

RESEARCH ARTICLE

4D aircraft trajectory prediction considering severe weather effects

H. Zhang  and Z. Liu 

Department of Land Surveying & Geo-Informatics, The Hong Kong Polytechnic University (PolyU), Hong Kong, People's of Republic China

Corresponding author: Zhizhao Liu; Email: lszzliu@polyu.edu.hk

Received: 13 October 2024; **Revised:** 8 July 2025; **Accepted:** 8 July 2025

Keywords: 4D trajectory prediction; CTGAN; ADS-B; deep learning

Abstract

The rapid growth of civil aviation has posed significant challenges to air traffic management (ATM), highlighting the need for accurate aircraft trajectory prediction (TP). Due to the scarcity of relevant data and the resulting class imbalance in the sample, aircraft TP under severe weather conditions faces significant challenges. This paper proposes an aircraft TP method framework consisting of trajectory data augmentation and TP networks to address this issue. To validate the effectiveness of this framework in solving the TP problem in severe weather, we propose an improved conditional tabular generative adversarial networks (CTGAN)-long short-term memories (LSTMs) hybrid model. We conduct comparative experiments of four LSTM-based models (LSTM, convolutional neural network (CNN)-LSTM, CNN-LSTM-attention, and CNN-BiLSTM) under this framework. The improved CTGAN is also compared with the commonly used data augmentation method, the Synthetic Minority Oversampling Technique (SMOTE). The results show that the TP accuracy can be effectively improved by enhancing the minority-class sample data; compared with SMOTE, the improved CTGAN is more suitable for minority-class sample data augmentation for aircraft TP, and it also shows that for minority-class sample data augmentation, data distribution characteristics are more important than the simple trajectory point accuracy. The hybrid modeling approach with the improved CTGAN as the data augmentation network proposed in this study provides valuable insights into addressing the data imbalance problem in aircraft TP.

Nomenclature

\oplus	vector concatenation operation
$\hat{\alpha}_i$	a scalar
$\hat{\beta}_i$	leave-one-out vector
BN	BN function
$C(\cdot)$	critic
$cond$	condition vector
\hat{d}_i	discrete value
$drop$	dropout
$FC_{u \rightarrow v}(x)$	application of a linear transformation to a u -dimensional input to obtain a v -dimensional output
$gumbel_{\tau}(x)$	application of the Gumbel softmax function with a parameter τ to a vector x
h	track angle
h_i	the i hidden layer
k	Gaussian kernel function
$leaky_{\gamma}(x)$	application of the leaky ReLU activation function on x with leaky ratio γ

m	the maximum number of variables of Y
M	number of historical tracks
n	the maximum number of variables of X
N	number of points in the track
r_i	i th random sample
s_i	average trajectory point of the generated dataset
\hat{s}_i	average trajectory point of the real dataset
σ	width parameter of the kernel function
t	time
\tanh	tanh function
v_{xy}	ground speed
x	latitude
X	a random variable in the generated distribution
y	longitude
Y	a random variable in the actual distribution
z	geometric height
θ	a parameter

Acronyms

<i>ADS-B</i>	automatic dependent surveillance-broadcast
<i>ATM</i>	air traffic management
<i>ATS</i>	air traffic service
<i>CGAN</i>	conditional generative adversarial networks
<i>CNN</i>	convolutional neural network
<i>CTGAN</i>	conditional tabular generative adversarial networks
<i>DSW</i>	dynamic space warp
<i>ET</i>	echo top
<i>GH</i>	geometric height
<i>GMM</i>	Gaussian mixture models
<i>GRU</i>	gated recurrent units
<i>GS</i>	ground speed
<i>HKO</i>	Hong Kong observatory
<i>HMM</i>	hidden Markov models
<i>ICAO</i>	International Civil Aviation Organization
<i>Lat</i>	latitude
<i>Lon</i>	longitude
<i>LSTM</i>	long short-term memory
<i>MED</i>	mean Euclidean distance
<i>MLP</i>	multilayer perceptron
<i>MMD</i>	maximum mean discrepancy
<i>RMSE</i>	root mean square error
<i>SA</i>	self-attention
<i>SMOTE</i>	Synthetic Minority Over-sampling Technique
<i>TA</i>	track angle
<i>TBO</i>	trajectory-based operations
<i>TGAN</i>	table generative adversarial networks
<i>TMP</i>	<i>atmospheric temperature</i>
<i>TP</i>	trajectory prediction
<i>VIL</i>	vertically integrated liquid

1.0 Introduction

In the past a few decades, civil aviation transportation has developed rapidly, posing significant challenges to the existing air traffic management (ATM) system. To address this issue, the International Civil Aviation Organization (ICAO) proposed the Global ATM Operational Concept, identifying

trajectory-based operations (TBO) as a core technology [1]. Within this model, precise aircraft trajectory prediction (TP) is crucial for understanding the overall operational status of the airspace, as well as for improving and ensuring the efficiency and safety of the ATM system. This prediction provides the most direct reference for aircraft surveillance and control forecasts [2].

The challenge in accurately predicting aircraft trajectories lies in the influence of environmental conditions [3]. These conditions include meteorological factors, airspace congestion and infrastructure service capacity. Convective weather and severe turbulence significantly impact flight safety, and they are the primary reasons for changes in aircraft trajectories. Typically, aircraft follow planned air traffic service (ATS) routes as outlined by the flight plan. However, when encountering convective weather or severe turbulence, ATS may direct en-route aircraft to alter their flight paths (either horizontally or vertically) and adjust their arrival times at given positions to avoid severe weather effects. Even if aircraft follow their planned routes, significant deviations can occur due to the impact of severe weather. Therefore, it is crucial to consider weather factors, particularly the impact of severe weather on aircraft TP [4].

Weather information has long been used in aircraft TP. However, due to the dynamic and uncertain nature of weather, effectively leveraging its impact to improve TP accuracy remains a challenge [5]. To address this issue, artificial intelligence technologies such as deep learning have been introduced to aircraft TP to enhance forecasting accuracy [6]. Nevertheless, most current studies fail to differentiate between weather types or fail to thoroughly investigate their data characteristics and impact on deep learning performance. In practical operations, severe weather significantly impacts ATM systems and is a critical issue for all sectors of air transportation. In TP, the focus is primarily on flyable weather conditions [7]. For description convenience, this paper classifies flyable weather conditions into two categories: normal and severe. Severe weather refers to meteorological conditions such as severe convective weather, heavy precipitation and low visibility, which have significant negative impacts on aviation safety and operations [7]. It excludes extreme weather conditions that render flight impossible. Other common weather conditions that affect flight but do not have a substantial impact are classified as normal weather. Due to the data scarcity, flight data under severe weather conditions account for only a small proportion compared to historical trajectory data under normal weather conditions, leading to an imbalanced dataset in aircraft TP. The flight data under severe weather conditions are considered minority-class samples. Existing TP models often perform poorly when learning features from minority-class samples in imbalanced datasets. This is because data-driven methods treat minority-class samples as outliers or fail to capture their features due to insufficient sample size, leading to decreased prediction accuracy.

Considering this situation, this paper proposes an aircraft TP framework consisting of data augmentation and TP models, with an improved conditional tabular generative adversarial network (CTGAN) used as the data augmentation network. To validate the effectiveness of this approach, the widely used long short-term memory (LSTM) model is incorporated into a hybrid model, termed CTGAN-LSTMs, to address the challenging task of four-dimensional (4D) aircraft TP, considering the influence of severe weather conditions. The main contributions of this paper are as follows:

- (1) A novel aircraft TP framework integrating data augmentation and prediction networks, specifically designed to address data imbalance in severe weather scenarios;
- (2) Development of the CTGAN-LSTMs hybrid model demonstrating significant accuracy improvements for short-term TP under severe weather;
- (3) The improved performance of CTGAN over traditional Synthetic Minority Over-sampling Technique (SMOTE) for augmenting minority-class trajectory data is validated.

