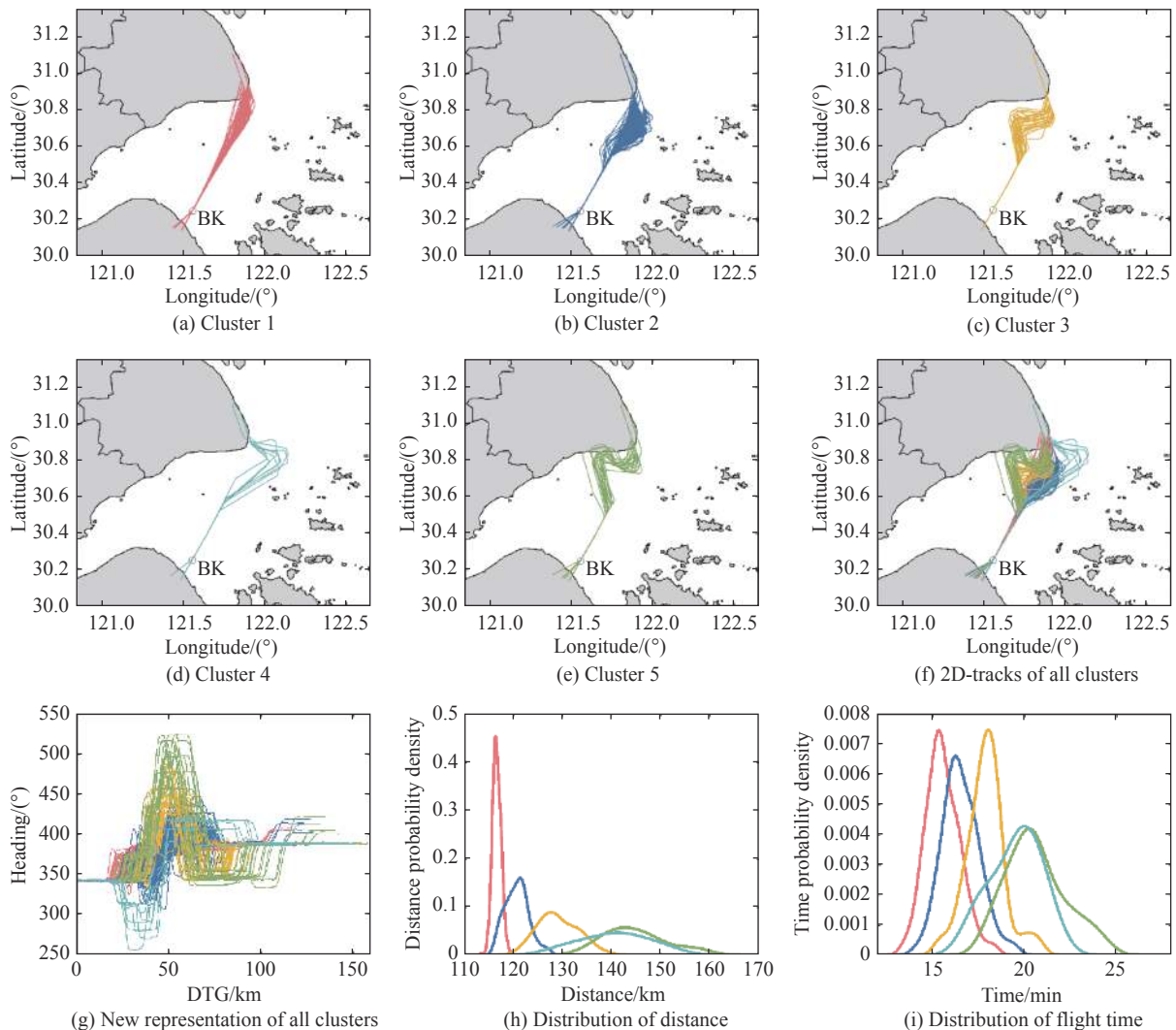

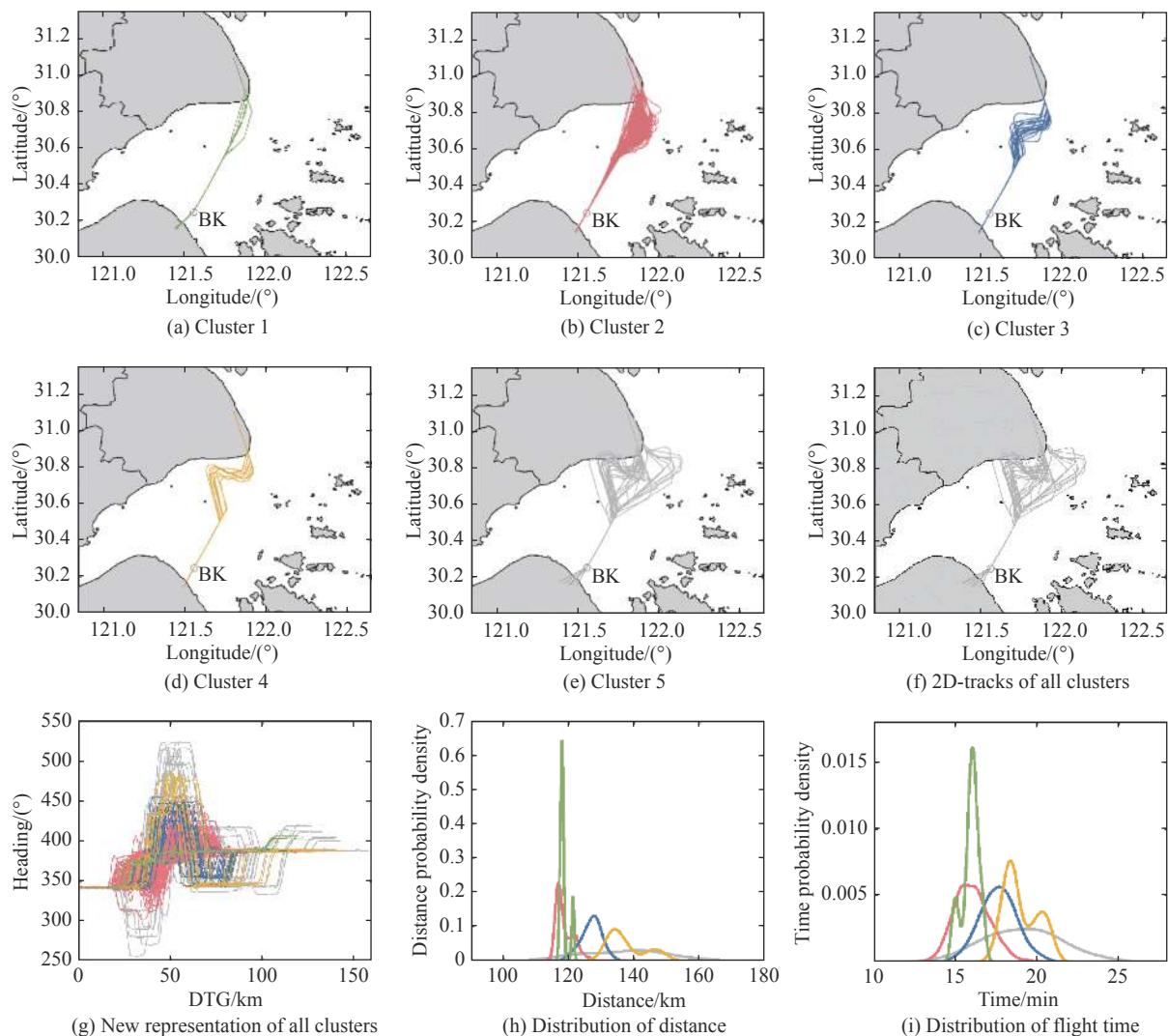Fig. 9   Illustration of clustering results via BK fix by the k-means algorithm

