A Hybrid Method Combining Improved *K*-means Algorithm with BADA Model for Generating Nominal Flight Profiles

Tang Xinmin[1,2*], *Gu Junwei*[1], *Shen Zhiyuan*[1], *Chen Ping*[2], *Li Bo*[1]

1. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China;
2. The 28th Research Institute of China Electronic Technology Group Corporation, Nanjing 210007, P. R. China

(Received 4 November 2015; revised 24 March 2016; accepted 25 April 2016)

Abstract: A high-precision nominal flight profile, involving controllers' intentions is critical for 4D trajectory estimation in modern automatic air traffic control systems. We proposed a novel method to effectively improve the accuracy of the nominal flight profile, including the nominal altitude profile and the speed profile. First, considering the characteristics of trajectory data, we developed an improved *K*-means algorithm. The approach was to measure the similarity between different altitude profiles by integrating the space warp edit distance algorithm, thereby to acquire several fitted nominal flight altitude profiles. This approach breaks the constraints of traditional *K*-means algorithms. Second, to eliminate the influence of meteorological factors, we introduced historical gridded binary data to determine the en-route wind speed and temperature via inverse distance weighted interpolation. Finally, we facilitated the true airspeed determined by speed triangle relationships and the calibrated airspeed determined by aircraft data model to extract a more accurate nominal speed profile from each cluster, therefore we could describe the airspeed profiles above and below the airspeed transition altitude, respectively. Our experimental results showed that the proposed method could obtain a highly accurate nominal flight profile, which reflects the actual aircraft flight status.

Key words: air transportation; flight profile; *K*-means algorithm; space warp edit distance (SWED) algorithm; trajectory prediction

CLC number: V355 **Document code**: A **Article ID**: 1005-1120(2016)04-0414-11

0 Introduction

Airspace resources and air traffic congestion have been rapidly decreasing due to the fast-expanding global aviation industry. Acknowledging this dilemma, the United States and Europe have planned and implemented a next generation air traffic management system, whose core technology is 4D trajectory-based operation[1]. As 4D trajectory prediction with high precision has become critical, its essential issue of mining the nominal flight profile, including intentions of controllers, has also become pivotal. It is necessary to develop a state-of-the-art method to improve the accuracy of the nominal flight profile, so that changes in the aircraft flight status can be predicted in advance and controllers can receive necessary information to handle aircraft conflicts timely.

In recent years, the important role of 4D trajectory prediction in air traffic automation systems has been widely studied. Two types of algorithms are employed for this study: (1) Methods based on an aircraft dynamic model. For example, Coppenbarger developed a method for 4D trajectory prediction based on an aircraft dynamic model in take-off and climbing phases[2]. Gong and Chan proposed a approach that employed a climbing timetable from the flight performance manual to obtain aircraft aerodynamic models and dynamic equations for trajectory estimation[3].

* Corresponding author, E-mail address: tangxinmin@nuaa.edu.cn.

How to cite this article: Tang Xinmin, Gu Junwei, Shen Zhiyuan, et al. A hybrid method combining improved *K*-means algorithm with BADA model for generating nominal flight profiles[J]. Trans. Nanjing Univ. Aero. Astro., 2016,33(4): 414-424.

http://dx.doi.org/10.16356/j.1005-1120.2016.04.414

Wu and Guo utilized a numerical integral algorithm and a simplified algebraic algorithm to calculate flight profiles in terminal areas[4]. A kinematics model established on the flight phase can allow trajectory prediction, but the results are not ideal due to the lack of consideration of meteorological factors and performance errors for different aircraft types. (2) Fitting methods based on the center trajectory using data mining technology. For example, Gariel et al. proposed a *K*-means clustering method based on radar data to monitor aircraft in terms of their actual operation status[5]. For busy areas with many airports, Leiden and Atkins used the clustering method based on grid tracking to analyze different airports, thereby to determine the average trajectories of different aircraft in the arrival and departure phases[6]. Wang and Huang analyzed aircraft flight data using the fuzzy clustering method in the arrival phase and calculated the average center trajectory[7]. However, the methods mentioned above are limited within the data gleaned only from arrival or departure phases rather than the entire flight profile. Therefore, a new concept called the basic flight model was proposed for building horizontal trajectories, altitude profiles, and speed profiles depending on the flight phases, but this method cannot obtain the entire flight profile in an effective manner due to the lack of meteorological factors[8]. Xing et al. proposed a novel method based on aircraft meteorological data relay (AMDAR) data to generate a flight profile, which can obtain the nominal flight altitude profile based on the dynamic space warping algorithm and the speed profile based on a hybrid method that combines the large oval distance algorithm with the base of aircraft data (BADA) model[9-10]. However, AMDAR data have a large sampling interval, so it is not accurate to use the large oval distance between two trajectory points as the aircraft flight distance. Furthermore, the time accuracy of AMDAR data is not adequate, and thus the nominal speed profile includes specific errors.

To determine the actual aircraft flight status, we proposed an improved *K*-means algorithm to cluster the altitude profiles of historical trajectory data, thereby to obtain several centers of the flight profiles, i. e., fitted nominal altitude profiles, without relying on the Newtonian mechanics model. Second, we developed a method for integrating the historical meteorological data and speed conversion model based on BADA to obtain the nominal speed profiles. Hence, combining these methods with airway meteorological forecasts could revise the estimated flight trajectory in order to obtain a more accurate 4D trajectory.

1 Generating Nominal Altitude Profiles

1.1 Flight altitude profiles and their characteristics

The flight altitude profile is a two-dimensional image that represents the relationship between the flight altitude and range, i. e., the trajectory in the vertical direction of a geographic coordinates system. As shown in Fig. 1, a typical flight altitude profile can be divided into three stages: *a*—*d* represents the climbing stage, *e*—*f* the cruise stage, and *g*—*j* the descent stage. The specific flight altitude parameters for each stage were described in Ref. [11].

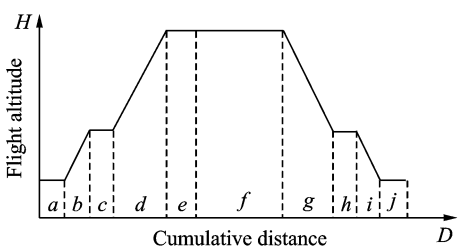


Fig. 1 A typical flight altitude profile

Secondary surveillance radar (SSR) can obtain real-time trajectory data, where the scan cycle is usually 3—5 s, which means that flight trajectory data are not updated continuously, but in a form of a series of discrete trajectory points. Let $M=\{L_1, L_2, \cdots, L_n\}$ be the set of n historical flight altitude profiles with the same origin and destination, where each profile $L_i=\{l_i^1, l_i^2, \cdots, l_i^m\}$

with unequal length comprises m historical points (i denotes a serial number of a profile). Let $l_i^j \in H \times D$ be an arbitrary point, where j is the point index, $H \subset R$ embeds the altitude-stamp variables, and $D \subset R$ the distance-stamp variable, i.e., $l_i^j = (h_i^j, s_i^j)$, where $h_i^j \in H$ and $s_i^j \in D$, with the constraint of $s_i^m > s_i^n$ whenever $m > n$. Therefore, the characteristics of the flight altitude profile can be summarized as: (1) Each profile comprises discrete points; (2) The length of each profile is unequal; (3) There is a delay that corresponds to the distance shaft between the two profiles.