Through these contributions, this research significantly enhances the ability to predict aircraft trajectories under severe weather conditions, providing a feasible approach to improve TP accuracy based on minority-class sample data. The structure of this paper is as follows. Section 2 provides a literature review on the current background of aircraft TP considering the influence of weather. Section 3

introduces the overall framework of the method developed in this paper, including data preprocessing, improvements to CTGAN, design of model evaluation metrics, model training process and trajectory prediction. Section 4 presents the experimental setup and results. Section 5 discusses our contributions and limitations of the research work. The conclusions of the study are presented in Section 6.

2.0 Introduction

2.1 Application of weather data

The commonly used aircraft TP methods considering the impact of weather can be summarised into two categories based on the prediction approach: traditional kinetic approach and machine learning approach.

Traditional kinetic approach includes state estimation methods and dynamic methods. State estimation methods establish motion equations based solely on aircraft attributes such as position, velocity and acceleration. Although this model is relatively simple, it may lead to significant errors due to its inability to accurately capture aircraft manoeuvring uncertainty over long term [3]. Dynamic methods build TP models based on aerodynamic principles. Modeling requires considering various factors, including aircraft performance, status, environmental conditions and intent. This involves a large number of parameters. Some parameters may be commercially sensitive and challenging to obtain. Additionally, specific parameters exhibit dynamic characteristics and change continuously during flight, making accurate estimation challenging.

Weather is a critical input parameter for TP and significantly influences prediction accuracy. Various methods have been proposed to reduce weather-induced uncertainties. For instance, Porretta et al. [8] utilised a standard civil aircraft 3D point mass model to simulate dynamics, decomposing wind fields into nominal and random components to improve TP accuracy. Tang et al. [9] proposed methods to correct meteorological forecasts and extract nominal flight profiles, significantly enhancing strategic 4D trajectory forecasts.

Despite these improvements, traditional methods struggle to meet the high-precision, multi-dimensional demands of the TBO environment. As a result, there is a growing interest in machine learning approaches that leverage historical flight and meteorological data. These methods, such as Gaussian mixture models (GMM), hidden Markov models (HMM), multilayer perceptrons (MLP), convolutional neural networks (CNN), LSTM and generative adversarial networks (GAN), have shown promise in distinguishing weather impacts on trajectories [10, 11].

Ayhan and Samet were the first to propose the concept of surrounding weather data (feature cubes) [12], significantly influencing the subsequent development of TP models under weather impact. Their approach involved conceptualising airspace as a 3D grid network and storing weather information in each grid. By establishing the relationship between each trajectory point and the weather grid ahead, these grids became joint cubes for 4D trajectories. They then used an HMM model to predict trajectories with weather uncertainty. Subsequent research on aircraft TP considering weather impact has broadly referenced the concept of weather cubes.

Liu and Hansen proposed a deep generative network framework for aircraft TP [11], which consists of an LSTM encoder network and a mixture density LSTM decoder network. The decoder network learns spatiotemporal correlations from historical flight trajectories and outputs parameters of a GMM. This model uses beam search, adaptive Kalman filtering and Rauch-Tung-Striebel smoothing algorithm during inference. It can take high-dimensional weather features (wind, temperature and convective weather) and the final flight plan as input conditions to predict the aircraft's 4D trajectory.

Inspired by Ref. [11], Pang et al. proposed a network architecture that embeds convolutional layers into LSTM units to predict trajectories based on convective weather conditions before aircraft takeoff and flight plans [13]. In this model, echo top (ET) data are used as weather condition data and flight plan input to the LSTM layer to predict the aircraft's position. The results showed an accuracy improvement in 47% of flight plans and the average accuracy increase is 12.3%. However, this result focuses on 3D trajectory prediction (ignoring altitude factors). Subsequently, Pang et al. replaced the previous LSTM

units with conditional generative adversarial networks (CGAN) architecture to predict trajectories based on ET data and flight plans [14]. This improved flight plan accuracy to 55.2%, with an overall variance reduction of 22.1%. Building upon this, Pang et al. extended their research to 4D TP [15], reporting errors in the forms of latitude, longitude and feet altitude. The errors typically are within 1 degree of latitude or longitude in horizon, approximately 60 nautical miles (nmi), and the vertical error is 100 feet (ft) in altitude. This study improved the accuracy of aircraft TP by explicitly considering weather uncertainty through weather feature extraction methods.

Recent research on weather-related aircraft TP is referenced in Ref. [6]. Schimpf et al. proposed a generalised hybrid-recurrent prediction model for aircraft TP to eliminate potential biases in model training across different routes and dates. This model combines CNN, self-attention (SA), LSTM and gated recurrent units (GRU), enabling route generalisation and improving flight TP accuracy. In this approach, the weather is processed into cubes using conventional methods. After feature extraction, it is connected with relevant trajectory point information as feature inputs for model learning and prediction. The study also comprehensively compares weather products used in previous literature and their combined effectiveness, suggesting that the ET feature is the most suitable overall weather product for aircraft TP.

To summarise, current research has significantly addressed weather-related aircraft TP tasks. However, these studies have not conducted an in-depth analysis of the data characteristics associated with different weather types, and the proposed methods fail to effectively address the TP challenges under severe weather conditions. High-precision aircraft TP under severe weather conditions remains a challenge.

There are two main challenges in predicting aircraft trajectories under severe weather conditions. First, the quantity of flight data corresponding to severe weather is scarce, accounting for a small proportion of the flight sample dataset, and belongs to the minority class. Commonly used deep neural networks typically employ small-batch random sampling methods to improve training efficiency and model generalisation ability. Since the probability of selecting minority-class samples is low, it is difficult to accurately capture the hidden complex relationships. The second challenge is that, compared to normal weather, severe weather has a greater impact on aircraft trajectories, leading to more significant trajectory deviations. Machine learning algorithms easily identify these deviations as outliers and may be ignored. Therefore, research on TP techniques based on imbalanced data is key to solving the TP problem under severe weather conditions.

2.2 Imbalanced data processing

2.2.1 Basic handling methods of minority-class samples

Minority-class samples refer to one or more types of samples that occupy a tiny proportion of the entire dataset. The causes of data imbalance can be attributed to two factors: first, the inherent imbalance in natural phenomena, and second, practical limitations in the data collection process, which result in fewer samples from one or more classes. The emergence of minority-class issues is closely related to how sample data is utilised during the training process of deep learning models. Common machine learning methods often use small-batch random sampling techniques to accelerate the training process and enhance the model's generalisation ability. As a result, the chances of selecting minority-class samples are lower, making extracting relevant feature knowledge challenging. This is especially true for deep neural networks, which learn complex nonlinear relationships from data distributions. A minimal number of minority-class samples makes it difficult for the model to accurately capture the hidden and complex relationships.

There are four main methods to address data imbalance: data-level methods, algorithm-level methods, cost-sensitive methods and ensemble methods [16]. Among these, data-level methods are the most basic because they do not require modifications or adjustments to software models. We have chosen this approach to handle aircraft TP problems under severe weather conditions. Data-level methods include oversampling and undersampling techniques. Undersampling reduces the majority class size, which is sometimes impractical in real-world scenarios due to the smaller resulting dataset. Oversampling

expands the minority class by adding more samples, which is widely used to balance class distribution in datasets [17]. Oversampling poses less risk of overfitting than undersampling when dealing with minority-class sample issues [18]. The over-sampling techniques are also known as data augmentation methods.

Minority-class samples are different from small samples because the former involves the issue of data imbalance. This requires a greater focus on the relationship between minority and majority class samples. When performing data augmentation, it is crucial to maintain a clear boundary between the minority and majority classes to avoid misclassification in future applications. Additionally, minority-class samples often consist of several smaller subclasses, and the relationships and distribution characteristics among these subclasses are vital for achieving finer-grained learning and improving the performance of machine learning models. This is an important aspect to consider when predicting minority-class samples.