Fig. 10   Illustration of clustering results via BK fix by the DBSCAN algorithm

(d) Cluster 4

(e) Cluster 5

(f) 2D-tracks of all clusters

(g) New representation of all clusters

(h) Distribution of distance

(i) Distribution of flight time
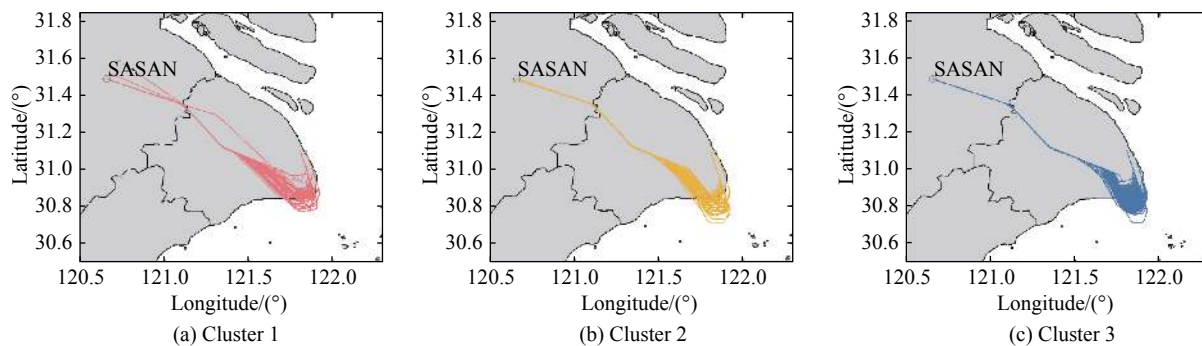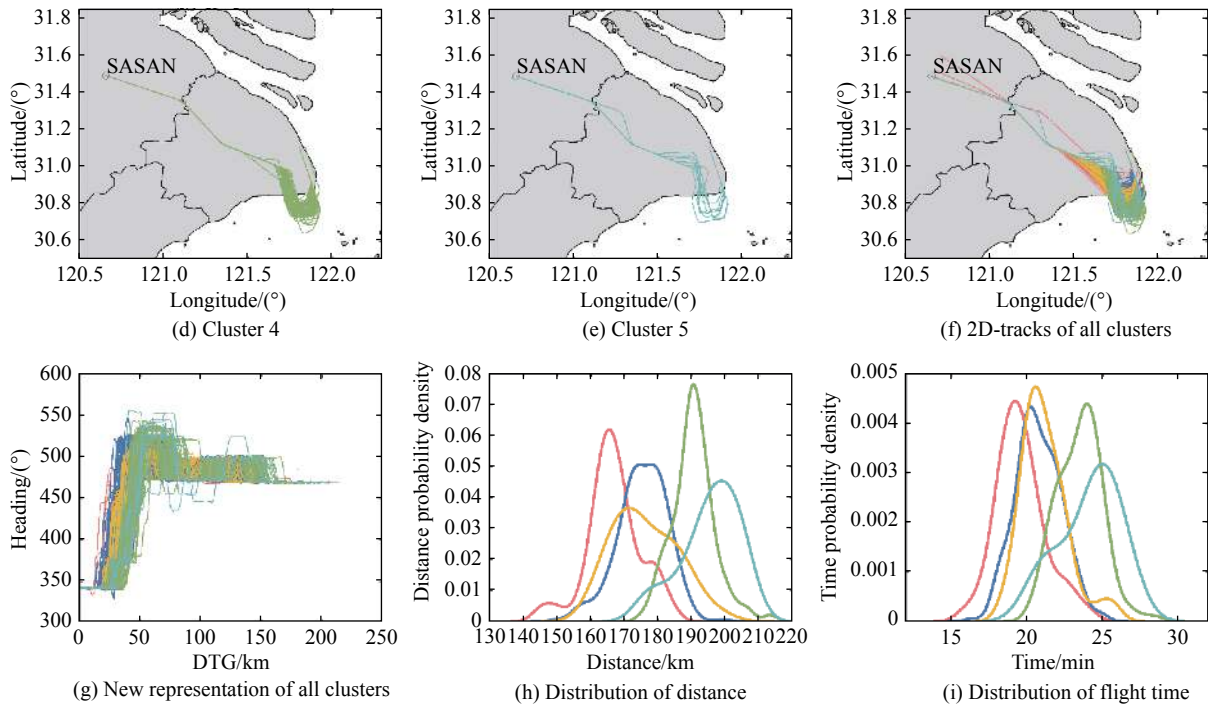
Fig. 11   Illustration of clustering results via SASAN fix by the k-means algorithm