1.2 Altitude profile measurment based on SWED

Marteau designed a time warp edit distance algorithm[12] based on the characteristics of historical trajectory data. The method performed better than classic dynamic time warping algorithms[13-14], because it involved the influence of the non-matching cost and distance deviation so as to measure the distance between time sequence trajectories. but it lacked the limitation of a time warping degree. Thus, considering the relationship between the flight altitude and distance, we proposed an improved space warp edit distance (SWED) algorithm to measure the distance between two profile point series. The proposed method constructs a cost function to calculate the influence of non-matching cost and distance deviations.

Let $L_u^{1,p}$ be the set of finite profile point series: $L_u^{1,p} = \{l_u^1, l_u^2, \cdots, l_u^p\}$, where the superscript is a discrete distance index that varies from 1 to p, and the subscript the serial number of a profile. Ω represents the empty distance series (with null length) and by convention, $L_u^{1,0} = \Omega$, so Ω is the member of set $L_u^{1,p}$. $|L_u^{1,p}|$ is the length of $L_u^{1,p}$. Λ denotes the null sample. For a pair of distance series samples, $(l_u^i, l_v^j) \neq (\Lambda, \Lambda)$ is written as $l_u^i \rightarrow l_v^j$, where if $l_u^i \neq \Lambda$ and $l_v^j \neq \Lambda$, $l_u^i \rightarrow l_v^j$ is called a match operation; if $l_v^j = \Lambda$, $l_u^i \rightarrow \Lambda$ is called a delete L_u operation; and if $l_u^i = \Lambda$, $\Lambda \rightarrow l_v^j$ is called a delete L_v operation.

We defined the similarity between any two profile point series L_1 and L_2 with finite lengths of p and q, respectively, as

$$\delta_{\lambda,\gamma}(L_1^{1,p}, L_2^{1,q}) = \min\begin{cases}\delta_{\lambda,\gamma}(L_1^{1,p-1}, L_2^{1,q}) + \Gamma(l_1^p \rightarrow \Lambda) & \text{Delete } l_1^p \\ \delta_{\lambda,\gamma}(L_1^{1,p-1}, L_2^{1,q-1}) + \Gamma(l_1^p \rightarrow l_2^q) & \text{Match} \\ \delta_{\lambda,\gamma}(L_1^{1,p}, L_2^{1,q-1}) + \Gamma(\Lambda \rightarrow l_2^q) & \text{Delete } l_2^q\end{cases} \quad (1)$$

$$\Gamma(l_1^p \rightarrow \Lambda) = d(l_1^p, l_1^{p-1}) + \lambda = d_{LP}(h_1^p, h_1^{p-1}) + \gamma \cdot (s_1^p - s_1^{p-1}) + \lambda$$

$$\Gamma(l_1^p \rightarrow l_2^q) = d(l_1^p, l_2^q) + d(l_1^{p-1}, l_2^{q-1}) = d_{LP}(h_1^p, h_2^q) + d_{LP}(h_1^{p-1}, h_2^{q-1}) + \gamma \cdot (|s_1^p - s_2^q| + |s_1^{p-1} - s_2^{q-1}|)$$

$$\Gamma(\Lambda \rightarrow l_2^q) = d(l_2^q, l_2^{q-1}) + \lambda = d_{LP}(h_2^q, h_2^{q-1}) + \gamma \cdot (s_2^q - s_2^{q-1}) + \lambda \quad (2)$$

where $p \geqslant 1$, $q \geqslant 1$, and Γ is an arbitrary cost function, which assigns a nonnegative real number $\Gamma(l_1^p \rightarrow l_2^q)$ to each edit operation $l_1^p \rightarrow l_2^q$. $\lambda (\lambda \geqslant 0)$ and γ denote the non-matching constant penalty and distance penalty coefficient, respectively. The recursion is initialized for $i > 1$ and $j > 1$, by setting

$$\begin{aligned}&\delta_{\lambda,\gamma}(L_1^{1,0}, L_2^{1,0}) = 0 \\ &\delta_{\lambda,\gamma}(L_1^{1,0}, L_2^{1,j}) = \infty \\ &\delta_{\lambda,\gamma}(L_1^{1,i}, L_2^{1,0}) = \infty\end{aligned} \quad (3)$$

The rules of the matching and deleting operation are as follows. (1) The editing process is performed from left to right, and if i is an index on the segments of L_1 and j on the segments of L_2, the initial process setting is $i = j = 1$. (2) A matching operation adds one to i and j simultaneously, which are denoted by $i \leftarrow i+1$ and $j \leftarrow j+1$. A deleting $L_1(L_2)$ operation adds one to $i(j)$ only, which is denoted by $i \leftarrow i+1 (j \leftarrow j+1)$. (3) After segment i (j) in $L_1(L_2)$ has been processed using either a matching or a deleting operation, it is impossible for it to be edited again. A recursive algorithm is conducted to complete the overall process therefore to obtain the final global minimum cost function value.

1.3 Clustering and fitting nominal altitude profiles based on an improved *K*-means algorithm

Traditional K-means clustering algorithms set the value k and initial altitude profile centers

randomly, usually leading to the clustering results trapped by local optima[15-16]. In addition, traditional algorithms are no longer applicable to generating cluster centers, given the characteristics of the altitude profile. Therefore, we proposed a nominal altitude profile clustering and fitting method based on an improved *K*-means algorithm to obtain the *k* fitted altitude profile centers in an effective manner. The technical framework of the method is shown in Fig. 2.