2.2.2 SMOTE data augmentation method

The most commonly used oversampling technique is SMOTE, which generates synthetic data by applying linear interpolation between minority class points and their K nearest neighbours [16]. Reducing the possible risk of overfitting, reducing information loss, and strong adaptability are the outstanding advantages of SMOTE. Unlike random oversampling, SMOTE minimises the risk of overfitting by synthesising new samples instead of simply copying existing samples and retaining more original data information, which can minimise information loss. In addition, SMOTE is simple in principle and easy to use. The number of synthesised samples and the number of neighbouring samples (K value) can be adjusted as needed, thereby flexibly controlling the number and quality of generated samples, and it has strong adaptability. Given the above advantages of SMOTE, it has been widely used in various industries. SMOTE also has some limitations, mainly the assumption of uniform distribution of samples, blurred category boundaries and noise sample problems. SMOTE assumes that minority-class samples are uniformly distributed in the feature space without considering that minority-class samples are composed of several slight separations or sub-concepts. However, in real-world datasets, minority-class samples may not be uniformly distributed. Secondly, when SMOTE generates new samples, it may create samples in the boundary area between different categories, causing the category boundaries to become blurred. Finally, if there are noise samples in the minority-class samples, SMOTE may generate new synthetic samples based on these noise samples, thereby introducing more noise. The main reason for the above limitations is that the data distribution patterns assumed by SMOTE may not match the actual ones because they are not generated from the original minority-class distribution. These deficiencies affect the quality of synthetic data and may cause a decline in the quality of subsequent machine learning model training and application performance [15, 18, 19].

2.2.3 Deep learning data augmentation methods

In recent years, GANs, a novel neural network technology, have been widely applied in fields such as image generation [20], data generation [21] and image resolution enhancement [22] through the competition between generators and discriminators to learn the distribution of given data. GANs are also suitable for data augmentation and trajectory synthesis [23]. Reference [24] is the only paper we have collected that applies data augmentation techniques to aircraft TP. Zhang et al. [24] aimed at application scenarios where the number of TP sample data is scarce, using WGAN-GP (a GAN model) to expand the trajectory sample data and training the LSTM model for aircraft TP based on the sample data sets before and after enhancement. Experiments show that as the sample data set expands, the TP accuracy of LSTM increases significantly. The model is designed for small sample scenarios rather than minority-class sample scenarios of data imbalance. It is based on automatic dependent surveillance-broadcast (ADS-B) data without considering the impact of weather factors. Nevertheless, it represents a valuable attempt to apply data augmentation methods to improve aircraft TP accuracy.

CTGAN [25] is the latest network model developed based on the GAN framework. It has achieved good application results in handling imbalanced data [26] and detecting network attacks [27]. The CTGAN architecture combines the advantages of CGAN and table generative adversarial networks (TGAN). It excels in handling unique challenges associated with tabular data, such as the non-Gaussian distribution of continuous columns, the imbalanced distribution of discrete columns and the mix of discrete and continuous columns, allowing tabular data to be used directly as input [28]. This model outperforms traditional GANs in data feature extraction [29].

Considering that the ADS-B data mainly used in aircraft TP tasks can be easily converted into a table form, and from the perspective of data distribution, ADS-B data lacks symmetry and does not conform to the Gaussian distribution characteristics; from the standpoint of sample data composition, the number of aircraft trajectory data samples under severe weather conditions is scarce and disproportionate to the number of samples under typical meteorological conditions, resulting in an unbalanced distribution of sample data; using CTGAN for data augmentation can effectively address the problem of severe weather datasets being minority-class samples that are difficult to recognise by commonly used machine learning TP algorithms. The authors' previous research has demonstrated that the CTGAN model is well-suited for aircraft TP using ADS-B data as the primary flight data [30]. In this context, CTGAN combined with an LSTM-based hybrid framework is expected to be a viable solution to the complex problem of deplaning TP tasks associated with severe weather conditions using a suitable data augmentation network.

3.0 Methods

Since weather parameters are continuous variables, severe weather is a fuzzy concept whose boundaries are artificially defined and have statistical significance. For individuals in severe weather conditions where flights are possible, some of their trajectories are similar to those in normal weather conditions, making it difficult to distinguish them manually. Due to severe weather severity and specific factors, there may be different subclass distributions in minority-class samples. SMOTE performs data enhancement based on the uniform distribution assumption rather than the actual distribution; CTGAN generates synthetic data using the distribution learned from the samples. At the same time, CTGAN can process unbalanced data, which increases the sampling frequency of minority-class samples. When the boundaries of minority-class samples are artificially defined, it can also increase the enhancement of subclasses with a smaller proportion, which helps to reduce the accuracy requirements for setting the boundaries of severe weather. Combining severe weather characteristics and the advantages and disadvantages of SMOTE and CTGAN, CTGAN is more suitable for data enhancement in severe weather aircraft TP. To verify the effect of data enhancement, based on the representativeness of LSTM in aircraft TP tasks, we designed a hybrid model based on CTGAN and LSTM for short-term TP experiments. LSTM can be replaced by other TP models (including hybrid models) to adapt to different application scenarios. Taking LSTMs as the aircraft TP model as an example, the TP method based on data enhancement considers severe weather conditions. It combines CTGAN with LSTM models. The overall framework of the proposed 4D aircraft TP method, considering severe weather conditions and combining CTGAN with LSTM-based models, is illustrated in Fig. 1. It comprises four parts: data preparation, data augmentation, model training and trajectory prediction.

First, we preprocess ADS-B and weather data, using the processed weather data as conditions combined with ADS-B data as input. Second, we use an improved CTGAN model to augment the small-proportion severe weather trajectory data, controlling the quantity of generated synthetic data to maintain the essential characteristics of the data distribution before and after merging with the original trajectories. This allows the LSTM-based model to learn more effectively.

Subsequently, we combine the synthetic data generated by CTGAN (augmented data) with the processed ADS-B data and weather cube feature data to form a new dataset, which is then input into the LSTM network for training. Finally, the trained LSTM-based model is used for aircraft TP. Of

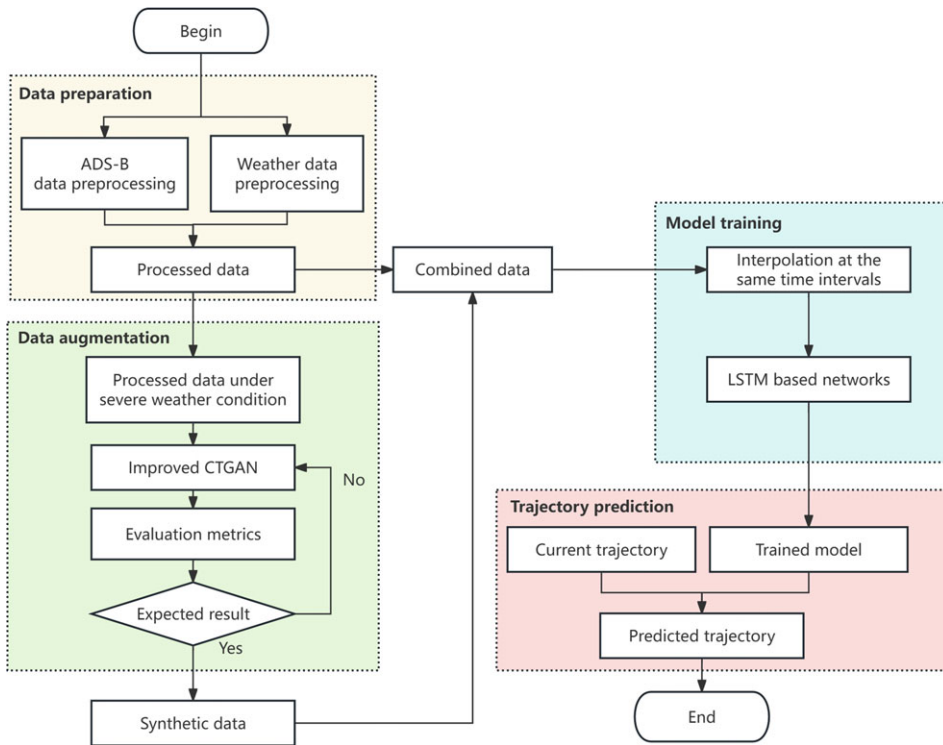


Figure 1. The general framework of the 4D aircraft TP method based on the improved CTGAN and LSTM-based architecture.

course, based on this architecture, other data augmentation networks and prediction networks can replace CTGAN and the LSTM-based model to form new hybrid models, with the basic process remaining unchanged.