(a) Cluster 1

(b) Cluster 2

(c) Cluster 3

(d) Cluster 4

(e) Cluster 5

(f) 2D-tracks of all clusters

(g) New representation of all clusters   (h) Distribution of distance   (i) Distribution of flight time

Fig. 12　Illustration of clustering results via SASAN fix by the DBSCAN algorithm

4.3　Results and discussion

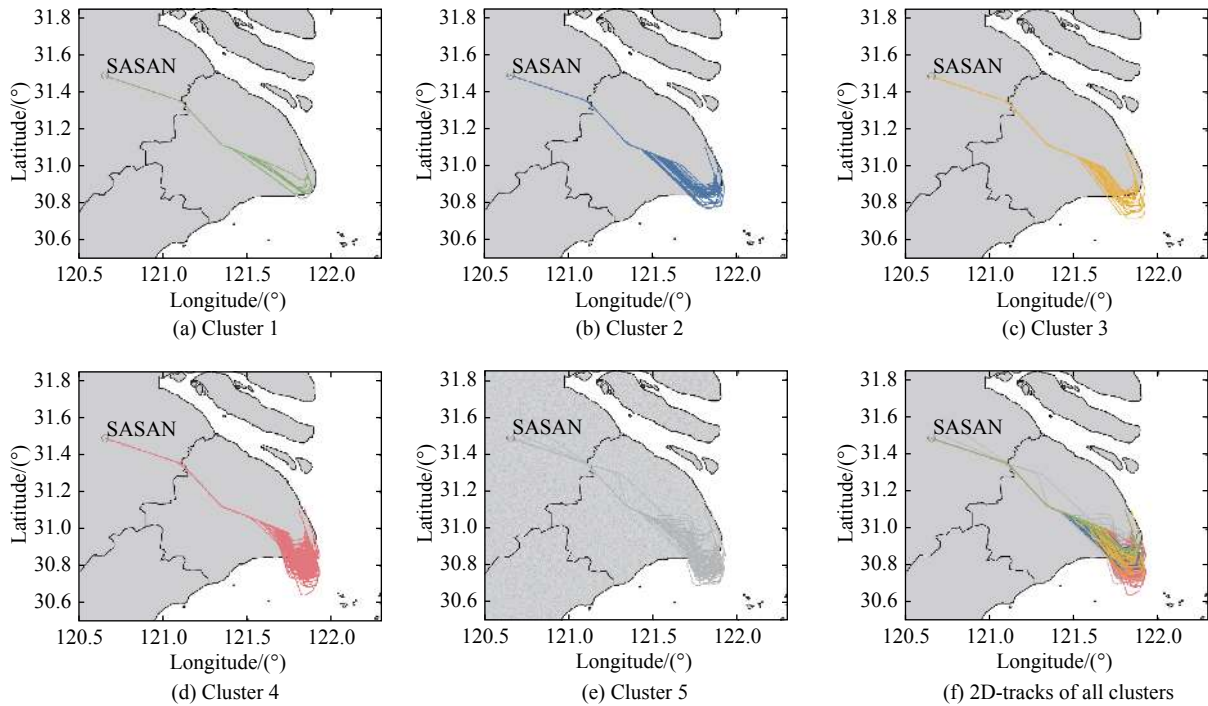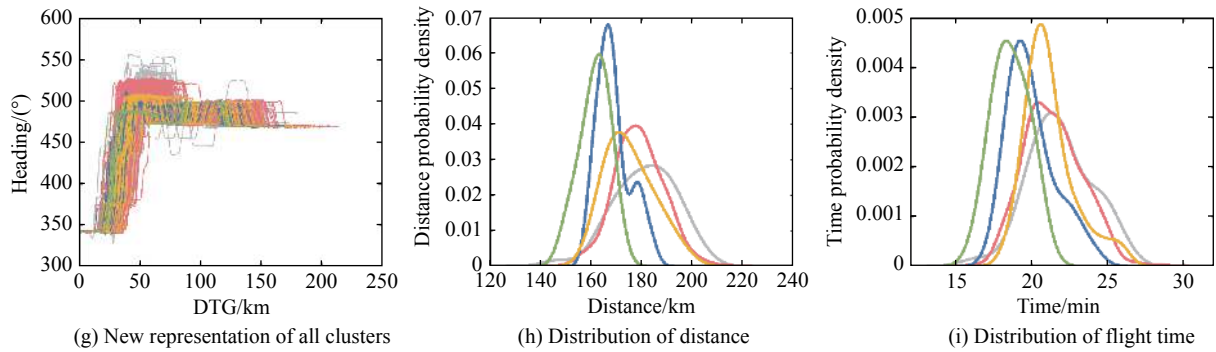One purpose of arrival trajectory clustering is to make the flight time within TMA more compact, so we present the flight time distribution of each cluster, including the total trajectories, as shown in Fig. 13.