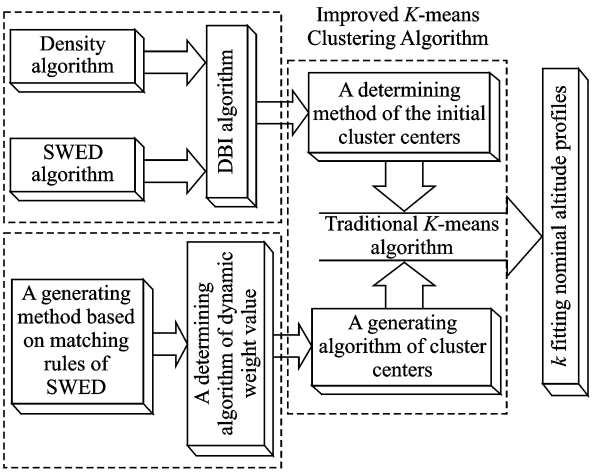


Fig. 2 Technical framework of the improved *K*-means algorithm

In order to determine *k* initial altitude profile centers, we proposed a novel concept , i. e. , an altitude profile object. For any altitude profile object L_u, if the condition $d_{\text{swed}}(L_u, L_v) < \xi$ is met by taking a threshold ξ, then the object number can be defined as the density of L_u. Let $M = \{L_1, L_2, \cdots, L_n\}$ be the set of altitude profiles and the profile with the maximum density in the set M is regarded as the first initial cluster center Z_1. Then, the second initial center Z_2 is determined by seeking the profile farthest from Z_1 and the initial altitude profile center Z_k is determined as

$$Z_k = Z_i : \forall j \leqslant n,\ \min\{d_{\text{swed}}(Z_i, Z_1), d_{\text{swed}}(Z_i, Z_2), \cdots\} \geqslant \min\{d_{\text{swed}}(Z_j, Z_1), d_{\text{swed}}(Z_j, Z_2), \cdots\} \quad 2 < i \leqslant n \tag{4}$$

Davies-Bouldin index (DBI) is used to synthetically measure the separation degree of the intra-cluster and the condensation degree of inter-clusters. The minimum DBI value, i. e. , $\text{DBI}_{\text{new}} > \text{DBI}_{\text{last}}$, can determine an optimal cluster number. The model equation is summarized as

$$\text{DBI} = \frac{1}{k}\sum_{u=1}^{k} \max\left\{\frac{S_u + S_v}{d_{\text{swed}}(Z_u, Z_v)}\right\} \tag{5}$$

where k is a cluster number, $d_{\text{swed}}(Z_u, Z_v)$ the distance between the cluster center Z_u in cluster u and Z_v in the cluster v, and S_u the standard error between each trajectory and the cluster center Z_u in the cluster u.

The method used for determining the cluster center based on the traditional *K*-means algorithm cannot be applied to altitude profiles, so we proposed a novel algorithm for generating cluster centers by integrating the weight and matching rules of SWED. The core strategy of the algorithm is to superimpose two arbitrary intra-cluster profiles repeatedly. Assuming that L_N denotes an altitude profile that is not superimposed in the cluster C_k and L_{last} a center that is superimposed, the algorithm for generating the center between the altitude profiles L_N and L_{last} is as

$$\begin{aligned} L_{\text{new}} &= \{l^i_{\text{new}} \mid i \in [1, |L_{\text{last}}|]\} = \\ &\begin{cases} \{\omega_1(h^i_{\text{last}}, s^i_{\text{last}}) + \omega_2(h^j_N + \gamma(s^i_{\text{last}} - s^j_N), s^j_N)\} & (\text{Ⅰ}) \\ \{\omega_1(h^i_{\text{last}}, s^i_{\text{last}}) + \omega_2(\omega_3[h^j_N + \gamma(s^i_{\text{last}} - s^j_N)] + \\ \quad \omega_4[h^{j+1}_N + \gamma(s^i_{\text{last}} - s^{j+1}_N)], \omega_3 s^j_N + \omega_4 s^{j+1}_N)\} & (\text{Ⅱ}) \end{cases} \\ &= \{\omega_1(x_i, s_{x_i}) + \omega_2(y_i, s_{y_i}) \mid i \in [1, |L_{\text{last}}|]\} = \\ &\omega_1 X + \omega_2 Y = Z \end{aligned} \tag{6}$$

Accordingly, L_{new} can be represented by: $Z = \omega_1 X + \omega_2 Y$, where (Ⅰ) denotes SWED matching, i. e. , one-to-one matching directly based on the segmented points, and (Ⅱ) SWED non-matching, employing a strategy for one-to-two matching in the interior points, where the two matching points are determined by the minimum distance and second minimum distance; h^i_{last} the altitude of a point in L_{last}; h^j_N and h^{j+1}_N are the altitudes of points in L_N; ω_1 is a weight determined by the superimposed intra-cluster profiles; ω_2 a weight determined by the remaining intra-cluster profile, and ω_3 and $\omega_4(1-\omega_3)$ are the weights determined by two profile points l^j_N and l^{j+1}_N, which are calculated by Eq. (8). Let $\boldsymbol{X}$ be the matching matrix, $\boldsymbol{Y}$ the matched matrix, and $\boldsymbol{Z}$ the center generating matrix

$$L_{\text{last}}:\boldsymbol{X}=\begin{bmatrix} x_1 & s_{x_1} \\ x_2 & s_{x_2} \\ \vdots & \vdots \\ x_n & s_{x_n} \end{bmatrix}_{2\times n},\ L_N:\boldsymbol{Y}=\begin{bmatrix} y_1 & s_{y_1} \\ y_2 & s_{y_2} \\ \vdots & \vdots \\ y_n & s_{y_n} \end{bmatrix}_{2\times n}$$

$$L_{\text{new}}:\boldsymbol{Z}=\begin{bmatrix} z_1 & s_{z_1} \\ z_2 & s_{z_2} \\ \vdots & \vdots \\ z_n & s_{z_n} \end{bmatrix}_{2\times n} \tag{7}$$

where $n=|L_{\text{last}}|$ denotes the length of L_{last}, Profile point (x_i, s_{x_i}) matches with (y_i, s_{y_i}), and i a serial number that varies from 1 to n.

$$\omega_3=\frac{|h^i_{\text{last}}-[h^{j+1}_N+\gamma(s^i_{\text{last}}-s^{j+1}_N)]|}{|h^i_{\text{last}}-[h^j_N+\gamma(s^i_{\text{last}}-s^j_N)]|+|h^i_{\text{last}}-[h^{j+1}_N+\gamma(s^i_{\text{last}}-s^{j+1}_N)]|} \tag{8}$$

In order to obtain the cluster centers for the altitude profile in an effective manner, the dynamic weight is determined based on a regression analysis algorithm to ensure that the generated center is the closest to each intra-cluster profile. The object S is calculated as

$$S=\frac{N-1}{N}d^2_{\text{swed}}(L_{\text{new}},L_{\text{last}})+\frac{1}{N}d^2_{\text{swed}}(L_{\text{new}},L_N) \tag{9}$$

where N represents the number of superimposed profiles. The function mentioned above is subject to the following constraints

$$\boldsymbol{Z}=\omega_1\boldsymbol{X}+\omega_2\boldsymbol{Y},\ \omega_1+\omega_2=1 \tag{10}$$

To minimize the value of object S, ω_1 and ω_2 can be determined by the partial derivative of function S, i.e., $S'(\omega_1)=S'(\omega_2)=0$. Finally, k clusters and the fitted nominal altitude profiles are obtained by applying the superimposition strategy repeatedly.