3.1 Data preparation

3.1.1 ADS-B data preprocessing

During the broadcasting process, ADS-B data may cause data errors due to system failures and sensor errors, so it needs to be preprocessed before use. Common ADS-B data errors include data duplication, anomalies, data missing, etc. The data preprocessing process used in this paper is shown in Fig. 2.

Unlike common ADS-B data preprocessing workflows, this study introduces another step of aligning aircraft time parameters. Since CTGAN does not require equal-length trajectory samples for data augmentation, there is no need to interpolate the original data, preserving the original data distribution and improving the accuracy of the CTGAN-generated data. Different flights on the same route typically have varying departure, arrival and waypoint times. To facilitate the training of the CTGAN model with more flight data, the time parameters of the trajectory data need to be aligned.

Given that the trajectory characteristics of inbound and outbound flights differ, time parameters must be processed separately for each. For outbound flights, the time of origin is set when the aircraft leaves the ground during takeoff. For inbound flights, the time origin is set when the aircraft passes a designated waypoint.

Since ADS-B data may experience gaps during transmission, these missing data points can increase the time intervals between some trajectory points [31]. For example, the interval between adjacent ADS-B data points is typically 0.5–2 seconds, but occasionally it can exceed 30 seconds. Therefore,

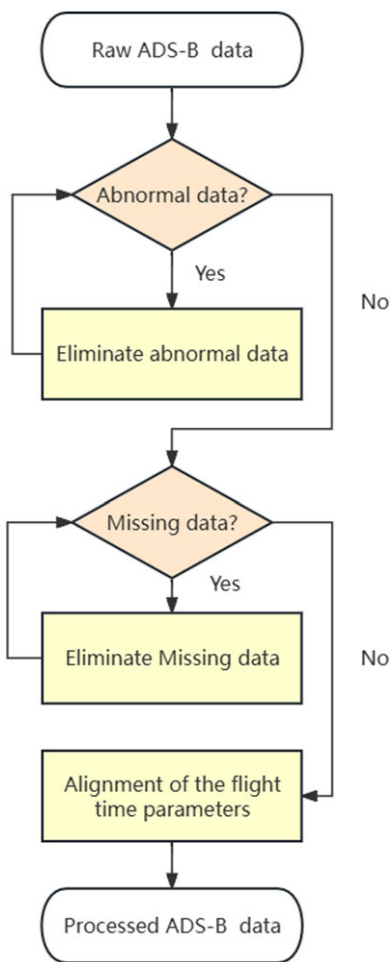


Figure 2. The ADS-B data preprocessing flow chart.

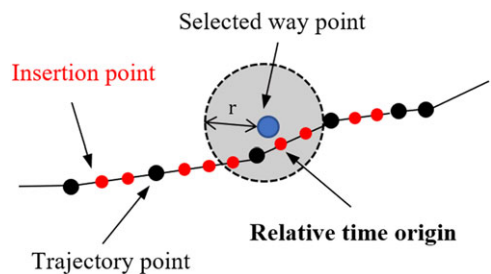


Figure 3. The diagram of alignment of arrival flight time parameters [30].

interpolating the flight data points near the designated waypoint is necessary when determining the relative time origin for inbound flights. The time of the trajectory point closest to the waypoint within a certain radius is chosen as the relative time origin. This alignment method is consistent with our previously published work [30]. The time parameter alignment process is illustrated in Fig. 3.

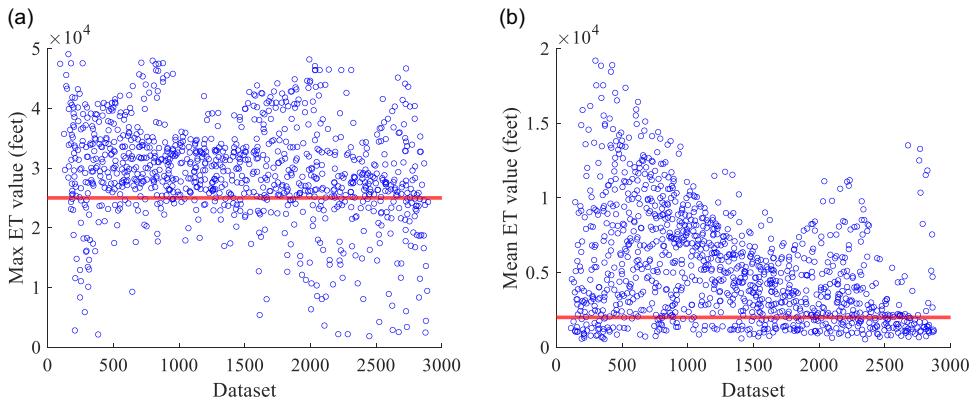


Figure 4. Visualisation of the data filtering process using thresholds on ET data. (a) represents maximum ET value. (b) represents mean ET value.

3.1.2 Weather data preprocessing and flight data selection under severe weather conditions

In current aircraft TP tasks, weather information commonly includes ET, vertically integrated liquid (VIL) features, atmospheric temperature (TMP) and U/V wind components. According to the study in Ref. [6], ET data is one of the most indicative of weather's impact on trajectories. Therefore, this study employs ET data as the model input for preprocessing and generating the feature cube.

We generate the weather feature cubes using the method described in Ref. [14]. We extract a weather cube ahead of the current position from the original weather file for each trajectory point and rotate it according to the aircraft's heading angle. Weather data is collected one layer above and below the current altitude. By scaling the original resolution of the weather file by a factor of 10, we can generate a $32 \times 32 \times 3$ feature cube for each trajectory point. In this way, the weather bandwidth can cover a sufficient area. Thus, we obtain a set of weather cubes that match the ADS-B trajectory data.

The feature extraction from the weather cubes follows the methods in Refs [6, 15]. Specifically, we use three convolutional layers and two dense layers, each followed by two dropout layers with a dropout rate of 1%. The convolution kernel sizes are $[6 \times 6, 3 \times 3, 1 \times 1]$, the convolution stride lengths are $[1 \times 2 \times 2]$, and the convolution filter sizes are $[1 \times 2 \times 4]$. The dense layer sizes are $[3 \times 15]$. This allows us to obtain weather cube feature data that matches the ADS-B trajectory data.

In this study, we selected flight data under severe weather conditions. First, based on historical weather forecasts from the Hong Kong Observatory (HKO), we identified dates with unfavorable weather and extracted flight data and corresponding ET (convective weather) data for those days. Previous studies on convective weather datasets indicate that flight trajectories are categorised as encountering weather deviations only when they experience convective weather at 25,000 feet or for at least 2 minutes [32]. Therefore, we filtered the flight data based on the maximum and average values of the corresponding ET data. The threshold for the ET maximum value is set to 25,000, and the average value threshold is set to 2,000. By this method, we screened flight data under severe weather conditions. Figure 4 visualises the data filtering process, adopting the method outlined in Ref. [15]. The X-axis represents the data pairs, and the Y-axis represents the maximum and average values of each data pair. The red line shows the thresholds used in different indicators.

3.2 Data augmentation

3.2.1 Improved CTGAN architecture

The architecture of the improved CTGAN for TP tasks under severe weather conditions referenced our previous study [30]. There are two significant differences from the original CTGAN to adapt to the features of ADS-B data.