The probability densities of flight time are obtained by using the kernel density estimation method, in which Gauss kernel is adopted. It should be noted that the probability density in Fig. 13 is based on the amount of each group. Moreover, Fig. 13 only provides the results by the k-means algorithm. For comparison reasons, Table 1 lists the mean/median values, lower/upper limits, and intervals of 95% and 75% confidence levels for the flight time distribution.
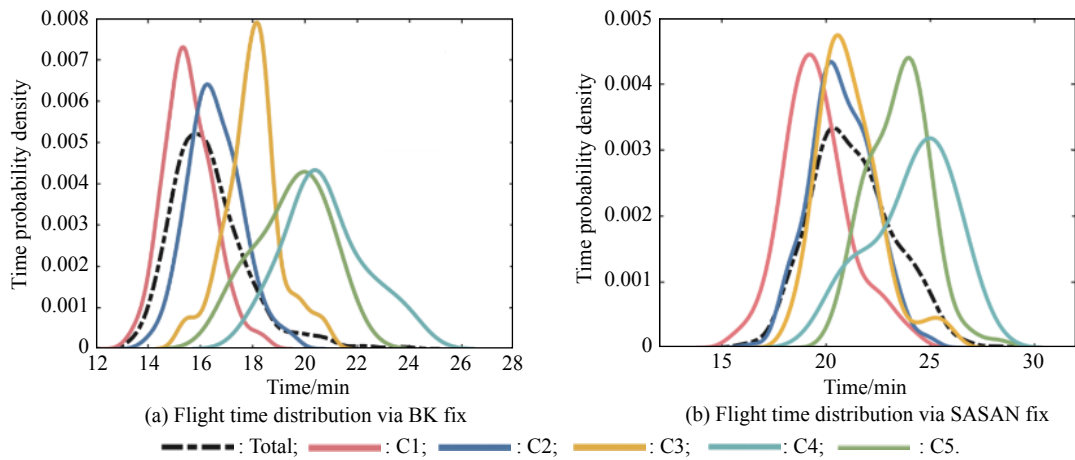

(a) Flight time distribution via BK fix   (b) Flight time distribution via SASAN fix

- - - : Total;　——— : C1;　——— : C2;　——— : C3;　——— : C4;　——— : C5.

Fig. 13　Illustration of flight time distribution of each cluster by the k-means algorithm

Table 1　Statistical analysis of flight time distribution

min

Fix	Group	Mean	Median	Confidence level (95%)			Confidence level (75%)		
				Lower limit	Upper limit	Interval	Lower limit	Upper limit	Interval
BK	Total	16.3	16.1	13.4	19.3	5.9	14.6	18.1	3.5
	C1	15.6	15.5	13.7	17.4	3.7	14.5	16.6	2.1
	C2	16.6	16.5	14.5	18.6	4.1	15.4	17.7	2.3
	C3	18.0	18.1	15.9	20.1	4.2	16.8	19.2	2.4
	C4	20.8	20.6	17.7	23.8	6.1	19.0	22.6	3.6
	C5	19.5	19.8	16.9	22.0	5.1	18.0	21.0	3.0
SASAN	Total	21.3	21.0	17.4	25.2	7.8	19.0	23.6	4.6
	C1	19.6	19.4	16.4	22.7	6.3	17.7	21.4	3.7
	C2	21.1	20.9	18.2	24.1	5.9	19.4	22.9	3.5
	C3	20.7	20.5	17.8	23.6	5.8	19.0	22.4	3.4
	C4	23.5	23.7	20.7	26.3	5.6	21.8	25.1	3.3
	C5	24.1	24.5	20.3	27.9	7.6	21.9	26.3	4.4