2 Nominal Speed Profile Generation

2.1 Flight speed profiles and their characteristics

A flight speed profile is a two-dimensional image that describes the relationship between speed and altitude during the flight process. Due to changes in the air density and sound velocity with increases in altitude, the calibrated airspeed (CAS), true airspeed (TAS) and Mach number will also change, as shown in Fig. 3.

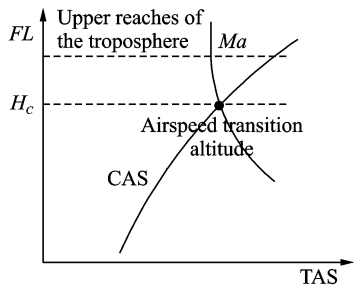


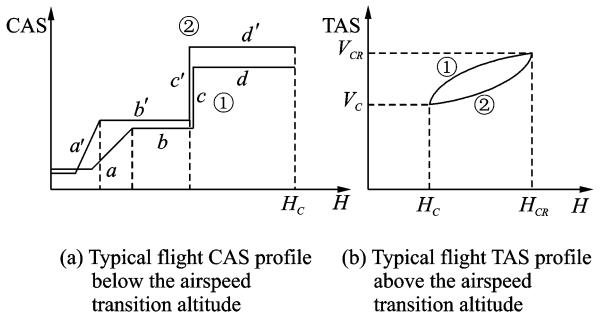
Fig. 3 Airspeed transition altitude

In practice, the airspeed transition altitude written as H_C is usually set for convenient flight control. In general, the aircraft operates with constant CAS below the airspeed transition altitude. However, considering the effects of air compressibility, the aircraft flies with a constant Mach number above the airspeed transition altitude.

Therefore, according to the airspeed transition altitude, the aircraft flight speed profile is divided into two stages. The phase below the airspeed transition altitude is shown in Fig. 4(a). In curve ① of the climbing phase, a represents the take-off and acceleration stage, b the climbing stage with constant CAS, c the acceleration stage at constant altitude, and d the climbing stage with constant CAS. Similarly, as shown in curve ②, the changing trend in the descending stage is equivalent to the reverse of the climbing stage process. However, considering the relationship between the true airspeed and altitude above the airspeed transition altitude, as shown in Fig. 4 (b), curve ① is a two-dimensional image describing the relationship between TAS and altitude (transition altitude to cruising altitude), whereas curve ② describes the relationship between TAS and altitude (cruising altitude to transition altitude).

In Fig. 4, H_{CR}, V_{CR} represent the cruising altitude and the cruising speed of aircraft, respectively; V_C is the speed at transition altitude H_C.

Similar to the altitude profile, let $S_{\text{NOM}}=\{s^1_{\text{NOM}}, s^2_{\text{NOM}}, \cdots, s^m_{\text{NOM}}\}$ be the set of the nominal speed profile, where $s^i_{\text{NOM}}(v^i_{\text{NOM}}, h^i_{\text{NOM}})\in S_{\text{NOM}}$ denotes an arbitrary speed profile point, $v^i_{\text{NOM}}\in R$ represents a one-dimensional speed, and $h^i_{\text{NOM}}\in R$



(a) Typical flight CAS profile below the airspeed transition altitude (b) Typical flight TAS profile above the airspeed transition altitude

Fig. 4 A typical flight speed profile

denotes a one-dimensional altitude.

2.2 Airway meteorological interpolation based on IDW

In order to obtain more accurate flight speed profiles, it is necessary to eliminate the influence of meteorological factors on the ground speed. Radar trajectory data do not include airway meteorological data, so gridded binary (GRIB) forecast data gleaned by World Area Forecast System, which are promoted worldwidely by the International Civil Aviation Organization and World Meteorological Organization, are introduced to supply service for aviation. GRIB data with a resolution of 0.5°×0.5° are updated every 6 h, thereby providing meteorological information according to the pressure distribution at 37 levels, with factors such as the air pressure, wind speed, air temperature[17-18]. Therefore, we proposed a data fusion method based on GRIB data and radar data to obtain airway meteorological data via the inverse distance weighted(IDW) interpolation algorithm.

The IDW algorithm is an exact local spatial interpolation method, which estimates values to forecast the summation of linear points using observation values. Then, we can use the inverse of the distance between points with observation values to weight the forecast values[19-20]. Let $L_{NOM}=\{l^1_{NOM}, l^2_{NOM}, \cdots, l^m_{NOM}\}$ be the nominal altitude profile, then the location (latitude, longitude, and altitude) and weather (east-west wind, north-south wind, and temperature) for an arbitrary sample l^i_{NOM} are denoted by $z^i_{NOM}(\varphi^i_{NOM}, \lambda^i_{NOM}, h^i_{NOM})$ and $m^i_{NOM}(u^i_{WS}, v^i_{WS}, t^i_{NOM})$, respectively. The IDW interpolation formula is

$$m^i_{NOM}(u^i_{WS}, v^i_{WS}, t^i_{NOM}) = \sum_{k=1}^{n} \frac{m^k_{grib}(u^k_{grib}, v^k_{grib}, t^k_{grib})}{(D_k)^{\lambda}} \Big/ \sum_{k=1}^{n} \frac{1}{(D_k)^{\lambda}} \tag{11}$$

where m^i_{NOM} represents a forecast value, $m^k_{grib}(u^k_{grib}, v^k_{grib}, t^k_{grib})$ the kth observation value, n the number of observation values, λ the exponent sign, which is the main factor that affects the estimation of IDW, and D_k the great circle distance from the kth observation value to the forecast value, which is calculated as

$$D_k = (R+h^i_{NOM})\arccos[\sin\varphi^i_{NOM}\sin\varphi^k_{Grib} + \cos\varphi^i_{NOM}\cos\varphi^k_{grib}\cos(\lambda^k_{grib}-\lambda^i_{NOM})] \tag{12}$$

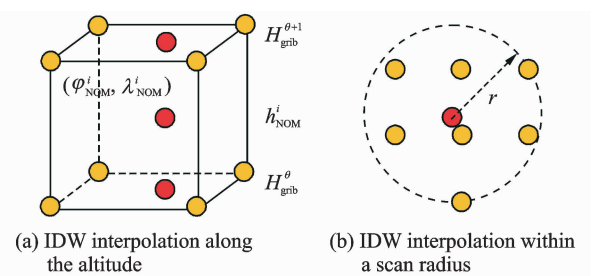
where R represents the radius of the earth, and $(\varphi^k_{grib}, \lambda^k_{grib})$ the longitude and latitude of the meteorological observation value.