Table 1. Training parameters for the improved CTGAN

Parameter	Description	Value
Embedding dimension	Size of the random sample passed to the Generator.	128
Generator dimension	Size of the output samples for each one of the residuals.	(256,256)
Discriminator dimension	Size of the output samples for each one of the Discriminator layers.	(256,256)
Generator learning rate	Learning rate for the generator.	$5 \cdot 10^{-4}$
Generator decay	Generator weight decay for the Adam Optimiser.	$1 \cdot 10^{-6}$
Discriminator learning rate	Learning rate for the discriminator.	$5 \cdot 10^{-4}$
Discriminator decay	Discriminator weight decay for the Adam Optimiser.	$1 \cdot 10^{-6}$
Batch size	Number of data samples to process in each step.	750
Step ratio	Number of discriminator updates to do for each generator update.	1
Epochs	Number of training epochs.	50,000
Pac	Number of samples to group together when applying the discriminator.	10

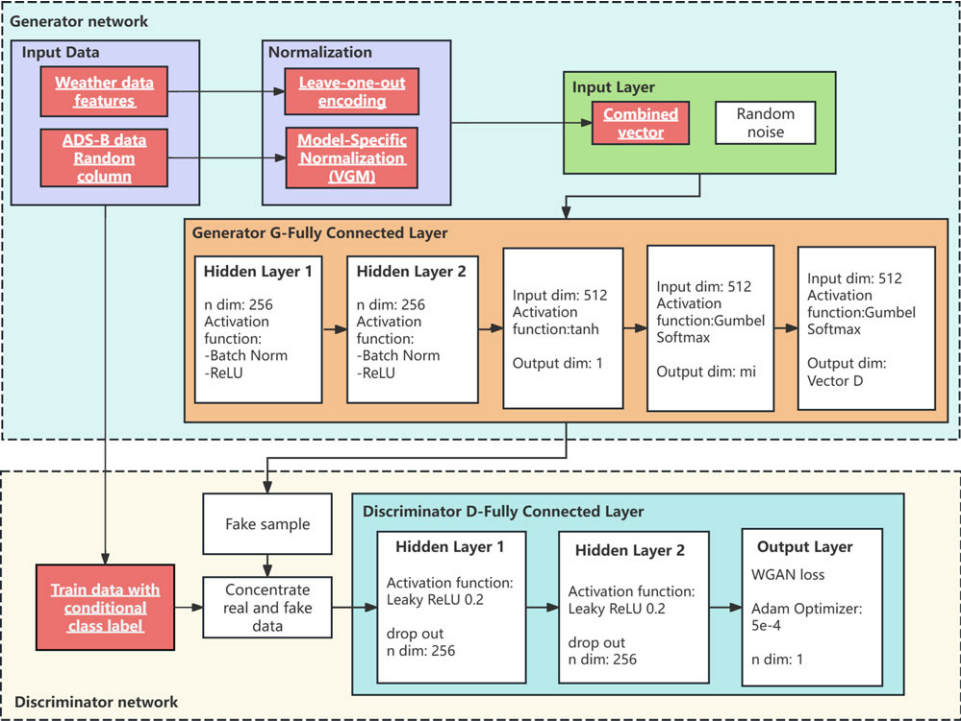


Figure 5. The improved CTGAN architecture considering severe weather conditions.

The first improvement involves the aircraft characteristics contained in ADS-B data. We sample using actual data frequencies rather than logarithmic frequencies. This is done to maintain the diversity of the enhanced data types and prevent the generator from unrealistically oversampling trajectories under severe weather conditions, significantly affecting model convergence [33].

The second improvement is replacing the original CTGAN's one-hot encoding with leave-one-out encoding to avoid problems of parallelism and multicollinearity in high-dimensional situations, where one-hot encoding can lead to too many and overly sparse generated features.

Unlike previous studies, this research introduces weather cube feature data as conditions. It aims to generate adverse weather data as much as possible during synthetic data generation to enhance data. Figure 5 illustrates the architecture of the improved CTGAN, with differences from the original CTGAN marked in red.

The improved CTGAN architecture consists of two parts: the generator network and the discriminator network. We perform data enhancement by adding features of aircraft trajectories under severe weather conditions and random noise. In the generator network, the random column vectors of ADS-B data are normalised using a variational Gaussian mixture model. After decoding the weather cube feature data using leave-one-out encoding, it is concatenated with the normalised column vectors to form a combined vector. This vector and random noise are input into the fully connected generator layer.

In the discriminator network, we use weather cube feature data as a condition to bias the network towards generating aircraft trajectory data under severe weather conditions, achieving the effect of data enhancement. The parameter settings of the entire architecture are the same as in our previous study, with specific parameters shown in Table 1. Formulas 1 and 2 represent the generator and discriminator networks, respectively.

$$\left\{ \begin{array}{l} h_0 = z \oplus cond \\ h_1 = h_0 \oplus ReLU(BN(FC_{|cond|+|z| \rightarrow 256}(h_0))) \\ h_2 = h_1 \oplus ReLU(BN(FC_{|cond|+|z|+256 \rightarrow 256}(h_1))) \\ \hat{\alpha}_i = \tanh(FC_{|cond|+|z|+512 \rightarrow 1}(h_2)) \quad 1 \leq i \leq N_c \\ \hat{\beta}_i = gumbel_{0.2}(FC_{|cond|+|z|+512 \rightarrow m_i}(h_2)) \quad 1 \leq i \leq N_c \\ \hat{d}_i = gumbel_{0.2}(FC_{|cond|+|z|+512 \rightarrow |D_i|}(h_2)) \quad 1 \leq i \leq N_d \end{array} \right. \quad (1)$$

$$\left\{ \begin{array}{l} h_0 = r_1 \oplus \dots \oplus r_{10} \oplus cond_1 \oplus \dots \oplus cond_{10} \\ h_1 = drop(leaky_{0.2}(FC_{10|r|+10|cond| \rightarrow 256}(h_0))) \\ h_2 = drop(leaky_{0.2}(FC_{256 \rightarrow 256}(h_1))) \\ C(\cdot) = FC_{256 \rightarrow 1}(h_2) \end{array} \right. \quad (2)$$

where \oplus represents the vector concatenation operation; z represents random noise; $FC_{u \rightarrow v}(x)$ denotes the application of a linear transformation to a u -dimensional input to obtain a v -dimensional output; $leaky_{\gamma}(x)$ represents the application of the leaky ReLU activation function on x with leaky ratio γ ; and $gumbel_{\tau}(x)$ represents the application of the Gumbel softmax function with a parameter τ to a vector x . The \tanh , BN and $ReLU$ functions are used for batch normalisation.

3.2.2 Evaluation metrics

During the data augmentation process, the CTGAN treats each row of input tabular data as an independent sample for random sampling without considering the order between rows. It learns the overall distribution characteristics of the data through these samples. Based on the input data and learned distribution characteristics, it outputs generated data in tabular form. For the aircraft TP model, the input to the CTGAN generator is usually actual time series trajectory data. The generated output data consists of non-time-series random data reflecting the distribution characteristics of the trajectories. Therefore, there is no direct one-to-one correspondence between them. For evaluating model training results, if the distributions of two datasets are very close, then the means of their trajectory parameters will also be very close. The mean can approximately reflect the difference between actual and generated distribution. Therefore, we use the mean value to evaluate the model.

According to the requirements of the data augmentation task described in this paper, evaluation metrics need to meet two conditions: first, to determine the quality of the generated data, ensuring that the data distribution of the new samples is similar to the original data; second, to eliminate the noise introduced by the generated data to avoid the fusion of generated data and original data leading to classification ambiguity. Therefore, to accurately evaluate the model, we use the maximum mean discrepancy (MMD) [34] from the perspective of data distribution and the mean Euclidean distance (MED) from the standpoint of numerical prediction to calculate the trajectory distance and eliminate the noise introduced by the generated data.

(1) Maximum mean discrepancy

The MMD is a commonly used GAN evaluation indicator. It finds continuous functions f in the sample space and calculates the mean difference among different distribution functions f to obtain the average f between the analysed datasets. To determine whether the two distributions are similar, we use the mean of the generated data to calculate the MMD. If a Gaussian kernel function is used to replace f , the formula can generally be interpreted as:

$$MMD^2(X, Y) = \left\| \frac{1}{n^2} \sum_i^n \sum_{i'}^n k(x_i, x_{i'}) - \frac{2}{nm} \sum_i^n \sum_j^m k(x_i, y_j) + \frac{1}{m^2} \sum_j^m \sum_{j'}^m k(y_j, y_{j'}) \right\|_H \quad (3)$$

$$k(u, v) = e^{-\frac{u-v^2}{2\sigma^2}} \quad (4)$$

where X refers to a random variable in the generated distribution; Y refers to a random variable in the actual distribution; n represents the maximum number of variables of X ; m represents the maximum number of variables of Y ; k represents the Gaussian kernel function; and σ is the width parameter of the kernel function. The smaller the MMD is, the more accurate the model is.

