

Research Article

Aircraft Trajectory Prediction Based on Bayesian Optimised Temporal Convolutional Network–Bidirectional Gated Recurrent Unit Hybrid Neural Network

Jin Huang and Weijie Ding 

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

Correspondence should be addressed to Weijie Ding; dingweijie163@163.com

Received 18 October 2022; Revised 26 November 2022; Accepted 28 November 2022; Published 15 December 2022

Academic Editor: Kenneth M. Sobel

Copyright © 2022 Jin Huang and Weijie Ding. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Efficient and accurate flight trajectory prediction is a key technology for promoting intelligent and informative air traffic management and improving the operational capabilities and predictability of air traffic. To address the problems in extracting hidden information from historical trajectory information, the approach must accurately select high-dimensional features related to the prediction target and overcome the short-term memory of the time series. Herein, we present a novel trajectory prediction model based on a dual-self-attentive (DSA)-temporal convolutional network (TCN)-bidirectional gated recurrent unit (BiGRU) neural network. In this model, the TCN provides highly stable training, high parallelism, and a flexible perceptual domain. The self-attentive mechanism of the TCN structure can focus on features that contribute the most to the output. After the TCN, the BiGRU network combined with the self-attentive mechanism is used to further bidirectionally mine the connections between the features and outputs of the trajectory sequence, and a Bayesian algorithm is used to optimise the hyperparameters of the model for optimal performance. A comparison and validation based on current well-known neural network models (i.e., CNN, TCN, GRU, and their variants) shows that the DSA-TCN-BiGRU model based on Bayesian hyperparameter optimisation has the best performance. Therefore, the improved predictive model is applicable and valuable, providing a basis for future decision trajectory-based operations.

1. Introduction

The continuous growth of air traffic has raised critical challenges to current air traffic control (ATC) systems, and the international civil aviation community is undergoing a new round of ATC system upgrades. In this new generation of ATC automation systems, a trajectory-based operation mode has been proposed as the core technology for the next phase [1]. In response to the global ATC integration requirements proposed by the International Civil Aviation Organisation [2, 3], the United States and Europe have taken the lead in developing future ATC plans including the Single European Sky Program, U.S. Central Terminal Area Control Automation System, and Distributed Ground and Air Traffic Management System. Together, they support the operational concepts for trajectories and performance, forming a

trajectory-based ATC system for providing flight efficiency, energy savings, and capacity assessments. The key technical basis of the trajectory-based operation concept is accurate trajectory prediction [4], as such predictions provide the most direct reference for aircraft surveillance and control prognosis. In addition, the management of airspace is changing from static to dynamic, and the use of automatic dependent surveillance-broadcast (ADS-B) systems for real-time surveillance. The need for more detailed airspace management has become even more pressing in light of airspace delineation and use characteristics. Therefore, the next-generation ATC system should not only update the concept of operation, popularise and standardise ADS-B surveillance technology but also strengthen the independent research and development of critical technologies to achieve more automated, intelligent, accurate and safe and efficient

air traffic management. There is a clear, practical need, and essential application value for real-time and effective track prediction through ADS-B data.

Thus, trajectory prediction is an important technology for new air traffic management automation systems, and the importance of trajectory prediction technology has made it a research topic of considerable interest in the context of ATC systems in recent years. Research into trajectory prediction technology has led to the proposal of many models for the prediction task. They can be categorised as follows.

Kinetic model: a kinematic model must consider the performance parameters of the aircraft, stage the flight process, and establish the corresponding motion equations. Fan et al. [5] proposed a multiobjective constraint-based aircraft descent segment track prediction method for improving the accuracy of aircraft prediction tracks by considering aircraft performance data, characteristic parameters, flight path restrictions, and cost index requirements. Li [6] combined aircraft energy states to simplify the equations of motion and obtain trajectory characteristics at minimum cost, and then used an equirectangular trajectory computational model to fit the flight parameters with the great likelihood estimation method to achieve a cost-optimised 4D trajectory prediction. Du et al. [7] constructed a prediction method for the vertical profile of a continuous aircraft climb while considering thrust intent using the aircraft full energy equation and considering the wind speed vector and temperature information. Pan et al. [8] used a hidden Markov model (which can predict the future position of an aircraft being flown every minute in the future), and the validity of the horizontal and vertical errors was significantly improved relative to the baseline model. However, the existence of idealised prerequisite assumptions for building a kinematic discrimination model limits its application scenarios and prediction accuracy

Data mining models: data mining-based trajectory prediction techniques have become popular as a ‘big data’ research subject in ATC. Ma and Gao [9] clustered historical trajectories generated by a certain route with typical trajectories through a clustering algorithm, and then performed a trajectory correction on the position information at a certain moment to predict the position information at the next moment. Zazzaro et al. [10] integrated a clustering algorithm, classification-based supervised learning, and an uncertainty model to calculate the probability of collision using Opensky public data, and implemented conflict risk detection in the terminal area. Carlos et al. [11] obtained more variable features by preprocessing the trajectory coordinates and flight plan data, and applied them to a k-means clustering algorithm (to the extent possible) to support a supervised trajectory classification. Choi et al. [12] combined a data-driven and physics-based state estimation model and compared it with a machine learning-only approach, substantially improving accuracy. However, algorithms based on state estimation can produce significant errors when faced with long-term forecasts as they need to consider the long-term movement features of the model capture model over time. Ma [13] proposed classifying the trajectories of

different flights using a generalised cluster analysis method combined with the hidden Markov model to obtain a high-density 4D trajectory prediction. In general, a data mining approach can improve the prediction accuracy using different clustering algorithms, but it also generates problems concerning data storage and computational overhead to different degrees

Deep learning models: the development of neural networks in deep learning has provided new ideas for nonlinear models, especially those good at managing time series. Thus, such networks have been implemented in the study of trajectory prediction [14–16]. Zhou et al. [17] reconstructed and combined the predictive capabilities of multiple neural networks over different time spans to improve prediction accuracy. Sahadevan et al. [18] extracted data through sliding windows and improved the prediction accuracy using a bidirectional long short-term memory (BiLSTM) model to understand the dependencies of the trajectory data in two directions, which, in turn, improved the ground system trajectory prediction accuracy. However, the proposed model has yet to consider the spatial feature extraction of the data. Tran et al. [19] provided a reliable model for a conflict detection system by combining encoder–decoder modelling with tactical intent, and verified that the prediction accuracy exceeded those of existing models using actual data. In 2020, Lv et al. [20] proposed using a temporal convolutional network (TCN) for trajectory prediction, and compared it with a conventional model. They found that the short-term effect was comparable and the long-term prediction was improved. Currently, gated recurrent unit (GRU) neural networks are primarily used for trajectory predictions [21–23]. Wu et al. [24] proposed to combine GAN networks with Conv1D, Conv2D, and LSTM networks and verified that Conv1D-GAN could achieve accurate long-term trajectory prediction with actual data. The great promise of combined neural networks in the field of trajectory prediction is shown. Shafiya and Regan [25] proposed a CG3D model that fully extracts ADS-B data’s spatial and temporal features by combining CNN-GRU and 3D-CNN networks. The uncertainty of the model is also considered, and the combined model has higher prediction accuracy than a single model. Chen et al. [26] proposed a combination of inception and LSTM modules for prediction, effectively extracting the data’s spatio-temporal features. However, the model mainly considers the reliability condition of localization and the error is significant. Han et al. [27] used k-means to cluster similar trajectories, followed by online prediction via GRU to achieve real-time prediction of trajectories. Ma et al. [28] proposed a CNN combined with an LSTM network, which achieved good performance in short-term prediction but still had a significant error in multistep prediction as no attention mechanism was used

To achieve more accurate short- and long-term forecasts and take into account computational efficiency, we adopted a novel TCN-bidirectional GRU (BiGRU) model based on an attention mechanism, with the following innovations.

- (a) In this study, the TCN model is fused with the BiGRU model, and a dual-self-attentive (DSA)

mechanism is introduced. The DSA-TCN-BiGRU model can extract multisource time-series features and maintain causal convolutional properties. High-order trajectory features are extracted by introducing a dilation convolution to improve the processing efficiency of long-time-span memory units. In addition, the model uses residual linking to transfer the underlying complete features across layers, thereby enriching the feature results while optimising the overall network-learning process. Subsequently, the BiGRU model combined with the DSA mechanism can bidirectionally extract the impact of each time node in the hidden layer state on the prediction results. The important problems concerning the high-dimensional feature extraction and long-term dependence of the time series are effectively solved. To the best of our knowledge, this is the first time that a DSA-TCN-BiGRU model has been used to forecast aircraft trajectories

- (b) The mapping from points in the hyperparameter space to the model generalisation performance can be viewed as a complex black-box function with a high evaluation cost; it is difficult to apply with general optimisation methods. We used a Bayesian Optimisation (BO) algorithm to optimise the hyperparameters of the whole model to improve the model prediction performance and training speed, thereby overcoming the disadvantages concerning the time-consumption and low accuracy in manual parameter tuning
- (c) To verify the performance of the model in terms of track predictions, three sets of comparisons were made. Particularly, the TCN and DSA mechanisms were investigated to ensure that the important track history information was fully utilised in the extraction of the track data, and that the sequential nature of the tracks was preserved to compensate for the shortcomings of the convolutional neural network (CNN). By comparing the BiGRU model with the GRU and LSTM, the bidirectional temporal feature extraction capability of the BiGRU model was verified, and the time complexity was simultaneously reduced. In addition, we tested the model for single-step and multistep prediction, and the results verified that the DSA-TCN-BiGRU approach has high stability and track prediction capability

The remainder of this paper is organised as follows. Section 2 analyses the data and preprocessing. Section 3 presents the DSA-TCN-BiGRU model. Section 4 presents the experimental and comparative validation results. The final section concludes this paper.

2. Trajectory Description and data Analysis

As a new technology promoted by the Civil Aviation Administration of China, the ADS-B technology gener-

ates massive amounts of data that can hide important aircraft flight information [29]. The dynamic data (generated in real time) implies the future movement trend of the aircraft.

2.1. ADS-B Trajectory data Format. An ADS-B message returns the trajectory information of each vehicle at a certain point during flight as a series of discrete trajectory points aggregated into discontinuous trajectory data for each vehicle. The trajectory is represented as follows:

$$D = \{D_1, D_2, \dots, D_i, \dots, D_n\}. \quad (1)$$

Here, D denotes the set of aircraft trajectories, i denotes the trajectory number, $i \in [1, n]$, $n \in \mathbb{N}$, n denotes the total number of trajectories, and D_i denotes the flight path of flight i .

A trajectory consists of a series of multidimensional trajectory points ordered by time.

$$D_i = \{p_{i1}, p_{i2}, \dots, p_{ij}, \dots, p_{im}\}. \quad (2)$$

Here, p_{ij} denotes the j multidimensional trajectory point of trajectory i and m denotes the number of trajectory points.

The information contained in each waypoint is expressed as follows:

$$p_{ij} = \{lat_t, lon_t, alt_t, v_t, h_t\}. \quad (3)$$

Here, lat_t denotes the latitude of the aircraft at time t ; lon_t denotes the longitude of the aircraft at time t ; alt_t denotes the altitude of the aircraft at time t ; v_t denotes the speed of the aircraft at time t ; and h_t denotes the heading of the aircraft at time t .