From Fig. 13 and Table 1, we could find that on the one hand the mean or median flight time in each cluster is different from the value of the total trajectories, where the average flight time is 20.8 min for arrival aircraft via BK fix in Group C4 while 16.3 min for the total. There is a deviation of 4.5 min. On the other hand, most of the intervals in each cluster are smaller than the corresponding value of the total trajectories, which means that the clustered results could produce more compact flight time distribution.

Furthermore, the random forest (RF) regression method is adopted to carry out ETA prediction. We design a feature set for the regression task, such as entry states, current operation situations, historical flight time, and so on. The simulation results are shown in Fig. 14.
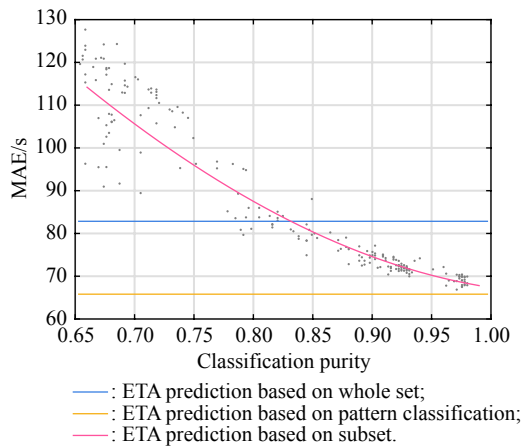

Fig. 14   Mean absolute error (MAE) of ETA prediction

The blue line shows that the MAE is about 83 s when using the RF method based on the whole data set. In order to reduce the MAE of ETA prediction, the total trajectories are clustered into different patterns, for each pattern, there is an individual RF trained for ETA prediction. If the classification purity reaches 100%, the MAE of ETA prediction could be reduced to 66 s, denoted by the yellow line in Fig. 14. Such grey dots line represents the MAE of ETA prediction according to different classification purities, while the pink line indicates that the higher the classification purity, the smaller the MAE of ETA prediction. Consequently, through our proposed method, we could estimate the flight time more reliably.

Subsequently, one example will be given to demonstrate why the proposed representation and DTW measurement are able to group the trajectories into different clusters, which could indicate the different flight patterns or air traffic controllers' intents. Suppose there is a runway (RWY09), as shown in Fig. 15(a), four arrival aircraft follow different instructions delivered by ATCOs (controllers' intents) to establish and maintain the landing sequence, including standard route, "short-cut" "dog-leg", and "parallel-offset". Fig. 15(b) presents the corresponding representation of the four kinds of trajectories. Finally, the similarity measurements based on DTW are listed in Table 2 between the different flight patterns.
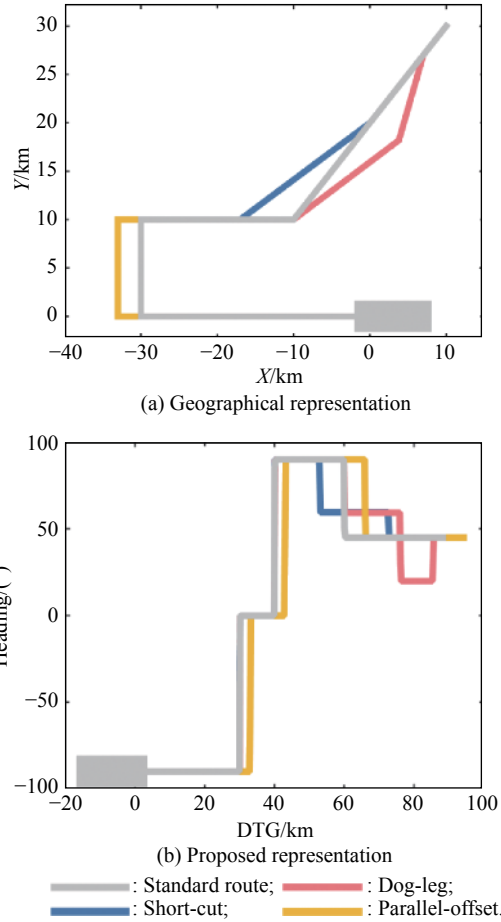