(2) Mean Euclidean distance

The MED is a typical metric for calculating the distance between trajectories. The ADS-B data can provide information such as longitudes, latitudes and altitudes (geometric altitudes in ADS-B) in the WGS-84 coordinate system. The measurement units of these parameters are different. Latitude and longitude are measured in degrees, and altitude is usually measured in feet. The latitude and longitude ranges are $[0, \pm 90]^\circ$ and $[0, \pm 180]^\circ$, respectively, and the altitude can be hundreds or thousands of times the magnitude of the longitude and latitude. Therefore, a unified dimension is needed to reduce dimensional differences. We first transform these three parameters into a Cartesian coordinate system. Then, based on a point-by-point calculation of the Euclidean distance between the two datasets, the MED is obtained by taking the average. The calculation equation is as follows:

$$MED = \frac{1}{n} \sum_{i=1}^n |s_i - \hat{s}_i| \quad (5)$$

where s_i represents the average trajectory point of the generated dataset and \hat{s}_i represents the average trajectory point of the real dataset.

3.3 Model training

During the model training process, the data augmented by the improved CTGAN will be mixed with the processed original ADS-B trajectory data, with a certain proportion of the augmented data representing trajectories under severe weather conditions. This mixture forms a new dataset. The latest mixed dataset will be interpolated at 1-second intervals for learning by LSTM-based models.

This paper uses four commonly used TP models, LSTM, CNN-LSTM, CNN-LSTM-Attention and CNN-BiLSTM, to learn from the new mixed dataset. This is done to validate the impact of data augmentation on the accuracy of aircraft TP.

In this study, the training objective is to minimise the multi-task loss combining:

(1) Trajectory prediction loss:

$$\mathcal{L}_{traj} = \frac{1}{N} \sum_{t=1}^N \|\hat{y}_t - y_t\|_2^2 + \lambda \cdot KL(p(\hat{y}_t) \| p(y_t)) \quad (6)$$

where \hat{y}_t and y_t are the predicted and ground-truth coordinates at time t , respectively. The second term introduces Kullback-Leibler divergence to capture uncertainty in dense traffic scenarios (weighted by $\lambda = 0.1$).

(2) Auxiliary regularisation:

To prevent overfitting on short-term patterns, we incorporate a velocity constraint term:

$$\mathcal{L}_{vel} = \frac{1}{N} \sum_{t=2}^N \|(\hat{y}_t - \hat{y}_{t-1}) - (y_t - y_{t-1})\|_1 \quad (7)$$

The final loss is $\mathcal{L} = \mathcal{L}_{traj} + 0.3 \cdot \mathcal{L}_{vel}$.

3.4 Trajectory prediction

The goal of 4D TP is to predict the spatiotemporal state sequence of an aircraft through a trained model. If every track point is regarded as S_i , the 4D trajectory is regarded as a random variable P , and we assume that T follows the distribution of $P = P(s_{1:N}; \theta)$, the 4D TP problem can be expressed as $P(s_{1:N}; \theta)$. According to the maximum likelihood estimation, the 4D trajectory distribution is expressed as:

$$S_i = \{t, x, y, z, v_{xy}, h, w\} \quad (8)$$

$$T = \{T_1 = s_{1:N}^1, \dots, T_k = s_{1:N}^k, \dots, T_M = s_{1:N}^M\} \quad (9)$$

$$\underset{\theta}{\operatorname{argmax}} P(D|\theta) = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^M P(s_{1:N}^i; \theta) \quad (10)$$

where t represents time, x represents latitude, y represents longitude, z represents geometric height, v_{xy} represents ground speed, h represents track angle and w represents the weather cube; N denotes the number of points in the track and M denotes the number of historical tracks; θ stands for a parameter.

The essence of model training is to learn the distribution of $P(s_{1:N}^i; \theta)$. In this paper, the trained LSTM-based models learn the distribution of $P(s_{1:N}^i; \theta)$. If we input the current aircraft trajectory through the model, we can get a predicted trajectory that corresponds to the input trajectory distribution and follows the time order.

4.0 Experiments and Analyses

4.1 Data preparation

We chose the flight segment near the Hong Kong International Airport (HKIA) as our experimental data. HKIA is one of the busiest airports in the world, handling many flight takeoffs and landings daily and providing abundant data to verify the performance of prediction models. The airspace structure around Hong Kong Airport is complex, containing multiple inbound and outbound flight routes and corridors, which helps test the model's applicability and accuracy in complex environments. Additionally, located in the subtropical region near the South China Sea, Hong Kong is significantly affected by extreme weather, such as typhoons and heavy rain every year, which substantially impacts aviation traffic. Choosing Hong Kong Airport allows effective research into the effects of weather factors on aircraft trajectories and optimising prediction models to improve accuracy under adverse weather conditions.

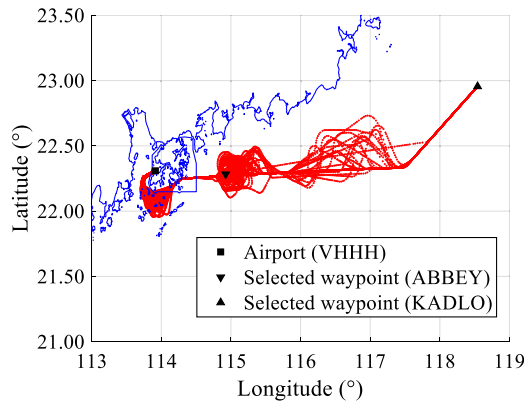


Figure 6. Inbound flight trajectories from KADLO to VHHH in the Hong Kong region.

Therefore, selecting flight segments near Hong Kong Airport as experimental data has practical significance and provides rich and complex experimental conditions, contributing to prediction models' comprehensive and reliable performance.

The ADS-B data in this study comes from the HKO and the OpenSky website, covering flight data with severe weather conditions from 2016 to 2022. The corresponding ET data is obtained from the China Meteorological Administration.

In this dataset, flights from the KADLO waypoint to the VHHH waypoint (airport) account for nearly 1/4 of all data, and flights on this route have a flight time of over 1 hour. Therefore, we focused on experiments and analysis on this route. Due to severe weather conditions, many aircraft did not follow the usual flight routes. Their trajectory data deviated significantly from the expected path and sometimes did not pass through essential waypoints. These significantly deviated trajectories are extreme cases and were not included in this study. We chose to retain trajectory data passing through the ABBEY waypoint on this route, excluding trajectories with significant deviations to ensure the quality of experimental data and the objectivity of model predictions. After processing, we selected 200 trajectories that met the criteria. The Fig. 6 shows a schematic of the trajectory of inbound flights on this route.

4.2 Test environment and model parameter settings

The experiment is performed on a PC with Microsoft Windows 10 (CPU: 12th-Gen Intel(R) Core i9-12900KF at 3.20 GHz, RAM: 64.0 GB, GPU: NVIDIA GTX 3,080). The improved CTGAN-based models were built using Keras and TensorFlow. The parameters of the improved CTGAN architecture are based on our previous research shown in Table 1.

In this study, we conducted a comparative analysis using MATLAB to run prediction algorithms based on LSTM models. The ratio of the training set to the test set was set to 0.8 and 0.2, respectively. We selected four commonly used neural network architectures for comparison: LSTM, CNN-LSTM, CNN-LSTM-Attention and CNN-BiLSTM. The specific parameter settings are shown in Table 2. With these model parameters, we could compare the performance of different neural network models in predicting aircraft trajectories under severe weather conditions and analyse their effects on the experimental dataset.