2.2. Data Preprocessing. The ADS-B data are often incomplete and inaccurate owing to receiving terminal failures, communication link delays, and navigation signal losses. Accordingly, in this study, data with duplicate 3D position points and time points are deleted to regenerate the trajectory sequence. Missing points were handled by recalculating the trajectory sequence points in 5-s units using three-spline interpolation, as shown in Figure 1.

In Figure 1, the red colour represents the trajectory information after the ADS-B data are parsed. The basic route characteristics can be obtained, but there are missing trajectory points in the middle; if there are trajectory missing points at the edge of the time period when the feature extraction is conducted, the statistics will have certain errors, and the validity of the experiment will be questioned. Therefore, all aircraft trajectory information is interpolated (as shown in black).

When the amount of data is small, the prediction accuracy may be poor owing to overfitting during the deep learning training process. We construct and expand features by adding the distance feature G and angle feature δ between the trajectory point (x_i, y_i, z_i) and centre reference point

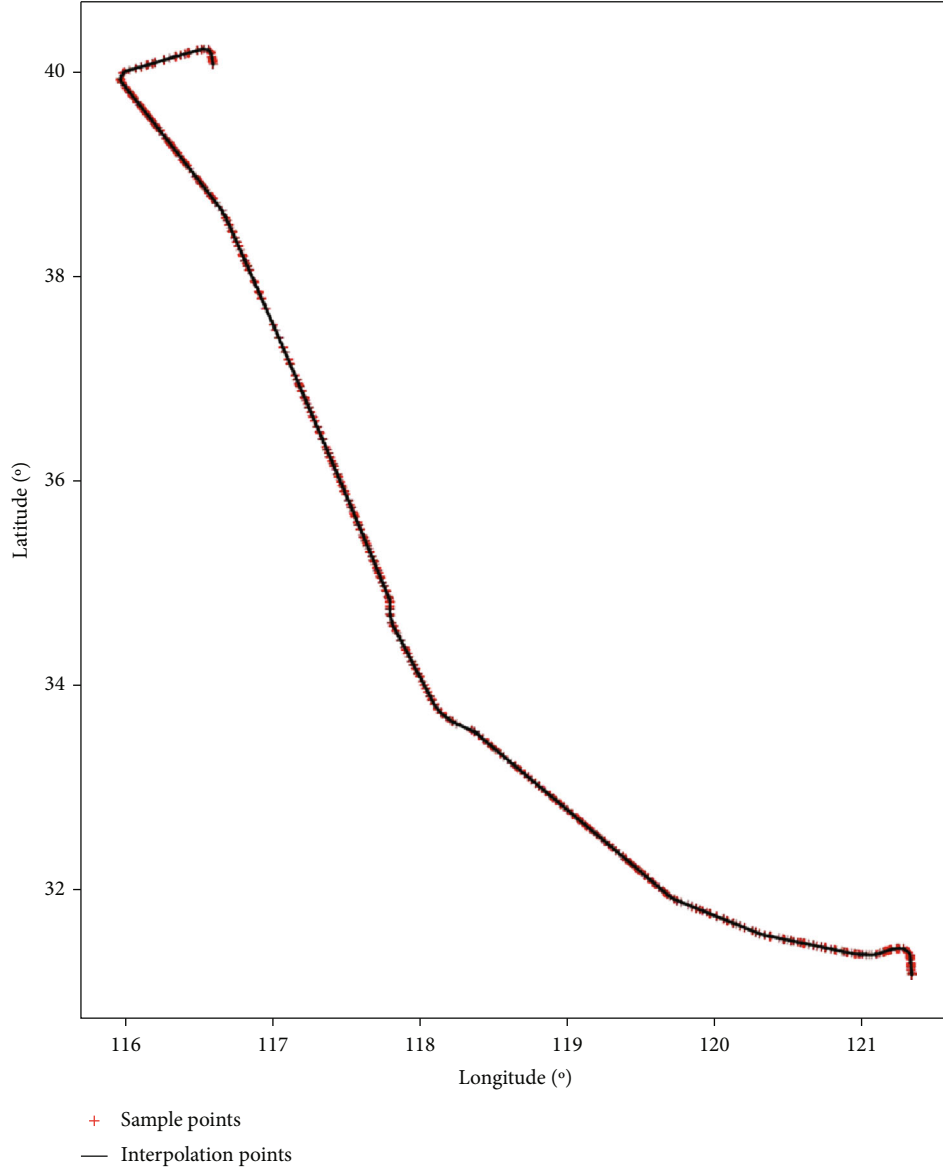


FIGURE 1: Three-spline interpolation.

(x_o, y_o, z_o) of the destination to reflect the pilot's intention, as shown in

$$G = \sqrt{(x_i - x_o)^2 + (y_i - y_o)^2 + (z_i - z_o)^2}, \quad (4)$$

$$\delta = \arctan \left(\frac{y_i - y_o}{x_i - x_o} \right). \quad (5)$$

Normalisation: the data are processed using normalisation, thereby solving the problem of the nonuniformity of the input data.

$$X = \frac{x - x_{\max}}{x - x_{\min}}. \quad (6)$$

Here, X is the normalised input data, and x is the original input data.

2.3. Sample Construction. In the sample construction, we select the data [30] using sliding windows with a sliding window size of six to achieve single- and multistep predictions, as shown in Figure 2.

3. Methods

3.1. Self-Attentive Mechanism. When using neural networks to process large amounts of sequence data, one can borrow from the attention mechanism of the human brain and selectively process certain primary information while skipping secondary information to improve the network performance. Sequence coding based on convolutional or recurrent neural networks can only establish local dependencies. One way to establish long-distance dependencies between the sequences is to increase the number of layers in the network to obtain long-distance information interactions through deeper networks. The other is to use full connectivity, which, although

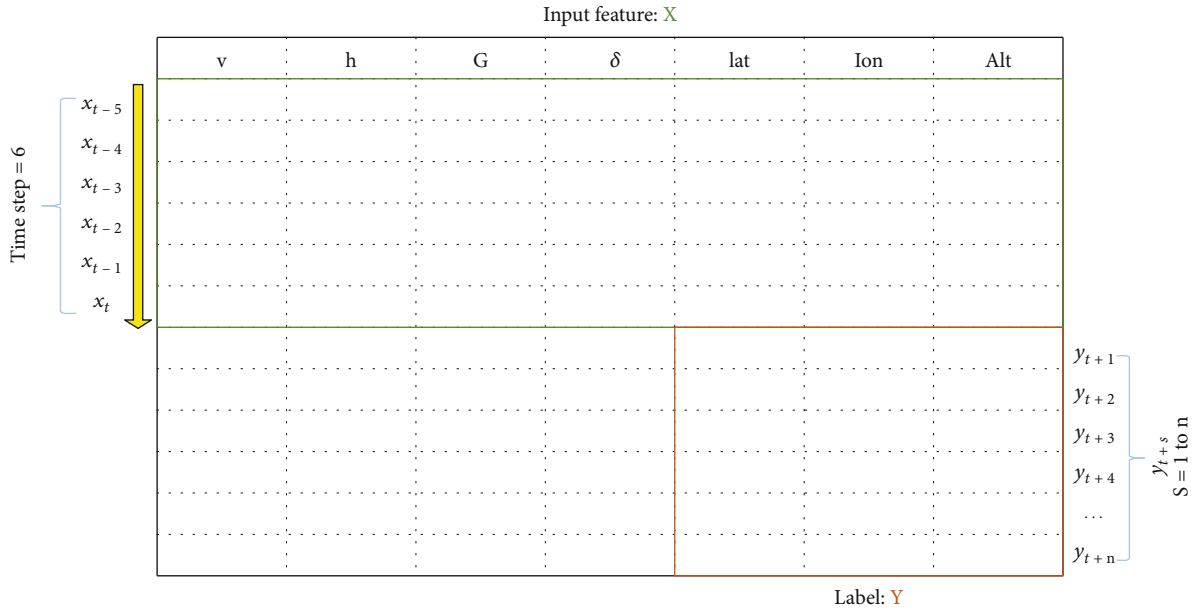


FIGURE 2: Sample division.

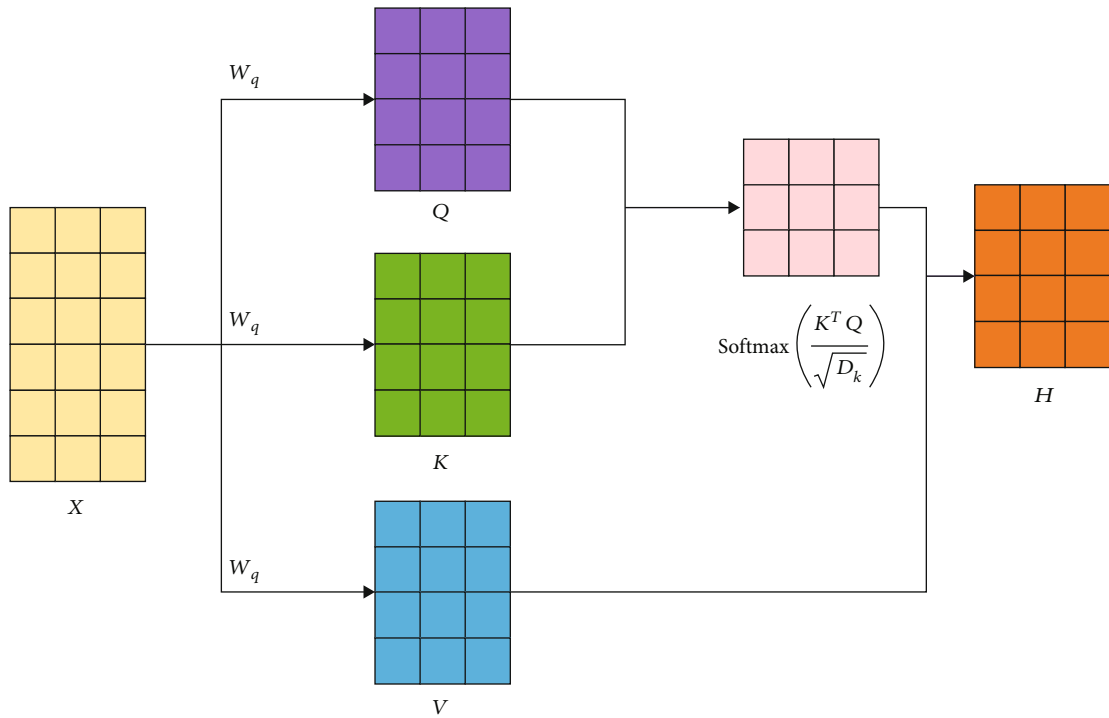


FIGURE 3: Self-attentive model calculation process.

capable of establishing long-distance dependencies, cannot handle variable-length input sequences. A self-attention model [31] can dynamically generate weights for handling variable-length sequences of information and can establish long-range dependencies.