The airway meteorological interpolation algorithm is described as following steps.

Step 1 Determine the GRIB altitude level where the altitude of the forecast value h^i_{NOM} is located. Assume that h^i_{NOM} is between the altitude levels of H^{θ}_{grib} and $H^{\theta+1}_{grib}$, as shown in Fig. 5(a).

Step 2 At the altitude H^{θ}_{grib} or $H^{\theta+1}_{grib}$, the observation values can be determined by the center $(\varphi^i_{NOM}, \lambda^i_{NOM})$ and the radius r (as shown in Fig. 5(b)), which can be used to obtain the forecast values $(\varphi^i_{NOM}, \lambda^i_{NOM}, H^{\theta}_{grib})$ and $(\varphi^i_{NOM}, \lambda^i_{NOM}, H^{\theta+1}_{grib})$ by IDW interpolation.

Step 3 The final forecast value $m^i_{NOM}(u^i_{WS}, v^i_{WS}, t^i_{NOM})$ at location z^i_{NOM} is calculated by IDW interpolation, according to the forecast values at the locations $(\varphi^i_{NOM}, \lambda^i_{NOM}, H^{\theta}_{grib})$ and $(\varphi^i_{NOM}, \lambda^i_{NOM}, H^{\theta+1}_{grib})$.



(a) IDW interpolation along the altitude (b) IDW interpolation within a scan radius

Fig. 5 Airway meteorological interpolation principle

Aiming at the forecast errors in historical GRIB data, we introduced a method of wind forecast error statistics that can effectively analyze

the existential errors[21]. Firstly, the forecast error of the wind is defined as a random field $\Delta \boldsymbol{w}: \boldsymbol{R} \times \boldsymbol{R}^3 \rightarrow \boldsymbol{R}^2$ where $\Delta \boldsymbol{w}(t, P)$ denotes the wind at time $t \in \boldsymbol{R}$ and at location point $P(x, y, z) \in \boldsymbol{R}^3$. For the sake of simplicity, wind in the vertical direction is neglected and the case focuses on where $\Delta \boldsymbol{w}(t, P) \in \boldsymbol{R}^2$ is Gaussian with zero mean and covariance matrix $R(t, P, t', P') \in \boldsymbol{R}^{2\times 2}$. Assume that the wind-field is isotropic and the east-west wind and south-north wind are uncorrelated. Therefore, according to these hypotheses, the covariance matrix $\boldsymbol{R}(t, P, t', P')$ can be simplified as

$$\boldsymbol{R}(t, P, t', P') = E[\Delta \boldsymbol{w}(t, P) \cdot \Delta \boldsymbol{w}^{\mathrm{T}}(t', P')] = \begin{bmatrix} r(t, P, t', P') & 0 \\ 0 & r(t, P, t', P') \end{bmatrix} \tag{13}$$

where $E[\cdot]$ represents expected value, and the covariance $r(t, P, t', P')$ can be expressed as

$$r(t, P, t', P') = \sigma(z)\sigma(z') r_t(|t - t'|) \cdot r_{xy}\left(\left\| \begin{matrix} x - x' \\ y - y' \end{matrix} \right\|\right) r_z(|AP(z) - AP(z')|) \tag{14}$$

where $\sigma(z)$ represents the standard deviation of the wind error at altitude z and $AP(z)$ the atmospheric pressure which can be calculated via using the standard atmosphere model. In addition, according to the reported data in Ref. [22], functions $r_t(\cdot)$, $r_{xy}(\cdot)$, $r_z(\cdot)$ all decay exponentially, and can be determined by the parameter values obtained from Ref. [22].

2.3 Nominal speed profile generation based on BADA model

The ground speed vector $\boldsymbol{v}_{\mathrm{GS}}^i$ can be obtained easily using a parser of radar data, but the aircraft's heading under the influence of upper-air wind, with wind speed u_{WS}^i and v_{WS}^i, needs to be calculated by GRIB data interpolation. As a result, the TAS vector $\boldsymbol{v}_{\mathrm{TAS}}^i$ can be calculated by the flight speed triangle comprising three vectors, as shown in Fig. 6, the ground speed vector, the TAS vector, and the wind speed vector.

In Fig. 6, α_i is the wind angle, i.e., the angle between the wind direction with the track line, and ε_i the drift angle, i.e., the angle of the track line off the course line that can be determined as

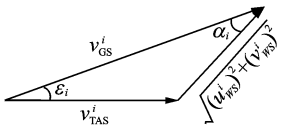


Fig. 6 Flight velocity triangle

$$\frac{\sin[\pi - (\alpha_i + \varepsilon_i)]}{|v_{\mathrm{GS}}^i|} = \frac{\sin\varepsilon_i}{\sqrt{(u_{\mathrm{WS}}^i)^2 + (v_{\mathrm{WS}}^i)^2}} \tag{15}$$

Then, TAS can be calculated by the resultant vector principle as

$$v_{\mathrm{TAS}}^i = \frac{v_{\mathrm{GS}}^i - \sqrt{(u_{\mathrm{WS}}^i)^2 + (v_{\mathrm{WS}}^i)^2} \cdot \cos\alpha_i}{\cos\varepsilon_i} \tag{16}$$

Base of aircraft data (BADA) is a collection of ASCII files which specifies operation performance parameters, airline procedure parameters and performance summary tables for more than 300 aircraft types, which can be used for trajectory simulation and prediction algorithms within the domain of Air Traffic Management (ATM)[23]. During the BADA modelling process, a variety of aircraft performance reference data is acquired from Aircraft Operation Manuals, Aircraft Performance Engineering Programs, and Jane's All the World's Aircraft. The RDAP validation is used to validate the behavior of identified models in respect to real data, including radar data, flight plan data, and meteorology data, which enables identifying all the parameters that describe the BADA aircraft performance model[24].