4.3 Model testing and analysis

4.3.1 Data augmentation validation based on the improved CTGAN model

Since this study introduces a weather module as a condition that differs from previous research, it is necessary to verify which encoder is optimal for the improved CTGAN model. We used three different encoders, One-hot, Target and Leave-one-out, and calculated their respective MMD and MED according

Table 2. *The parameter setting for different LSTM-based models*

Parameter Setting	LSTM	CNN- LSTM	CNN-LSTM- Attention	CNN- BiLSTM
Number of hidden layers	100	200	200	100
Training epochs	1,000	500	500	200
Learning rate	0.005	0.005	0.005	0.001
Learning rate drop period	800	400	400	160
Learning rate drop factor	0.9	0.9	0.9	0.1
Number of convolution layers	–	2	2	2
Convolution kernel size	–	[1,3]	[1,3]	[1,3]
Number of feature maps in first conv layer	–	32	32	32
Number of feature maps in second conv layer	–	64	64	64
Activation function	–	ReLU	ReLU	ReLU
Minibatch size	–	200	100	200
Validation frequency	–	–	10	–

Table 3. *Evaluation results of the improved CTGAN model under different encoding methods*

Evaluation Metrics	One-hot Encoding	Target Encoding	Leave-one-out Encoding
MMD	0.0382	0.0382	0.0382
MED	0.0087	0.0088	0.0086

to formulas (3) and (5). The results are shown in Table 3. Smaller MMD and MED values indicate higher accuracy of the generated trajectories. When both values equal 0, the generated data is identical to the actual data. As shown in Table 3, the MMD values for all three encoding methods are the same, indicating that the differences in data distribution between the generated and actual data are consistent across different encoders. However, the MED value for the Leave-one-out encoder is the smallest, suggesting that it produces the most accurate trajectory distances. Therefore, the Leave-one-out encoder is the most effective.

To intuitively observe the similarity between model-generated and real data distribution, we calculate the probability density distribution using the kernel function as the PDF for all parameters. This difference in probability density distribution visually reflects the distance between the generated and accurate data distribution.

Figure 7 shows the PDFs of the generated data for the variables of time, latitude, longitude, geometric height, ground speed and track angle. The blue line representing the actual data distribution fits well with the red line representing the generated data distribution. This result indicates that the distributions of the actual data and the generated data are highly similar and consistent with the calculated evaluation metrics. Therefore, the generated data can be used for data augmentation. Figure 8 shows the value of the loss function for the generator and discriminator with every 50 epochs for training. It shows that the model has converged.

4.3.2 Model comparison with different data percentages

- (1) Comparison of short-term prediction errors for LSTM-based models under severe weather conditions

LSTM-based models tend to be limited by the chaos phenomenon when predicting long time series data [35]. As the prediction period extends, the output probability density of the expected trajectory tends to become uniform due to the flight uncertainty. Therefore, most LSTM models struggle to provide

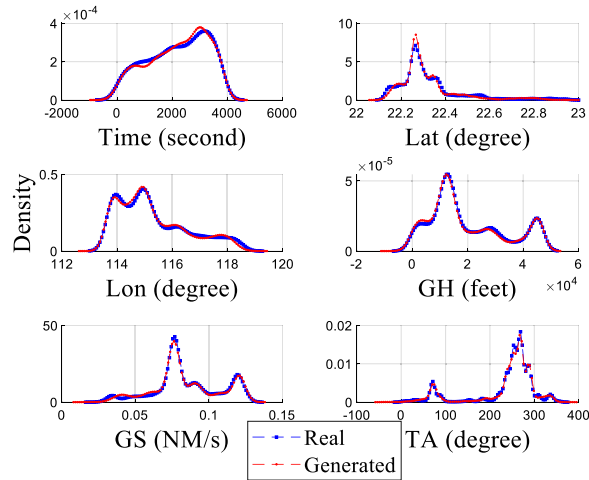


Figure 7. Distribution comparison between the generated and real data.

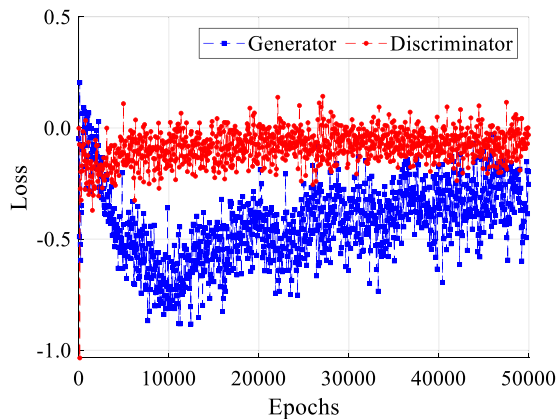


Figure 8. Loss function value of the generator and discriminator for training.

accurate medium to long-term predictions (prediction periods exceeding 10 minutes). Our comparative experiments only compare the short-term prediction accuracy under severe weather conditions.

We set the sliding window of the LSTM-based model to 40 and the prediction step to 5 minutes. Using our proposed improved CTGAN-LSTMs hybrid model, we enhance the proportion of trajectory data under severe weather conditions according to a specific ratio. We then conduct comparative experiments and calculate the average root mean square error (RMSE) of trajectory parameters. The results are shown in Tables 4–7. ‘Lat’ represents latitude, ‘Lon’ represents longitude, ‘GH’ represents geometric height, ‘GS’ represents ground speed and ‘TA’ represents track angle.

To more intuitively compare the impact of data augmentation on the prediction accuracy of aircraft trajectories under severe weather conditions, we plotted the data from Tables 4–8 into line graphs based on latitude, longitude, geometric height, ground speed and track angle, as shown in Fig. 9. These charts allow us to compare the average RMSE of different models under various data percentage conditions, illustrating the effect of data augmentation on short-term prediction tasks under severe weather conditions.

Overall, data augmentation significantly improved the prediction accuracy of all models under severe weather conditions. As the proportion of augmented data increased, the average RMSE of all

Table 4. *RMSE of LSTM model with different data percentages of severe weather conditions*

Data Percentage (%)	Lat (degree)	Lon (degree)	GH (feet)	GS (NM/s)	TA (degree)
4.7	0.0464	0.1947	1665.52	0.00552	63.73
9.1	0.0412	0.1524	1430.24	0.00547	57.62
13.0	0.0380	0.1247	1187.58	0.00536	52.47
16.7	0.0371	0.1184	1202.98	0.00533	53.76
20.0	0.0351	0.1401	1237.33	0.00541	49.50

Table 5. *RMSE of CNN-LSTM model with different data percentages of severe weather conditions*

Data Percentage (%)	Lat (degree)	Lon (degree)	GH (feet)	GS (NM/s)	TA (degree)
4.7	0.05568	0.1847	1808.35	0.00994	83.18
9.1	0.04863	0.1498	1772.69	0.00822	79.03
13.0	0.04219	0.1282	1704.75	0.00693	74.46
16.7	0.04130	0.1197	1306.17	0.00669	70.19
20.0	0.03407	0.1202	1312.64	0.00537	52.39

Table 6. *RMSE of CNN-LSTM-Attention model with different data percentages of severe weather conditions*

Data Percentage (%)	Lat (degree)	Lon (degree)	GH (feet)	GS (NM/s)	TA (degree)
4.7	0.05526	0.1849	2095.77	0.00957	85.65
9.1	0.05081	0.1759	1963.36	0.00848	80.39
13.0	0.04621	0.1423	1644.99	0.00764	75.75
16.7	0.04001	0.1243	1458.42	0.00665	71.29
20.0	0.04070	0.1154	1445.02	0.00699	70.84

Table 7. *RMSE of CNN-BiLSTM model with different data percentages of severe weather conditions*

Data Percentage (%)	Lat (degree)	Lon (degree)	GH (feet)	GS (NM/s)	TA (degree)
4.7	0.05547	0.1807	2165.73	0.00953	84.75
9.1	0.05135	0.1622	1912.20	0.00874	80.92
13.0	0.04595	0.1502	1728.44	0.00776	76.42
16.7	0.04064	0.1268	1454.66	0.00680	72.02
20.0	0.04036	0.1246	1419.62	0.00678	70.34

Table 8. *Comparison of mean horizontal and vertical errors of the LSTM model for flight CPA475; the prediction time span is 5 minutes*

Data Percentage (%)	Mean Horizontal Errors (nmi)	Mean Vertical Errors (m)
4.7	9.8422	413.14
13.0	6.4487	269.73

models decreased. For CNN-LSTM, CNN-LSTM-Attention and CNN-BiLSTM models, the CNN-LSTM model performed the best across most parameters, especially in predicting longitude, geometric height and ground speed. Although introducing attention mechanisms and bidirectional LSTM improved performance for specific parameters, the overall performance of the CNN-LSTM model remained superior.