The self-attention model typically uses a key-value pair model. The similarity function uses the scaled dot product as the attention scoring function, as shown in Figure 3. Assuming that the input sequence $X = [X_1, \dots, X_n]$ and out-

put sequence $H = [h_1, \dots, h_n]$, each input x_i is linearly mapped to three different spaces to obtain the query vector q_i , key vector k_i , and value vector v_i , respectively. The mapping is calculated as shown in

$$Q = W_q X \tag{7}$$

$$K = W_k X \tag{8}$$

$$V = W_v X \tag{9}$$

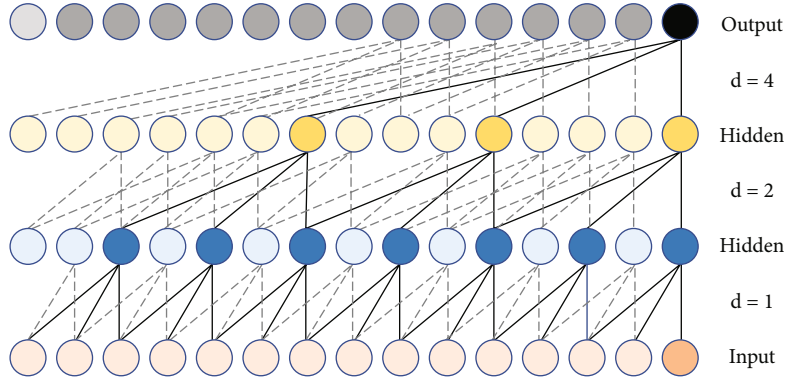


FIGURE 4: Causal dilated convolution structure.

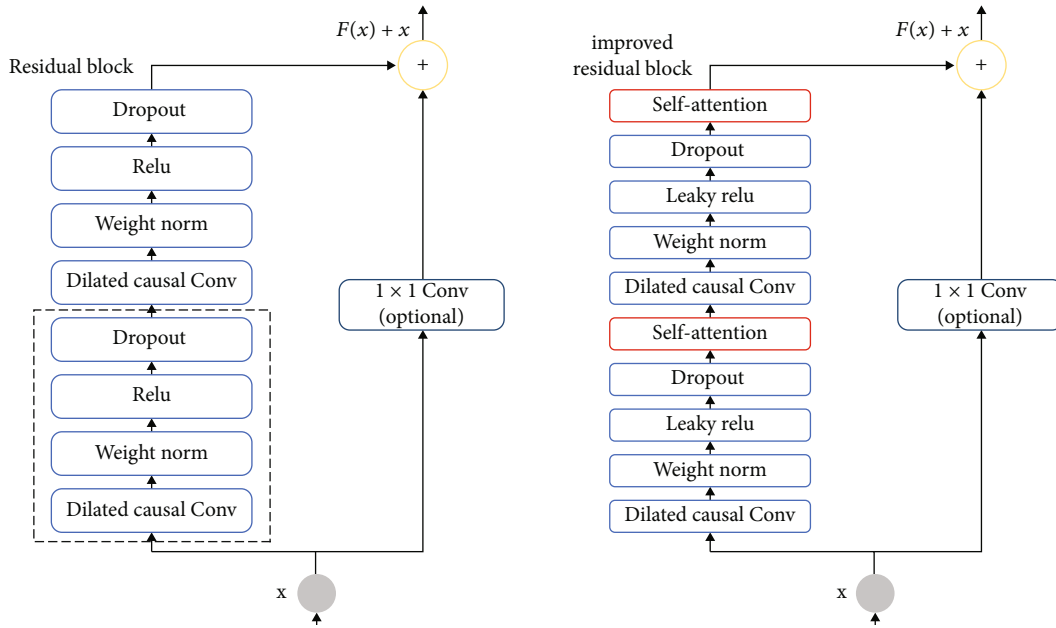


FIGURE 5: Architectures of temporal convolutional network (TCN) and self-attentive TCN (SATCN).

In these equations, W_q , W_k , and W_v are the linear mapping weight matrices, and Q , K , and V are the query, key, and value matrices, respectively.

When scaled clicks are used as the attention scoring function, the output vector sequence is calculated as follows:

$$H = V \text{softmax} \left(\frac{K^T Q}{\sqrt{D_k}} \right). \quad (10)$$

Here, D_k is the matrix dimension, and softmax is a function normalised by the column.

Self-attention models can effectively handle variable-length time series and establish long-range dependencies [32].

3.2. Self-Attention Temporal Convolutional network. The TCN networks [33] applied in trajectory prediction have the following main characteristics: (i) to meet the requirements for trajectory prediction, TCN networks have temporal causality, that is, the current model output is only related

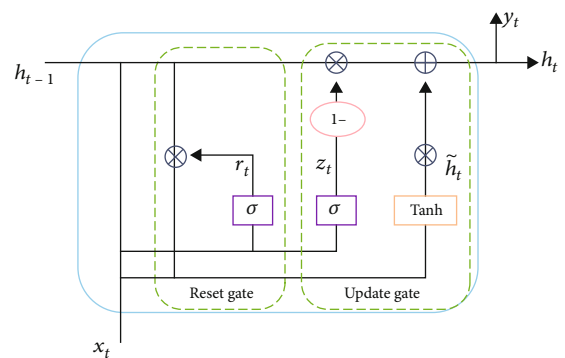


FIGURE 6: Structure of a standard gated recurrent unit (GRU).

to the past and not to the future; (ii) the output sequence length of TCN networks applicable to trajectories can be adjusted arbitrarily; and (iii) multilayer TCN networks have long memory distances despite their shallow depths.

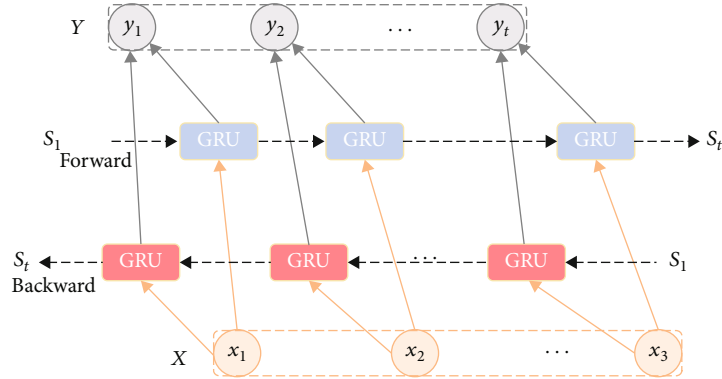


FIGURE 7: Structure of the bidirectional GRU (BiGRU) model.

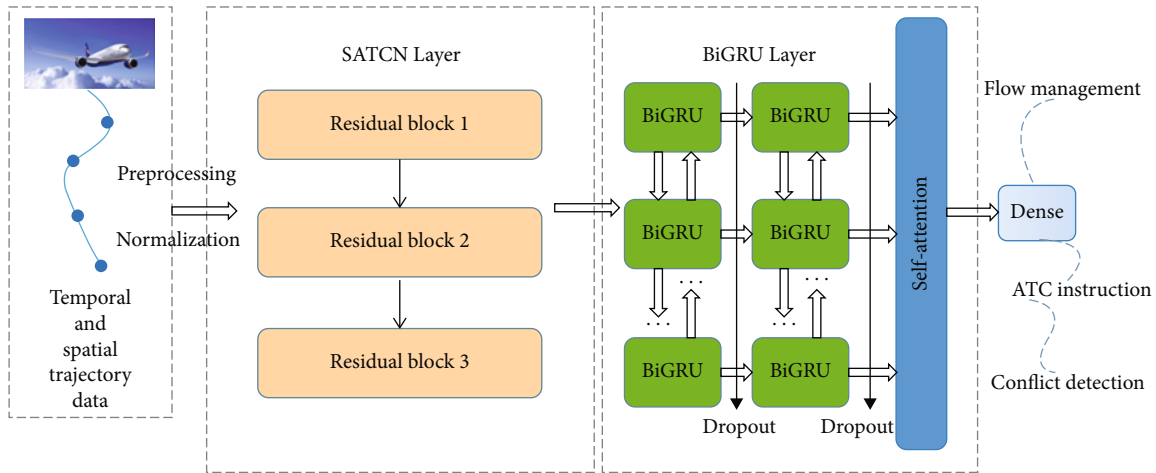


FIGURE 8: DSA-TCN-BiGRU trajectory prediction model.

3.2.1. Causal Dilated Convolution Network. The TCN causal dilation convolution structure is shown in Figure 4. The most unique feature of the TCN dilation convolutional structure is that its specific form is determined by the convolutional kernel and dilation coefficients. The convolution kernel and dilation coefficients change the form of the TCN in terms of the number of input elements and distance from the upper layers of the network, respectively. The expansion factor P increases exponentially with the depth of the network, i.e., $P = 2^r$, where r is the number of network layers. The larger the expansion factor, the larger is the information extraction range. Therefore, the TCN convolution can obtain a relatively large perceptual field by building a relatively small number of layers.

The receptive field of a TCN network neuron, i.e. the network memory length, is determined by the convolutional kernel size, dilation coefficient, and number of convolutional layers. The value of the s operation after the dilation convolution operation is as follows:

$$F(s) = (x * f)(s) = \sum_{i=0}^{u-1} f(i) \cdot x_{s-di} \quad (11)$$

Here, x is the input sequence, $*$ is the convolution operation, u is the convolution kernel size, $f(i)$ is the i th element

in the convolution kernel, and x_{s-di} is the element of the input sequence corresponding to the convolution kernel operation.

3.2.2. Residual Links. In the residual block [34], the output of a multilayer network F is added to the original input x and output through the activation function G . The computation is shown in

$$y = G(x + F(x)). \quad (12)$$

The self-attentive TCN (SATCN) residual module contains the underlying TCN causal expansion convolution layer, batch normalisation, activation function (LeakyReLU) and dropout layers. The N residual module structure is shown in Figure 5.