According to the speed conversion model between the CAS v_{CAS}^i and TAS v_{TAS}^i from BADA, one can obtain the nominal CAS profile

$$v_{\mathrm{CAS}}^i = \left\{ \frac{7p_0}{\rho_0} \left[\left(1 + \delta_i \left(\left(1 + \frac{\rho_0 T_0}{7 p_0 t_{\mathrm{NOM}}^i} (v_{\mathrm{TAS}}^i)^2 \right)^{3.5} - 1 \right) \right)^{1/3.5} - 1 \right] \right\}^{1/2} \tag{17}$$

where $T_0 = 288.15$ K is the standard atmospheric temperature at mean sea level (MSL), $p_0 = 101\,325$ Pa the standard atmospheric pressure at MSL, $\rho_0 = 1.225$ kg/m^3 the standard atmospheric density at MSL, t_{NOM}^i the air temperature at the aircraft's location, which is acquired from GRIB data interpolation, and

$$\delta_i = \begin{cases} (1-6.875\ 59\times10^{-6}\times h_{\mathrm{NOM}}^{i})^{5.255\ 88} \\ \qquad h_{NOM}^{i} \leqslant 360\ 89\ \mathrm{ft} \\ 0.223\ 360\ 9\times\exp\left(\dfrac{360\ 89-h_{\mathrm{NOM}}^{i}}{20\ 805.8}\right) \\ \qquad h_{\mathrm{NOM}}^{i} > 36\ 089\ \mathrm{ft} \end{cases} \tag{18}$$

Finally, based on the partition of the airspeed transition altitude, the speed profile is expressed as the relationship between the CAS and altitude, with the constraint that $h_{\mathrm{NOM}}^{i} < H_C$. By contrast, the speed profile is a two-dimensional relationship between TAS and altitude, with the constraint that $h_{\mathrm{NOM}}^{i} > H_C$, where the airspeed transition altitude can be calculated by

$$\begin{cases} \delta_1 > \delta & H_C = 145\ 442\times(1-\delta_1^{0.190\ 263\ 1}) \\ \delta_1 < \delta & H_C = 4\ 902-20\ 805.8\times\ln\delta_1 \end{cases} \tag{19}$$

where $\delta = 0.223\ 360\ 9$ represents critical pressure ratio, δ_1 a pressure ratio, can be calculated by

$$\delta_1 = \frac{P}{P_0} = \frac{[1+0.2(v_{\mathrm{CAS}}/a_0)^2]^{3.5}-1}{[1+0.2Ma^2]^{3.5}-1} \tag{20}$$

where $a_0 = 340$ m/s is the speed of sound under standard atmospheric pressure, and the parameters v_{CAS} and Ma can be obtained by analyzing the aircraft's speed profile, where the relationship between the TAS and Mach number can be calculated as

$$Ma^{i}_{\mathrm{NOM}} = \frac{v^{i}_{\mathrm{TAS}}}{a^{i}_{\mathrm{NOM}}} = \left(\frac{T_0}{t^{i}_{\mathrm{NOM}}}\right)^{0.5}\frac{v^{i}_{\mathrm{TAS}}}{a_0} \tag{21}$$

3 Simulation and Discussion

3.1 Simulation

In this study, we used 10 radar trajectories from the same flight, i.e., from Xiamen Airport to Wuxi Airport, as an example to discuss their flight profiles. The original relationship between the historical altitude profiles is shown in Fig. 7.

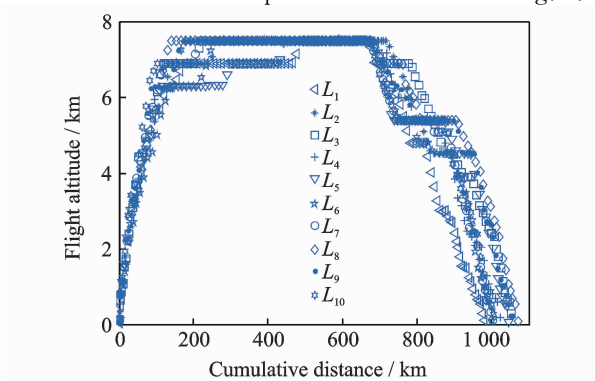


Fig. 7 Original flight altitude profile

First, the distance of the two trajectories was calculated using the SWED algorithm, as shown in Table 1. Second, the K initial cluster centers were determined by integrating the SWED algorithm, density algorithm, and DBI index. Finally, the results of the simulation showed that L_1 was an isolated trajectory and when $K=3$, DBI was minimized (the changing trend in DBI is shown in Table 2), so we determined L_2, L_5, and L_{10} as the initial cluster centers.

Table 1 Distance components based on SWED

Distance	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8
L_1	0	16 293	15 629	15 638	15 766	18 024	16 509	16 721
L_2	16 293	0	12 511	13 037	14 069	12 939	12 497	13 238
L_3	15 629	12 511	0	13 037	14 620	15 337	13 620	13 972
L_4	15 638	13 037	13 037	0	12 540	14 756	11 798	12 904
L_5	15 766	14 069	14 620	12 540	0	15 155	14 113	13 830

Table 2 Davies-Bouldin index (DBI)

K	2	3	4	5	6
DBI	1.845	1.819	1.857	1.865	1.881

We calculated the classification of the altitude profiles using the matching algorithm with SWED and the regression analysis algorithm, as shown in Table 3, where the nominal altitude profile, as shown in Fig. 8, was fitted via cluster II. In addition, the dynamic weight value ω_1 was optimized as 0.512, 0.659, 0.752, and 0.798 in each case, respectively.

Table 3 Classification of the altitude profiles

Cluster Ⅰ	Cluster Ⅱ	Cluster Ⅲ
(L_4, L_5)	(L_2, L_3, L_6, L_7, L_8)	(L_9, L_{10})

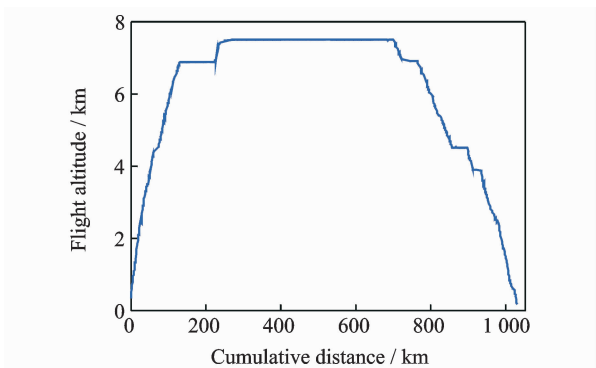


Fig. 8 Fitting of the nominal flight altitude profile

The meteorological information in the nominal trajectory of cluster II was calculated using the IDW interpolation algorithm, where the radius of the earth was 6 371 km and the two exponent signs were 2 and 1, respectively. The interpolation results are shown in Fig. 9.