(a) Geographical representation

(b) Proposed representation

Fig. 15   Illustration of different flight patterns

Table 2   Similarity measurements based on DTW between different flight patterns

Flight pattern	Standard	Short-cut	Dog-leg	Parallel-offset
Standard	0.00	130.00	358.86	21.50
Short-cut	130.00	0.00	317.50	129.62
Dog-leg	358.86	317.50	0.00	322.93
Parallel-offset	21.50	129.62	322.93	0.00

The similarity measurements listed in Table 2 indicate that the proposed trajectory representation method has a positive impact on arrival trajectory clustering. It is due to the fact that the dissimilarities between the different flight patterns could be easily captured through exerting DTW measurement on the new representation of these arrival trajectories. Such characteristics could help us distinguish the typical flight patterns or air traffic controllers' intents during arrival operation.

5. Conclusions

This paper presents a trajectory representation and clustering method for arrival aircraft to get access to a better understanding of the real operation and reliable flight time estimation. We substitute the distance-heading representation for the longitude-latitude representation since the proposed one could better describe the controller's intention, which in turn mostly defines a particular aircraft's trajectory. Furthermore, based on the proposed representation method, we develop a semantic trajectory clustering approach, which could not only easily distinguish the landing direction, landing mode, entry fix, and holding pattern, but also recognize the particular flight patterns. The case studies of the real operation in ZSPD validate the effectiveness and efficiency of our semantic trajectory clustering approach. We separate the trajectories according to their flight pattern and find that the distribution of flight time of different flight patterns is diverse. In other words, we establish the relation between flight time and flight patterns, since the airspace situation determines the flight time of arrival aircraft within the TMA. Such efforts, in return, could help to improve the performance of ETA prediction. Moreover, that semantic trajectory clustering makes it possible for us to describe the airspace situation with flight patterns.

However, some limitations are worth noting. Although there are some interesting work and impressive results in this study, we address only a 2-D spatial clustering problem in the arrival and approaching scenarios. Future work should, therefore, include the temporal information under the scenarios of inbound and outbound traffic within the TMA.

References

[1] KISTAN T, GARDI A, SABATINI R, et al. An evolutionary outlook of air traffic flow management techniques. Progress in Aerospace Sciences, 2017, 88: 15–42.

[2] ALESSANDRO G, ROBERTO S, TREVOR K. Multi-objective 4D trajectory optimization for integrated avionics and air traffic management systems. IEEE Trans. on Aerospace and Electronic Systems, 2018, 55(1): 170–181.

[3] LI M Z, RYERSON M S. Reviewing the DATAS of aviation research data: diversity, availability, tractability, applicability, and sources. Journal of Air Transport Management, 2019, 75: 111–130.

[4] REHM F. Clustering of flight tracks. Proc. of the American Institute of Aeronautics and Astronautics, 2010: 1–9.

[5] GARIEL M, SRIVASTAVA A N, FERON E. Trajectory clustering and an application to airspace monitoring. IEEE Trans. on Intelligent Transportation Systems, 2011, 12(4): 1511–1524.

[6] ANDRIENKO G, ANDRIENKO N, FUCHS G, et al. Clustering trajectories by relevant parts for air traffic analysis. IEEE Trans. on Visualization and Computer Graphics, 2017, 24(1): 34–44.

[7] OLIVE X, MORIO J. Trajectory clustering of air traffic flows around airports. Aerospace Science and Technology, 2019, 84: 776–781.