The normalisation of weights can eliminate the gradient explosion problem and effectively speed up the computation. To make the TCN network nonlinear while avoiding gradient disappearance, the LeakyReLU activation function is used and the dropout layer is added after the LeakyReLU activation layer to prevent overfitting. Accordingly, it is possible to achieve the regularisation effect. The self-attentive mechanism is introduced to focus on features that contribute more to the output, and adjusts the problem of the

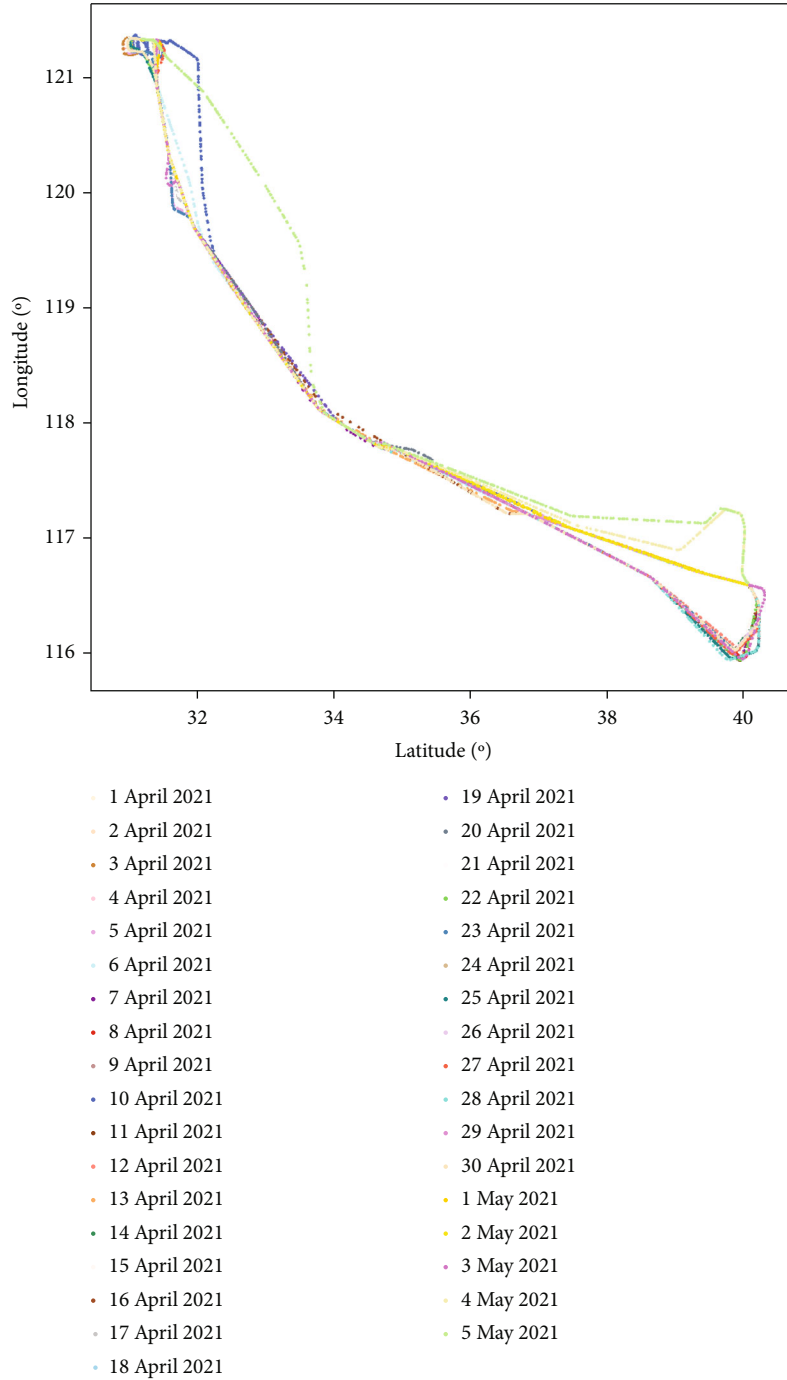


FIGURE 9: Lateral trajectory profile.

different dimensions of the residual tensor using 1×1 convolution.

3.3. Bidirectional Gated Recurrent Unit network. The GRU model [35] is a type of recurrent neural network (RNN); compared with the LSTM, GRU has only two gating structures, i.e., the reset gate and update gate. Thus, its structure is simpler, and comparable to its effect [36]. The structure of the GRU is shown in Figure 6. Equations (13)–(16)

describe the operation of each cell in the GRU network.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \quad (13)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]), \quad (14)$$

$$\tilde{h}_t = \tan h(W_{h_t} \cdot [r_t * h_{t-1}, x_t]), \quad (15)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t. \quad (16)$$

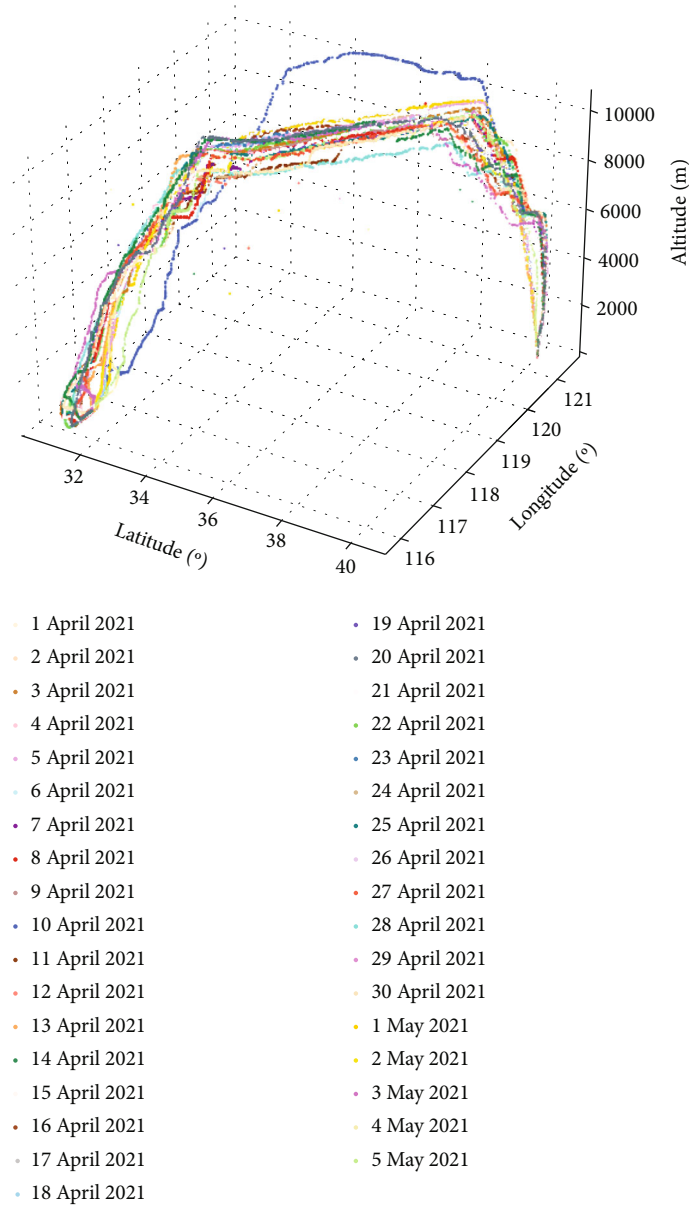


FIGURE 10: Vertical trajectory profile.

Here, x_t is the current moment input, h_{t-1} denotes the previous moment hidden state, h_t is the pass to the next moment hidden state, \tilde{h}_t denotes the candidate hidden state, r_t is the reset gate, z_t is the update gate, and σ is the sigmoid activation function.

The BiGRU [37] network is a two-layer GRU network consisting of a combination of forward and reverse GRU layers as shown in Figure 7.

3.4. DSA-TCN-BiGRU model. The insufficient extraction of the spatial and temporal features of trajectory data can cause complexity and difficulties in trajectory predictions. Accordingly, we propose a DSA-TCN-BiGRU trajectory prediction model, as shown in Figure 8. The model uses the SATCN

layer as a feature optimiser, as it can effectively extract hidden spatial feature information and temporal relationships and reduce redundant features. A memory function can be implemented relative to a simple CNN, and the different paths and sequence times for backpropagation enable it to avoid the gradient problem. In addition, if there is an excessive amount of data, it can be processed in parallel, thereby reducing the training time. The BiGRU model can perform bidirectional time-series feature extraction to further improve the global integrity of the time-series feature extraction. The self-attention mechanism can handle longer time-series data, strengthen long-range dependencies, and adjust the dynamic weights of the acquired serialised information to highlight the global key points of the trajectory information.

4. Experiments and Results

4.1. Dataset. The data preprocessing and training prediction of the network were based on the Tensorflow version 2.0 deep learning framework. The experimental dataset for this study was selected from 12 months of historical ADS-B one flight trajectory data from Beijing Daxing Airport to Shanghai Hongqiao Airport, with an average flight time of approximately two hours. Part of the trajectories was profiled laterally and vertically, as shown in Figures 9–10. When compiling the model, 80% of the data is used as the training set, and the rest is used for model testing.

4.2. Evaluation Metrics. To quantitatively evaluate the performance of the proposed DSA-TCN-BiGRU network prediction model, the root mean square error (RMSE) and mean absolute error (MAE) were used as error evaluation metrics. The equations are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i|, \quad (17)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2}.$$

Here, y_i and \tilde{y}_i represent the true and predicted values, respectively. Both indicate the stability and accuracy of the model; the smaller the value, the higher is the accuracy of the model.

4.3. Determination of model Parameters

4.3.1. Bayesian Hyperparameter Optimisation. Bayesian optimisation (BO) [38] uses Bayes' theorem to obtain most of the evaluation information from the previous functions, and then selects the next most promising sampling point based on the posterior distribution of the objective function. It can find the global optimal solution of a function with less evaluation and can also achieve good results for black-box functions, making it suitable for the hyperparametric optimisation problems of deep learning models [39]. The hyperparametric optimisation process is shown in Figure 11.

- (1) Parameter optimisation: the neural network parameters to be optimised include the residual block number, convolution kernel number and step size of the one-layer TCN, the number of neurons of the two-layer BiGRU, the number of neurons of the Dense, as well as the dropout rate, learning rate, batch size, and time step size

- (2) Objection function: MSE. Number of iterations: 100

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2. \quad (18)$$

y_i and \tilde{y}_i represent the true and predicted values, respectively.

- (3) Bayesian hyperparametric optimisation of the optimal range: $1 \leq n \leq 5$, $16 \leq t \leq 128$, $1 \leq k \leq 3$, $1 \leq u_1$

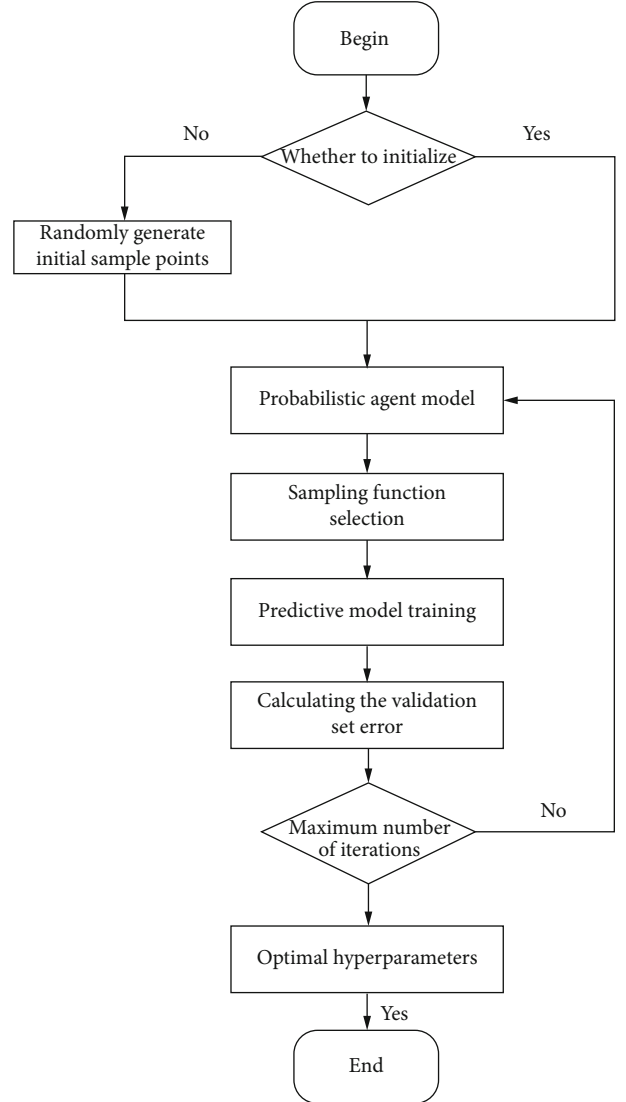


FIGURE 11: Bayesian hyperparameter optimisation flowchart.

$$\leq 100, \quad 1 \leq u_2 \leq 100, \quad 10 \leq e \leq 40, \quad 0.1 \leq d \leq 0.3, \\ 0.001 \leq l \leq 0.01, \quad 128 \leq b \leq 512, \quad \text{and} \quad 1 \leq s \leq 8$$

In the aforementioned range, n is the number of TCN residual blocks, t is the numbers of TCN residual block convolution kernel, k are the TCN residual block convolution kernel step, u_1 and u_2 are the numbers of BiGRU neurons, e is the numbers of Dense neurons, d is the dropout rate, l is the learning rate, and b is the number of batch processes, and s is the time step size.