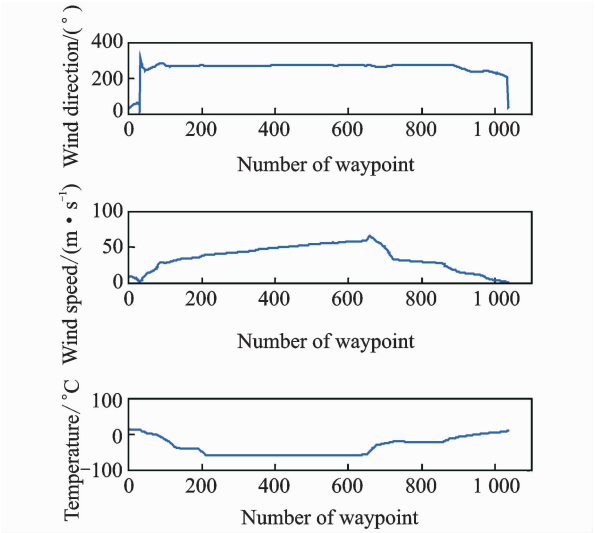


Fig. 9 Meteorological diagram based on IDW interpolation

In order to discuss the forecast error of GRIB data, the covariance matrix of wind-field forecast errors can be determined by using Eqs. (13)—(14), and the results are displayed in Fig. 10, which can meet requirements of interpolation.

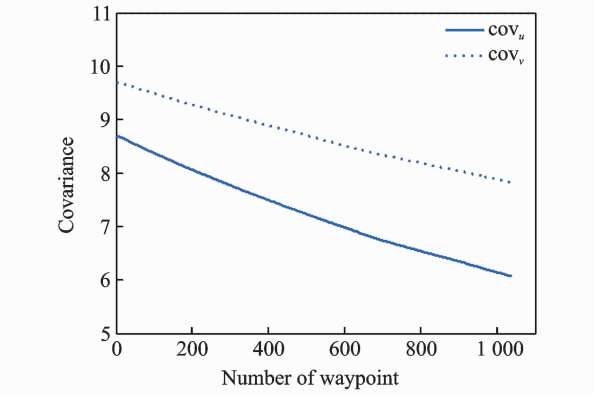


Fig. 10 Covariance of wind-field forecast from the nearest GRIB interpolation point

Finally, facilitating the flight speed triangle and CAS/TAS conversion model and the relationship between the aircraft's speed and altitude, we calculated the airspeed transition altitude as 6 880 m using Eqs. (19)—(21). Figs. 11—12 show the nominal airspeed profiles based on the demarcation point H_C, where the flight speed in the cruising stage was maintained at a Mach number of 0.82.

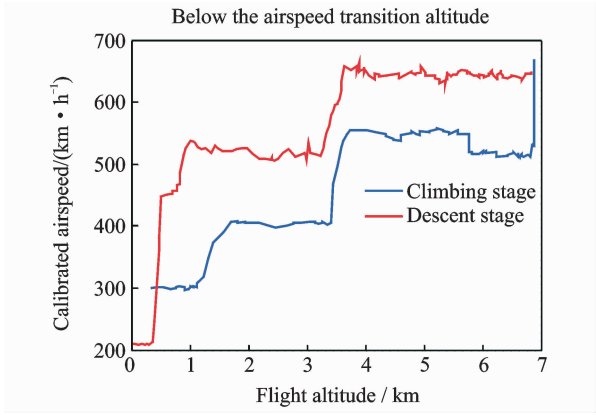


Fig. 11 Flight CAS profiles below the airspeed transition altitude

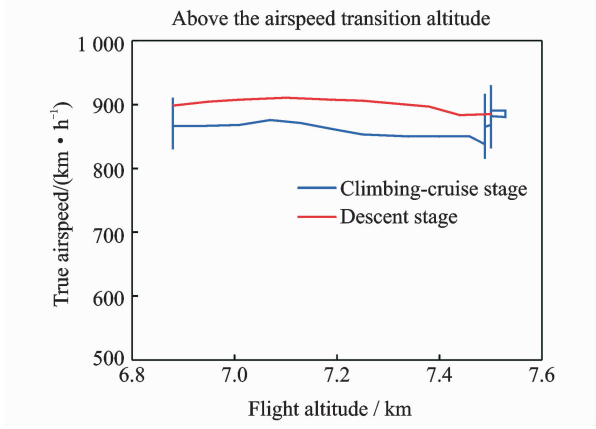


Fig. 12 Flight TAS profiles above the airspeed transition altitude

3.2 Result analysis

In order to verify the feasibility of the nominal flight profile, we estimated three trajectories using the flight profiles in clusters I, II, and III for the same flight, where the arrival times at given waypoints were determined and compared with the actual arrival times, which we obtained from SSR or Automatic Dependent Surveillance-Broadcast, as shown in Fig. 13.

Based on this comparison, we found that the arrival times obtained from the flight profiles in cluster II were the closest to the actual flight times, where the total flight time was 2.4 minutes slower than the actual time. In addition, we determined the maximum time error over the cumulative distance of 569.23 km, where the predicted time of arrival was 3.9 minutes later than

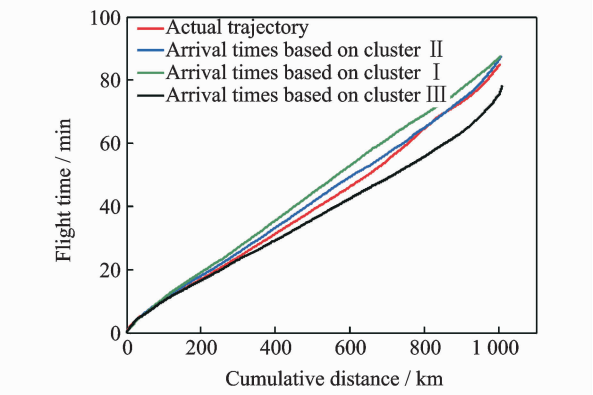


Fig. 13 Comparison of flight times based on 4D trajectory prediction

the actual time. The relative errors in different flight stages are shown in Table 4.

Table 4 Flight time in different flight stages

Flight stage	Flight time obtained from prediction/ min	Flight time obtained from actual SSR data/ min	Relative error/ %
Climbing	16.66	17.91	−6.98
Cruising	35.98	33.83	6.36
Descending	35.05	33.53	4.53

Finally, given the condition of a low real-time requirement before take-off, we found that the flight profile obtained by the improved algorithm proposed in this study could satisfy the requirements for trajectory prediction in the future.

4 Conclusions

In order to mine nominal flight profiles containing the intentions of controllers, we proposed an algorithm for generating altitude profiles based on an improved *K*-means clustering algorithm and an algorithm for generating speed profiles based on the BADA model using actual aircraft operation characteristics. Taking a flight from Xiamen Airport to Wuxi Airport as an example, we found that this method could reflect the flight intentions of controllers and eliminate the randomness in the trajectory caused by meteorological factors, thereby illustrate the effectiveness of the final results.