[8] MARZUOLI A, GARIEL M, VELA A, et al. Data-based modeling and optimization of en route traffic. Journal of Guidance, Control, and Dynamics, 2014, 37(6): 1930–1945.

[9] BOMBELLI A, SOLER L, TRUMBAUER E, et al. Strategic air traffic planning with Fréchet distance aggregation and rerouting. Journal of Guidance, Control, and Dynamics, 2017, 40(5): 1117–1129.

[10] BOMBELLI A, TORNÉ A S, TRUMBAUER E, et al. Improved clustering for route-based eulerian air traffic modeling. Journal of Guidance, Control, and Dynamics, 2019, 42(5): 1064–1077.

[11] GERDES I, TEMME A, SCHULTZ M. Dynamic airspace sectorisation for flight-centric operations. Transportation Research Part C: Emerging Technologies, 2018, 95: 460–480.

[12] MURÇA M C R, HANSMAN R J, LI L, et al. Flight trajectory data analytics for characterization of air traffic flows: a comparative analysis of terminal area operations between New York, Hong Kong and Sao Paulo. Transportation Research Part C: Emerging Technologies, 2019, 97: 324–347.

[13] MURCA M C R, HANSMAN R J. Identification, characterization, and prediction of traffic flow patterns in multi-airport systems. IEEE Trans. on Intelligent Transportation Systems, 2018, 20(5): 1683–1696.

[14] CARMONA M A A, NIETO F S, GALLEGO C E V. A data-driven methodology for characterization of a terminal manoeuvring area in multi-airport systems. Transportation Research Part C: Emerging Technologies, 2020, 111: 185–209.

[15] HONG S, LEE K. Trajectory prediction for vectored area navigation arrivals. Journal of Aerospace Information Systems, 2015, 12(7): 490–502.

[16] WANG Z Y, LIANG M, DELAHAYE D. A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area. Transportation Research Part C: Emerging Technologies, 2018, 95: 280–294.

[17] ZHENG Y. Trajectory data mining: an overview. ACM Trans. on Intelligent Systems and Technology, 2015, 6(3): 1–41.

[18] YUAN G, SUN P H, ZHAO J, et al. A review of moving object trajectory clustering algorithms. Artificial Intelligence Review, 2017, 47(1): 123–144.

[19] MORRIS B T, TRIVEDI M M. Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2287–2301.

[20] ZHAO L B, SHI G Y. A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition. Ocean Engineering, 2019, 172: 456–467.

[21] SALAUN E, GARIEL M, VELA A E, et al. Aircraft proximity maps based on data-driven flow modeling. Journal of Guidance, Control, and Dynamics, 2012, 35(2): 563–577.

[22] LEE J G, HAN J, WHANG K Y. Trajectory clustering: a partition-and-group framework. Proc. of the ACM SIGMOD International Conference on Management of Data, 2007: 593–604.

[23] MAHBOUBI Z, KOCHENDERFER M J. Learning traffic patterns at small airports from flight tracks. IEEE Trans. on Intelligent Transportation Systems, 2017, 18(4): 917–926.

Biographies
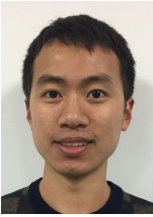
GUI Xuhao was born in 1995. He received his B.S. degree in communications and transportation from Civil Aviation University of China in 2018. He is pursuing his M.S. degree in Nanjing University of Aeronautics and Astronautics. His research interests are air traffic management and machine learning.
E-mail: ShowhowGui@outlook.com

ZHANG Junfeng was born in 1979. He received his Ph.D. degree in control theory and control engineering from Nanjing University of Aeronautics and Astronautics in 2008. He is currently an associate professor in College of Civil Aviation at Nanjing University of Aeronautics and Astronautics. His research interests include 4D trajectory generation, prediction and optimization, as well as decision support tool for the air traffic control.
E-mail: zhangjunfeng@nuaa.edu.cn

PENG Zihan was born in 1997. She received her B.S. degree in communications and transportation from Nanjing Agricultural University in 2019. She is pursuing her M.S. degree in Nanjing University of Aeronautics and Astronautics. Her research interests are air traffic management and machine learning.
E-mail: fluff9797@163.com