- (4) BO parameters: probabilistic agent model uses the Gaussian Process (GP); Acquisition function uses the Upper Confidence Bound (UCB); Number of iterations: 100

According to Figure 12, the loss function of the DSA-TCN-BiGRU model tends to zero after the hyperparameter search by BO, confirming the effectiveness of the hyperparameter search using BO. The final model parameters are listed in Table 1.

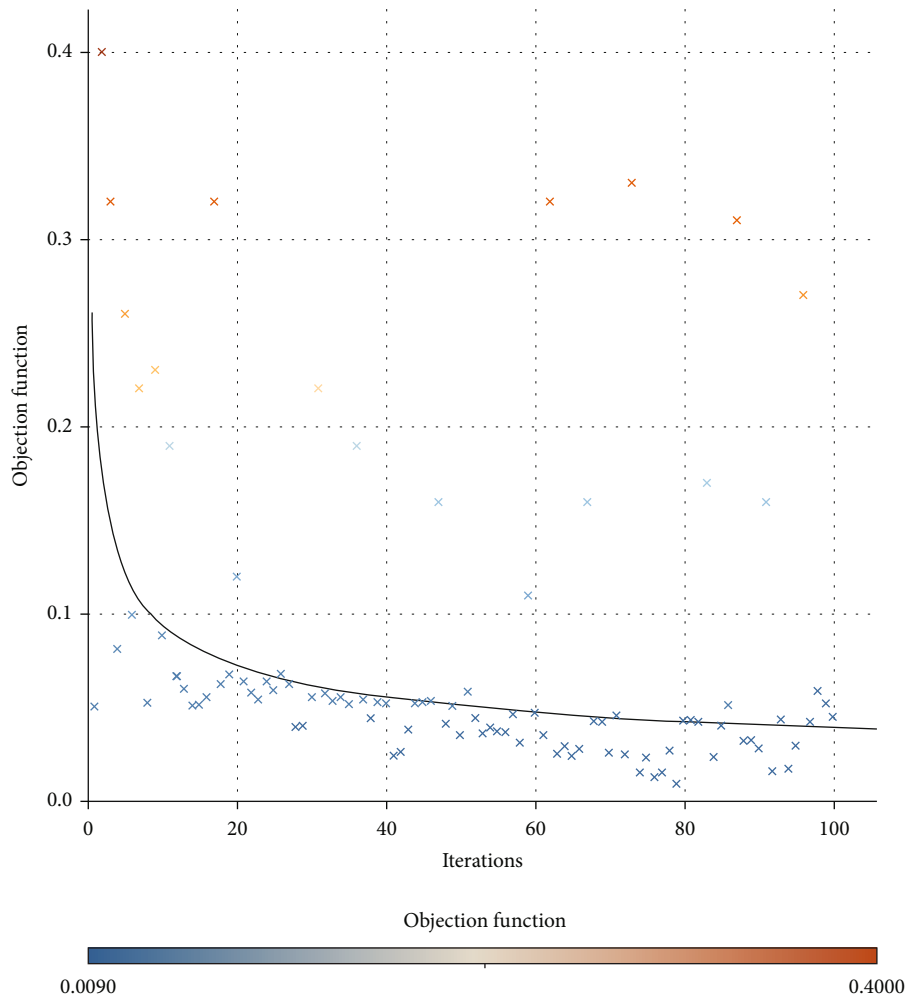


FIGURE 12: Bayesian hyperparametric optimisation search.

TABLE 1: Parameter settings.

Model			
DSA-TCN-BiGRU	TCN	Residual block 1	Filter = 128 Kernel size = 3 Dilation rate = 4
		Residual block 2	Filter = 64 Kernel size = 3 Dilation rate = 2
		Residual block 3	Filter = 32 Kernel size = 3 Dilation rate = 1
DSA-TCN-BiGRU	BiGRU	Units 1	Units 90
		Dropout	0.2561
		Units 2	Units 90
		Dropout	0.2561
		Dense	30
		Output	3
			Epochs = 100 Batchsize = 512 Learning rate = 0.001514

4.3.2. *Optimiser Selection.* The optimiser directly affects the speed of parameter optimisation and model accuracy, and the selection of a suitable optimiser can significantly

improve the model training efficiency. In this study, the sgd, rmsprop, and Adam optimisers were selected for comparison experiments, and the other parameters were

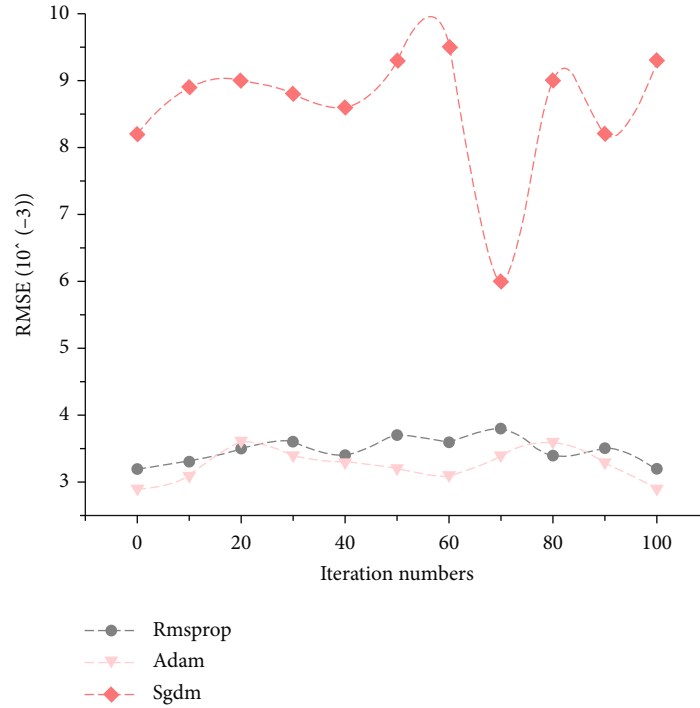


FIGURE 13: Visualisation of the impact of optimiser selection on model performance.

TABLE 2: Comparison of trajectory prediction errors.

	Hybrid model					Single model		
	DSA-TCN-BiGRU	TCN-BiGRU	DSA-CNN-BiGRU	CNN-BiGRU	SATCN	TCN	BiGRU	GRU
RMSE								
Alt/(m)	20.14	28.29	52.72	85.12	39.83	59.52	191.42	425.20
Lat/(°)	0.004	0.008	0.032	0.056	0.020	0.031	0.088	0.185
Lon/(°)	0.009	0.014	0.023	0.063	0.025	0.032	0.068	0.079
MAE								
Alt/(m)	14.64	18.50	34.74	70.33	23.20	44.64	153.11	325.13
Lat/(°)	0.003	0.006	0.028	0.042	0.018	0.026	0.071	0.116
Lon/(°)	0.009	0.008	0.019	0.053	0.021	0.029	0.059	0.052

kept the same as those mentioned earlier for the model training. The RMSE values of the validation set under the different optimisers are shown in Figure 13. We can see that both the rmsprop and Adam optimisers have high stability and accuracy. The accuracy of the rmsprop optimiser is approximately the same as that of the Adam optimiser; however, the training time consumed by the Adam optimiser is much lower than that of the rmsprop optimiser. The Adam optimiser was chosen based on a comprehensive consideration of the model accuracy and time efficiency.

4.4. Experiments and Comparisons. To verify the accuracy of the DSA-TCN-BiGRU network for single-step prediction, the same dataset and simulation parameters were used for comparison with existing trajectory prediction models (BiGRU, GRU, TCN, SATCN, CNN-BiGRU, DSA-CNN-BiGRU, TCN-BiGRU, and DSA-TCN-BiGRU). The prediction results for each model are shown in Table 2 and

Figure 14. To further illustrate the prediction accuracy of the proposed models, 2D and 3D zoomed-in plots are shown in Figures 15–16.

The aforementioned prediction results show that the DSA-TCN-BiGRU based on the BO model performs very well and shows the actual flight of the air vehicle in its entirety. The predicted 2D and 3D trajectory curves of the model are in good agreement with the actual flight trajectory of the air vehicle. The altitude, latitude, and altitude RMSE reach a minimum of 20.14 m, 0.004°, and 0.009°, respectively. In summary, the proposed model can fully extract the spatiotemporal characteristics of the trajectory data with a better generalisation effect and lower prediction error.