However, for the generated nominal flight profile, relative errors occurred for two reasons: (1) The airway meteorological data obtained from historical GRIB data using interpolation algorithm were not very accurate due to long refreshing cycle of the GRIB data. (2) During the actual aircraft operation, flight conflict resolution using various strategies, such as adjusting the airspeed and altitude, was considered in order to guarantee the flight safety, which influenced the accuracy of the 4D trajectory estimation compared with that obtained without considering conflicts.

In future research, we will focus on developing an airway meteorological revision model to amend the flight profile by improving the precision of airway meteorological data, as well as a strategic conflict-free 4D trajectory planning method for aircraft to guarantee flight safety, thereby obtaining more accurate 4D trajectory estimation.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Nos. 61174180, U1433125), the Jiangsu Province Science Foundation (No. BK20141413), and the Chinese Postdoctoral Science Foundation (No. 2014M550291).

References:

[1] HARRY S, RICHARD B, MICHAEL L. Next generation air transportation system (NGATS) air traffic management (ATM)—airspace project[R]. USA: NASA, 2006: 25-28.

[2] COPPENBARGER R A. Climb trajectory prediction enhancement using airline flight planning information [C]// Proceedings of the AIAA Guidance, Navigation, and Control Conference. Portand: AIAA, 1999: 1077-1087.

[3] GONG C, CHAN W N. Using flight manual data to derive aero-propulsive models for predicting aircraft trajectories[C]// AIAA Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical. California: AIAA, 2002: 1-7.

[4] WU Shufan, GUO Suofeng. Synthesis of aircraft vertical flight profile based on four-dimensional guidance in terminal airspace[J]. ACTA Aeronautical ET Astronautical Sinica, 1993, 14(5): 261-268. (in Chinese)

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

[6] LEIDEN K, ATKINS S. Trajectory clustering for

metroplex operations [C] // Proceedings of 11th AIAA Aviation Technology, Integration, and Operation (ATIO) Conference. USA: AIAA, 2011:7066.

[7] WANG Taobo, HUANG Baojun, Application of improved fuzzy clustering algorithm to analysis of trajectory[J]. China Safety Science Journal, 2013, 31(6):38-42. (in Chinese)

[8] WANG Chao, GUO JiuXia, SHEN Zhiping. Prediction of 4D trajectory based on basic flight models[J]. Journal of Southwest Jiaotong University, 2009, 44(2):295-300. (in Chinese)

[9] XING Jian, TANG Xinmin, Han Songchen, et al. Method for generating flight profile based on AMDAR data[J]. Journal of Nanjing University of Aeronautics and Astronautics, 2015, 47(1): 64-70. (in Chinese)

[10] TANG X M, XING J, CHEN P. Generating nominal flight profile for air traffic control system based on AMDAR data[C] // 2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC). USA: IEEE, 2014:2644-2649.

[11] DING Songbin. Flight performance and flight plan [M]. Beijing: Science Press, 2013. (in Chinese)

[12] MARTEAU P F. Time warp edit distance with stiffness adjustment for time series matching[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence archive, 2009,31(2):306-318.

[13] KEOGH E, RATANAMAHATANA C A. Exact indexing of dynamic time warping [J]. Knowledge and Information System, 2005:358-368.

[14] CHEN Q, HU G Y, GU F L, et al. Learning optimal warping window size of DTW for time series classification[C] // The 11th International Conference on Information Sciences, Signal Processing and their Applications, Special Sessions. [S. l.]: [s. n.], 2012:1272-1277.

[15] WANG S Y. An improved K-means clustering algorithm based on dissimilarity[C] // 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC). Shenyang, China: IEEE, 2013:20-22.

[16] MEHAR A M, MATAWIE K, MAEDER A. Determining an optimal value of K in K-means clustering [C] // 2013 IEEE International Conference on Bioinformatics and Biomedicine. USA: IEEE, 2013:52-55.

[17] SÖKMEN N. GRIB message resolution and accuracy analysis of upper-level wind[J]. Air Traffic Management, 2010, 190(1):51-63.

[18] WU Xiaoguang, ZHANG Junfeng, JIANG Haixiang. Application of GRIB data for 4D trajectory prediction [J]. Aeronautical Computing Technique, 2013,43(6):59-62. (in Chinese)

[19] EMRE O, SERDAR B, USTUNDAG B B, et al. Land surface temperature-based spatial interpolation using a modified inverse distance weighting method[C] // 2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics). [S. l.]:[s. n.],2013:110-115.

[20] DONG Zhinan, ZHEN shuanning, ZHAO Huibing. Comparative analysis of wind field simulation method based on spatial interpolation[J]. Journal of Geo-information Science, 2015,17(1):37-44. (in Chinese)

[21] LYMPEROPOULOS I, LYGEROS J. Sequential Monte Carlo methods for multi-aircraft trajectory prediction in air traffic management[J]. International Journal of Adaptive Control & Signal Processing, 2010,24(10):830-849.

[22] COLE R E, RICHARD C, KIM S, et al. An assessment of the 60 km rapid update cycle (RUC) with near real-time aircraft reports[R]. USA: NASA/A-1, MIT Lincoln Laboratory, 1998.

[23] EURO-CONTROL Experimental Centre. User Manual for the Base of Aircraft Data [R]. Germany: [s. n.], 2010.

[24] EURO-CONTROL Experimental Centre. Base of aircraft data (BADA) aircraft performance modelling report[R]. Germany: [s. n.], 2009.

Prof. **Tang Xinmin** is currently a professor of College of Civil Aviation in Nanjing University of Aeronautics and Astronautics. His research interests include (1) Petri Net and Discrete Event Dynamic System theory; (2) Hybrid System Theory. (3) Intelligent Air Traffic System.

Mr. **Gu Junwei** is currently a postgraduate of College of Civil Aviation in Nanjing University of Aeronautics and Astronautics. His research interest is next-generation Air Traffic Control Automation System.

Dr. **Shen Zhiyuan** is currently an assistant professor of college of civil aviation at Nanjing University of Aeronautics and Astronautics (NUAA). His research interests include ADS—B technique, 4D trajectory prediction.

Mr. **Chen Ping** is currently a chief expert in the 28th Research Institute of China Electronic Technology Group Corporation, and academic leader in the National Air Traffic Control Engineering Center. His research interests include (1) Airspace Management; (2) Traffic Management; (3) Data Fusion Disposal; (4) Intelligent Aid Decision Making System.

Mr. **Li Bo** is currently a graduate of College of Civil Aviation in Nanjing University of Aeronautics and Astronautics. His research interest is Low Airspace Air Traffic Management.

(Executive Editor: Zhang Bei)