4.5. Performance Analysis

4.5.1. Self-Attentive Mechanism and TCN Performance. To compare the different effects of the self-attentive mechanism

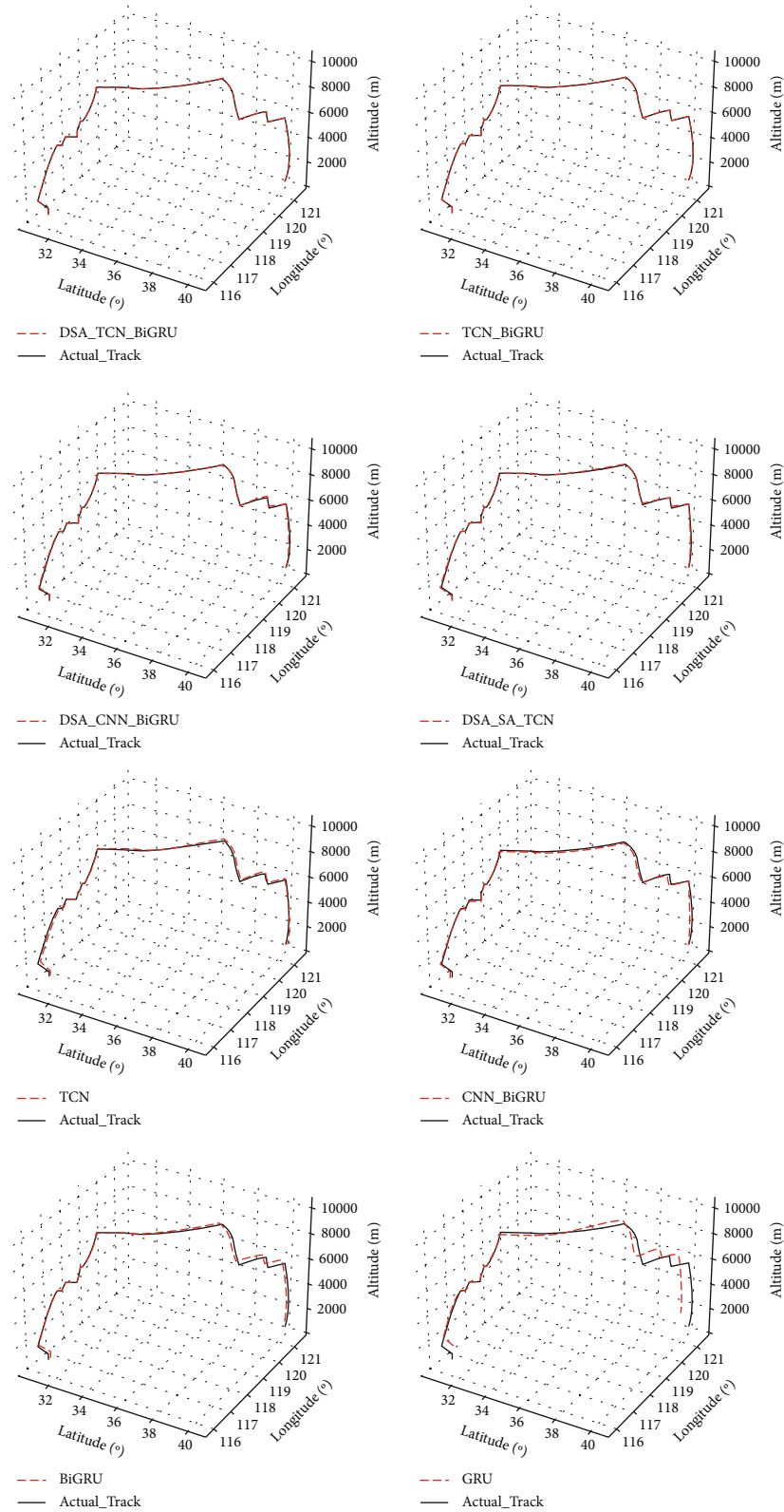


FIGURE 14: Visualisation of prediction results.

and TCN networks compared to the conventional CNN on the prediction models in trajectory prediction, the following model comparison groups were introduced: DSA-TCN-BiGRU vs. CNN-BiGRU, SATCN vs. TCN, and TCN-

BiGRU vs. CNN-BiGRU. The results are shown in Figure 17. We can see that, among these three groups of models, the DSA-TCN-BiGRU has the best prediction performance and lowest RMSE, with an altitude RMSE of

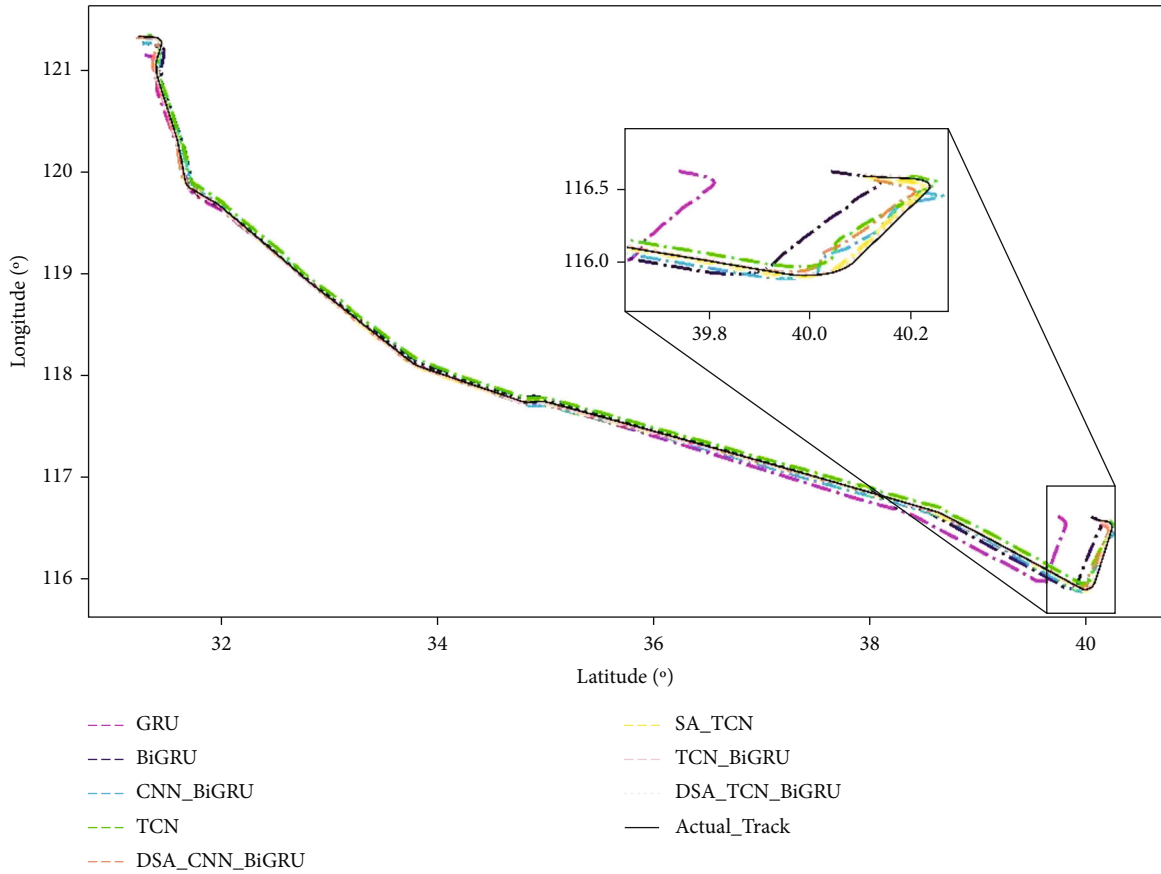


FIGURE 15: 2D comparison of DSA-TCN-BiGRU with baselines.

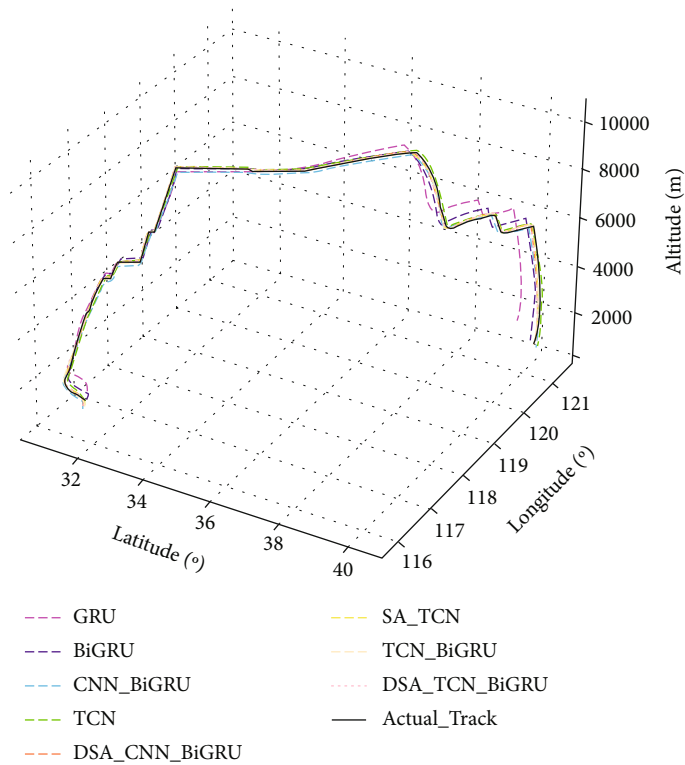


FIGURE 16: 3D comparison of DSA-TCN-BiGRU with baselines.

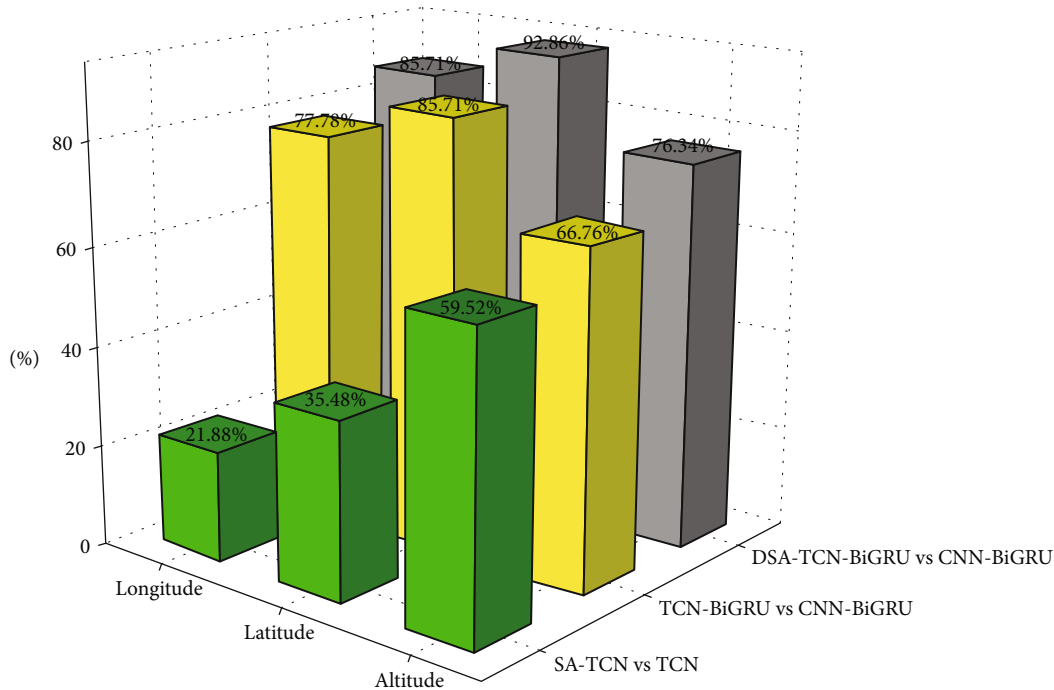


FIGURE 17: Model comparison results.

20.14 m, latitude RMSE of 0.004° , and a longitude RMSE of 0.009° .

Compared with the CNN-BiGRU, the DSA-TCN-BiGRU shows RMSE improvements of 85.71%, 92.86%, and 76.34% for the longitude, latitude, and altitude, respectively. Compared with the TCN, the RMSE improvement rates of the SATCN for the longitude, latitude, and altitude are 21.88%, 35.48%, and 59.52%, respectively. Compared with the CNN-BiGRU, the RMSE improvement rates of the TCN-BiGRU for the longitude, latitude, and altitude are 77.78%, 85.71%, and 66.76%, respectively. We can conclude that the hybrid model using the self-attentive mechanism and TCN has higher accuracy. The TCN can more effectively explore the potential features of the trajectory information, and the self-attentive mechanism focuses on the information that is more relevant to the results. Its ability to capture remotely dependent information can be further enhanced by modelling sequence data of variable length. In summary, the DSA-TCN-BiGRU can extract trajectory features from data more effectively, which is important for improving the accuracy of short-term trajectory predictions.

4.5.2. BiGRU Performance. To verify the advantages of using the BiGRU network, the DSA-TCN network was compared with a combination of the BiGRU, BiLSTM, GRU, and LSTM approaches. As shown in Table 3, the DSA-TCN-BiGRU and DSA-TCN-BiLSTM models have comparable performance with all hyperparameters tuned, and both had lower RMSE than the combined GRU and LSTM neural networks. This verifies that the introduction of a bidirectional mining mechanism can effectively improve the prediction accuracy. For a single-step prediction, the DSA-TCN-GRU has the lowest time of 40.1 ms, and DSA-TCN-BiLSTM has

TABLE 3: Comparison of models.

	RMSE			
	LAT/($^\circ$)	Lon/($^\circ$)	Altitude/(m)	Time/(ms)
DSA-TCN-BiGRU	0.004	0.009	20.14	47.2
DSA-TCN- BiLSTM	0.006	0.007	20.09	60.2
DSA-TCN-GRU	0.018	0.021	32.9	40.1
DSA-TCN-LSTM	0.013	0.019	31.8	51.3

TABLE 4: Future s-step prediction results.

Future s steps	RMSE		
	Lat/($^\circ$)	Lon/($^\circ$)	Altitude/(m)
1	0.004	0.009	20.14
2	0.003	0.008	19.64
6	0.005	0.010	23.48
12	0.006	0.011	27.64
18	0.008	0.013	29.11

the longest time of 60.2 ms. The BiGRU model is a special type of RNN with the advantages of fast convergence and fewer parameters compared with an LSTM. Thus, it simplifies the structure and improves the operation speed while guaranteeing the prediction performance.

4.5.3. Multistep prediction. In the aforementioned comparative analysis, the results from the model's single-step forecasts are shown. This section provides statistics on the characteristic mean error values of the model for single-step and multistep forecasts. In addition, the results from the model's single-step and multistep forecasts are quantitatively compared using the RMSE forecast error evaluation

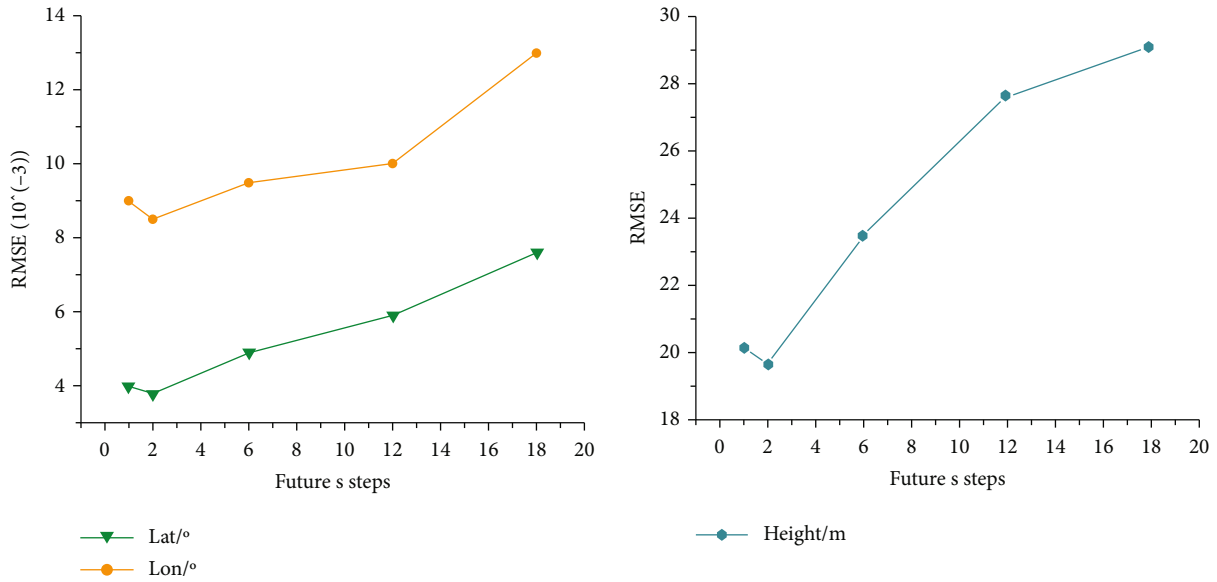


FIGURE 18: Effect of prediction step size on model performance.

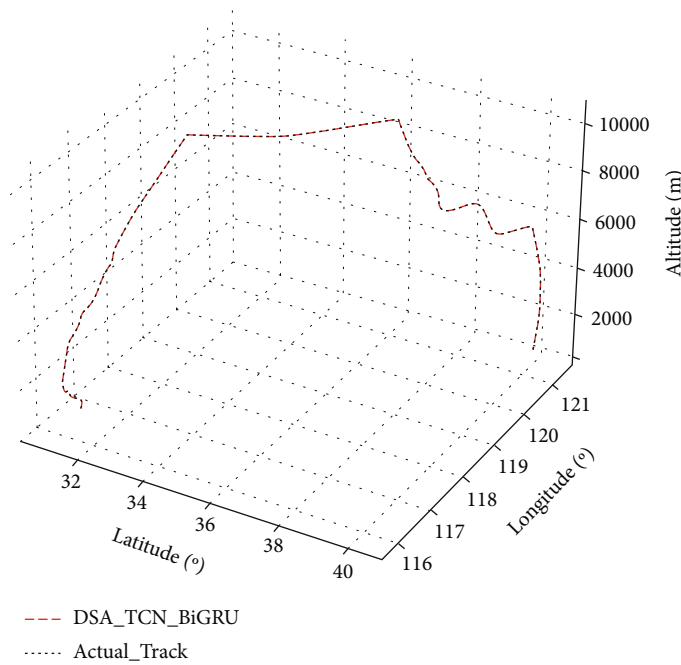


FIGURE 19: Comparison results for the proposed model and actual trajectory.

metric. Table 4 shows the results for the DSA-TCN-BiGRU model for predicting future s-steps, and Figure 18 shows the multistep prediction level line. As the number of prediction steps increases, the error remains within a certain range. This verifies that our proposed DSA-TCN-BiGRU model still performs well in multistep prediction, and can effectively improve the accuracy and practical requirements of trajectory prediction.

4.6. *Further Study.* We chose another flight along the same route to validate the robustness of the proposed DSA-TCB-

BiGRU model. The dataset contained data from the period 2021–2022, and the prediction results and evaluation metrics comparing the DSA-TCN-BiGRU based on the BO algorithm model with other models (DSA-CNN-BiGRU, CNN-BiGRU, TCN-BiGRU, SA-TCN, TCN, BiGRU, and GRU) are shown in Figures 19 and 20, respectively.

Our proposed DSA-TCN-BiGRU model has the lowest MAE and RMSE for the new dataset, and the fitted curve is closest to the actual route. This verifies that the DSA-TCN-BiGRU model based on BO has good robustness and accuracy for different datasets.

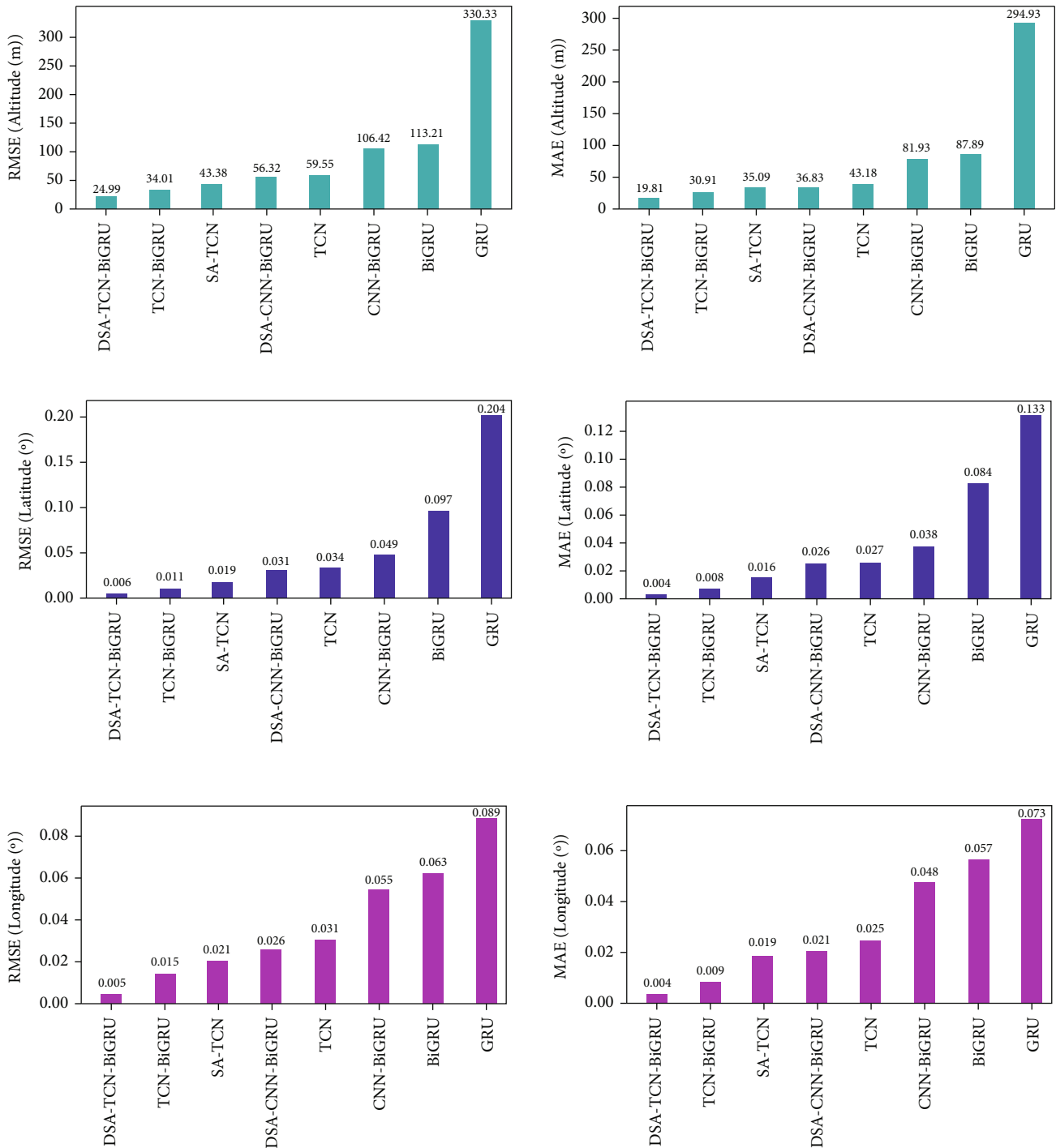


FIGURE 20: Results for evaluation indicators of the prediction model.

5. Conclusion

In this paper, we propose a DSA-TCN-BiGRU-based trajectory prediction method by fusing time-domain convolutional networks and two-way gated recurrent networks, and by introducing an attention mechanism and Bayesian hyperparameter optimisation algorithm. This combines the advantages of the TCN in extracting time series features with the BiGRU's ability to learn a sequence before and after con-

ditions to better extract the spatial and temporal features of trajectories while reducing the time complexity. Subsequently, an attention mechanism is introduced to assign different weights to the attributes, and the Bayesian optimisation algorithm is used to optimise the model hyperparameters. We compared the DSA-TCN-BiGRU model with a typical neural network and show the lowest RMSEs of 0.004°, 0.009°, and 20.14 m for the latitude, longitude, and altitude, respectively, demonstrating considerable improvements

in aircraft trajectory prediction accuracy. In addition, we verified the model's robustness and predictability performance using different datasets and multistep predictions. The proposed model assists controllers in decision-making by providing accurate information on aircraft trajectories over a while, calculating and identifying abnormal behaviour such as altitude anomalies and trajectory deviations, respectively, and alerting them to potential abnormal behaviour, helping controllers to be informed in advance of aircraft operating situation on the route and the terminal area. As a result, using the model can reduce the controller's workload and improve air traffic safety. The trajectory data also receives influence from factors such as meteorological controls. In the future, our approach will be to fuse weather, control, and other uncertainties. AMDAR can provide real-time weather data for aircraft, with the same real-time performance as ADS-B, and is the preferred choice for weather data fusion, considering the time delays associated with data fusion and forecasting. In addition, we can develop a trajectory data visualisation system that further explores the potential value of trajectory data by integrating the trajectory prediction module and displaying it visually on a map, which will significantly improve ATC's performance and predictive capability.

Data Availability

The ADS-B data used in this paper are from the publicly available dataset of Flightradar24 Technologies, Inc. (<https://www.flightradar24.com/data>).

Conflicts of Interest

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

Acknowledgments

This study was cosupported by the CAAC Vertical Project (Project No. 0242109) and (Project No. 180Z0550), the Civil Aviation Air Traffic Management Authority Horizontal Project (Project No. H2021-14) and (Project No. H2021-57), and the Institute of New Technologies for Civil Aviation Communications Navigation Surveillance (Project No. JG2022-20).

References

- [1] S. Mondoloni and N. Rozen, "Aircraft trajectory prediction and synchronization for air traffic management applications," *Progress in Aerospace Sciences*, vol. 119, article 100640, 2020.
- [2] K. Yang, M. Bi, Y. Liu, and Y. Zhang, "LSTM-based deep learning model for civil aircraft position and attitude prediction approach," in *2019 Chinese control conference (CCC)*, pp. 8689–8694, Guangzhou, China, 2019.
- [3] Y. Pang, X. Zhao, J. Hu, H. Yan, and Y. Liu, "Bayesian spatio-temporal graph transformer network (b-star) for multi-aircraft trajectory prediction," *Knowledge-Based Systems*, vol. 249, no. - article 108998, 2022.
- [4] W. Ding, J. Huang, G. Shang et al., "Short-term Trajectory prediction based on Hyperparametric optimisation and a dual attention mechanism," *Aerospace*, vol. 9, no. 8, p. 464, 2022.
- [5] Z. Fan, Z. Zhang, and H. Du, "Research on multi-objective constraint-based trajectory prediction method for descending segments [J/OL]," *Electro-Optics and Control* 1–7 <http://kns.cnki.net/kcms/detail/41.1227.TN.20220712.1851.016.html>.
- [6] Z. C. Li, *Research on Four Dimensional Flight Path Calculation Model Based on Aircraft Performance [D]*, Civil Aviation University of China, 2015.
- [7] Z. M. Du, J. F. Zhang, H. L. Miao, and B. Wang, "Aircraft vertical profile prediction for continuous climb based on thrust intention [J/OL]," *Journal of Beijing University of Aeronautics and Astronautics*, pp. 1–10, 2022.
- [8] Y. A. Pan, M. A. Nascimento, and J. Sander, "Online stochastic prediction of mid-flight aircraft trajectories," in *Proceedings of the 12th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, pp. 1–10, Chicago, Illinois, USA, 2019.
- [9] L. Ma and Y. Gao, "Four-dimensional trajectory prediction method based on ADS-B data mining," *Journal of Civil Aviation University of China*, vol. 37, no. 4, pp. 1–4, 2019.
- [10] G. Zazzaro, F. Martone, G. Romano, A. Vitale, and E. Filippone, "A data-driven methodology for pre-flight trajectory prediction," in *VEHITS*, pp. 188–197, 2022.
- [11] C. Morales and S. Moral, "Flight Trajectory Clustering: a framework that uses planned route data," in *2019 IEEE 14th international symposium on autonomous decentralized system (ISADS)*, pp. 1–6, Utrecht, Netherlands, 2019.
- [12] H. C. Choi, C. Deng, and I. Hwang, "Hybrid machine learning and estimation-based flight trajectory prediction in terminal airspace," *IEEE Access*, vol. 9, pp. 151186–151197, 2021.
- [13] Y. Ma, *Research on Precise Prediction Method of 4D Trajectory Based on Data Mining [D]*, Nanjing University of Aeronautics and Astronautics, 2016.
- [14] W. Zeng, Z. Xu, Z. Cai, X. Chu, and X. Lu, "Aircraft trajectory clustering in terminal airspace based on deep autoencoder and gaussian mixture model," *Aerospace*, vol. 8, no. 9, p. 266, 2021.
- [15] Y. Pang, X. Zhao, H. Yan, and Y. Liu, "Data-driven trajectory prediction with weather uncertainties: a Bayesian deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 130, article 103326, 2021.
- [16] A. Zhang, B. Zhang, W. Bi, and Z. Mao, "Attention based trajectory prediction method under the air combat environment," *Applied Intelligence*, vol. 52, pp. 1–15, 2022.
- [17] J. Zhou, H. Zhang, W. Lyu, J. Wan, J. Zhang, and W. Song, "Hybrid 4-dimensional Trajectory prediction model, based on the reconstruction of prediction time span for aircraft en Route," *Sustainability*, vol. 14, no. 7, p. 3862, 2022.
- [18] D. Sahadevan, P. Ponnusamy, V. P. Gopi, and M. K. Nelli, "Ground-based 4d trajectory prediction using bi-directional LSTM networks," *Applied Intelligence*, vol. 52, no. 14, pp. 16417–16434, 2022.
- [19] P. N. Tran, H. Q. V. Nguyen, D. T. Pham, and S. Alam, "Aircraft trajectory prediction with enriched intent using encoder-decoder architecture," *IEEE Access*, vol. 10, pp. 17881–17896, 2022.
- [20] Z. Lv, J. Li, C. Dong, and W. Zhao, "A deep spatial-temporal network for vehicle trajectory prediction," in *International Conference on Wireless Algorithms, Systems, and Applications*, pp. 359–369, Springer, Cham, 2020.

- [21] H. Zhang, Y. Peng, and D. Yang, "Research on trajectory prediction method of high mobility aircraft based on GRU model," *Changjiang Information Communications*, vol. 35, no. 4, pp. 7–9, 2022.
- [22] L. Qian and W. Wang, "A C-GRU based flight trajectory prediction method," *Electronic Measurement Technology*, vol. 45, no. 10, pp. 87–92, 2022.
- [23] H. Zhang, C. Huang, Y. Xuan, and S. Tang, "Real-time prediction of air combat flight trajectory using GRU," *Systems Engineering and Electronics*, vol. 42, no. 11, pp. 2546–2552, 2020.
- [24] X. Wu, H. Yang, H. Chen, Q. Hu, and H. Hu, "Long-term 4D trajectory prediction using generative adversarial networks," *Transportation Research Part C: Emerging Technologies*, vol. 136, article 103554, 2022.
- [25] H. Shafienya and A. C. Regan, "4D flight trajectory prediction using a hybrid deep learning prediction method based on ADS-B technology: A case study of Hartsfield-Jackson Atlanta International Airport (ATL)," *Transportation Research Part C: Emerging Technologies*, vol. 144, article 103878, 2022.
- [26] Y. Chen, J. Sun, Y. Lin, G. Gui, and H. Sari, "HybridN-Inception-LSTM-Based aircraft coordinate prediction method for secure air traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2773–2783, 2022.
- [27] P. Han, W. Wang, Q. Shi, and J. Yue, "A combined online-learning model with K-means clustering and GRU neural networks for trajectory prediction," *Ad Hoc Networks*, vol. 117, article 102476, 2021.
- [28] L. Ma and S. Tian, "A hybrid CNN-LSTM model for aircraft 4D trajectory prediction," *IEEE access*, vol. 8, pp. 134668–134680, 2020.
- [29] S. J. Corrado, T. G. Puranik, O. P. Fischer, and D. N. Mavris, "A clustering-based quantitative analysis of the interdependent relationship between spatial and energy anomalies in ADS-B trajectory data," *Transportation Research Part C: Emerging Technologies*, vol. 131, article 103331, 2021.
- [30] L. Dong, D. Fang, X. Wang et al., "Prediction of streamflow based on dynamic sliding window LSTM," *Water*, vol. 12, no. 11, p. 3032, 2020.
- [31] E. Fu, Y. Zhang, F. Yang, and S. Wang, "Temporal self-attention-based conv-LSTM network for multivariate time series prediction," *Neurocomputing*, vol. 501, pp. 162–173, 2022.
- [32] Z. Shang, B. Zhang, W. Li, S. Qian, and J. Zhang, "Machine remaining life prediction based on multi-layer self-attention and temporal convolution network," *Complex & Intelligent Systems*, vol. 8, no. 2, pp. 1409–1424, 2022.
- [33] H. Wang and Z. Zhang, "TATCN: time series prediction model based on time attention mechanism and TCN," in *2022 IEEE 2nd international conference on computer communication and artificial intelligence (CCAI)*, pp. 26–31, Beijing, China, 2022.
- [34] W. Sheng, K. Liu, D. Jia, S. Chen, and R. Lin, "Short-term load forecasting algorithm based on LST-TCN in power distribution network," *Energies*, vol. 15, no. 15, p. 5584, 2022.
- [35] Y. Cao, Y. Hao, B. Li, and J. Xue, "Depression prediction based on BiAttention-GRU," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 11, pp. 5269–5277, 2022.
- [36] F. Liu, Z. Zhang, and R. Zhou, "Automatic modulation recognition based on CNN and GRU," *Tsinghua Science and Technology*, vol. 27, no. 2, pp. 422–431, 2022.
- [37] Y. Xu and X. Jiang, "Short-term power load forecasting based on BiGRU-attention-SENet model," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 44, no. 1, pp. 973–985, 2022.
- [38] M. Munem, T. R. Bashar, M. H. Roni, M. Shahriar, T. B. Shawkat, and H. Rahaman, "Electric power load forecasting based on multivariate LSTM neural network using Bayesian optimization," in *2020 IEEE Electric power and energy conference (EPEC)*, pp. 1–6, Edmonton, AB, Canada, 2020.
- [39] A. H. Victoria and G. Maragatham, "Automatic tuning of hyperparameters using Bayesian optimization," *Evolving Systems*, vol. 12, no. 1, pp. 217–223, 2021.