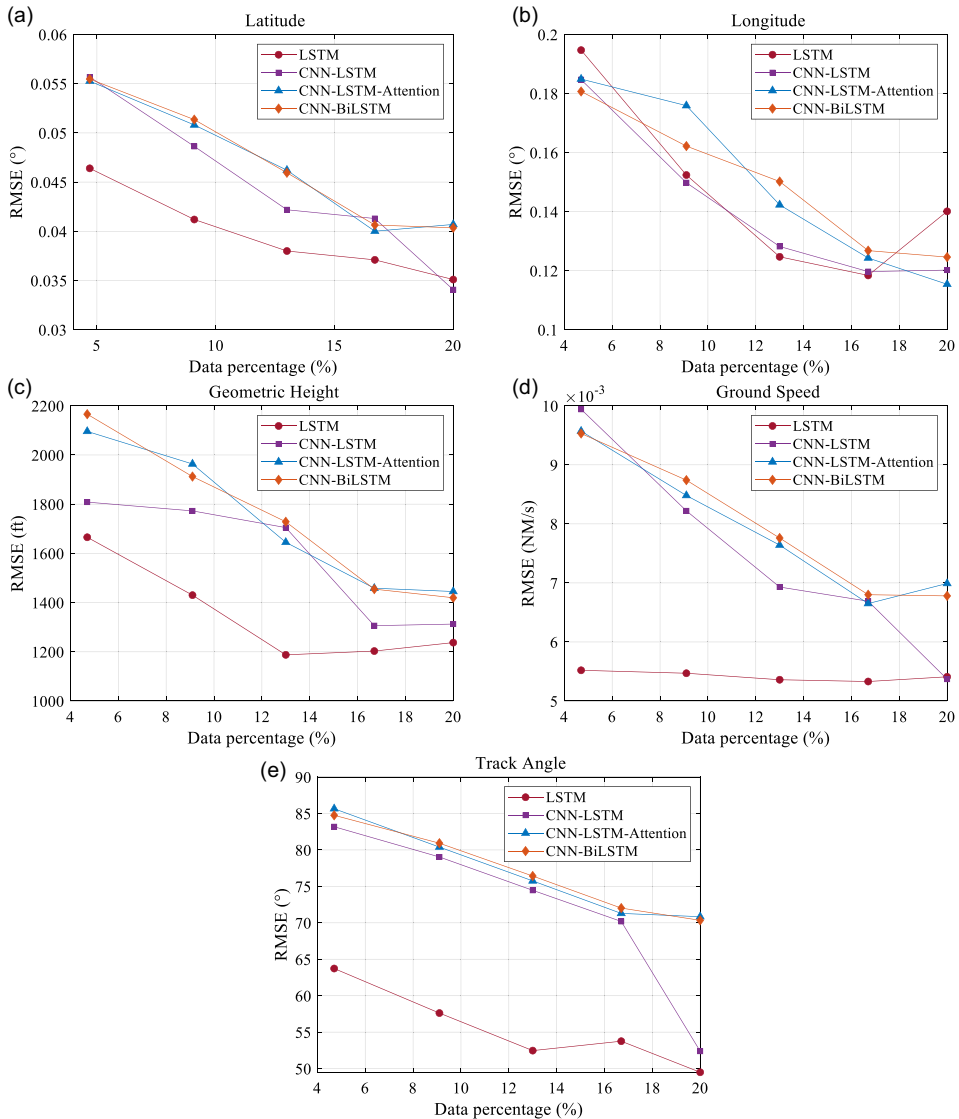


Figure 9. Comparison of the prediction RMSE of the LSTM, CNN-LSTM, CNN-LSTM-Attention and CNN-BiLSTM models for different data percentages.

When comparing CNN-LSTM and LSTM models, despite the CNN-LSTM model showing more significant accuracy improvements with increased data proportions – especially in longitude, geometric height and ground speed predictions – the LSTM model performed well and stably under minor sample conditions for some indicators. Therefore, while the CNN-LSTM model may perform better on comprehensive metrics after data augmentation reaches a certain percentage, using the LSTM model can yield better prediction results when handling minority-class sample severe weather aircraft TP tasks.

From Fig. 9, we can observe that for the LSTM model, the average RMSE of each model decreased with increasing data proportions. The prediction accuracy improvement, around 13%, was the most significant in data augmentation. Specifically, when the data proportion increased from 4.7% to 13%, the RMSE for latitude (Lat) decreased by about 18.10%, longitude (Lon) by approximately 35.93%, geometric height (GH) by around 28.68%, ground speed (GS) by the smallest margin of about 2.90%

Table 9. *RMSE of LSTM model with different data percentages using SMOTE data augmentation*

Data Percentage (%)	Lat (degree)	Lon (degree)	GH (feet)	GS (NM/s)	TA (degree)
4.7	0.0464	0.1947	1665.52	0.00552	63.73
9.1	0.0515	0.1842	1434.17	0.00591	72.76
13.0	0.0434	0.1507	1314.68	0.00552	62.87
16.7	0.0418	0.1706	1326.72	0.00493	60.03
20.0	0.0434	0.1255	1489.81	0.00553	61.65

and track angle (TA) by about 17.68%. On average, the RMSE decreased by about 20.46%, according to the LSTM model’s results. This indicates that increasing the data volume positively impacts model performance, enabling more accurate predictions of target variables.

(2) Visual analysis of typical trajectory prediction

To intuitively compare the effect of our proposed improved CTGAN-LSTM framework on TP tasks under severe weather conditions, we selected flight CPA475 as an example to analyse the TP parameter differences of the LSTM model. Specifically, we analysed and visually compared the changes in data proportions of 4.7% and 13%. The results are shown in Fig. 10. For the LSTM model, when CTGAN enhanced the minority-class sample trajectory data under severe weather conditions, the prediction accuracy for all parameters saw significant improvement. This result is consistent with the findings shown in Table 4. We also calculated the mean horizontal and vertical errors in Table 8 to show the difference further. It should be noted that this comparative experiment uses a prediction time span of 5 minutes, which causes an apparent vertical error. However, this error will be significantly reduced if we shorten the prediction time span.

4.3.3 Comparison of data enhancement quality with SMOTE

The quality of synthetic data is a crucial factor influencing the effectiveness of data augmentation. Therefore, it is necessary to compare the quality of data generated by the improved CTGAN with that generated by the commonly used SMOTE technique. Considering that the LSTM model yields better prediction results when handling minority-class sample TP tasks under severe weather conditions, the comparative experiment was conducted using the LSTM model with the same settings as in Table 2. Here, the number of nearest neighbours for SMOTE was set to 5.

The MMD for the enhanced data generated is 0.0442, and the MED is 0.00075. Compared to the improved CTGAN model (MMD is 0.0382, MED is 0.0087), the data generated by SMOTE has a larger MMD but a more petite MED. This indicates that from the data distribution perspective, the data generated by the improved CTGAN is closer to the actual data distribution. However, regarding trajectory point distance, the data generated by SMOTE has a more minor error.

Table 9 shows the impact of different data percentages on the average RMSE of five variables predicted by the LSTM model after using the SMOTE data augmentation technique. Lat represents latitude, lon represents longitude, GH represents geometric height, GS represents ground speed and TA represents track angle.

For latitude, when the data percentage increases from 4.7% to 9.1%, the RMSE increases from 0.0464 to 0.0515. As the data percentage rises to 13.0% and 16.7%, the RMSE decreases to 0.0434 and 0.0418, respectively. However, when the data percentage increases to 20%, the RMSE rises to 0.0434 again. The RMSE shows a trend of first increasing, then decreasing and slightly increasing, indicating that increasing the data volume and using SMOTE can improve the model’s prediction performance for latitude to a certain extent, but there are fluctuations.

For longitude, when the data percentage increases from 4.7% to 13.0%, the RMSE decreases rapidly from 0.1947 to 0.1507. As the data percentage increases to 16.7%, the RMSE rises to 0.1706. However,

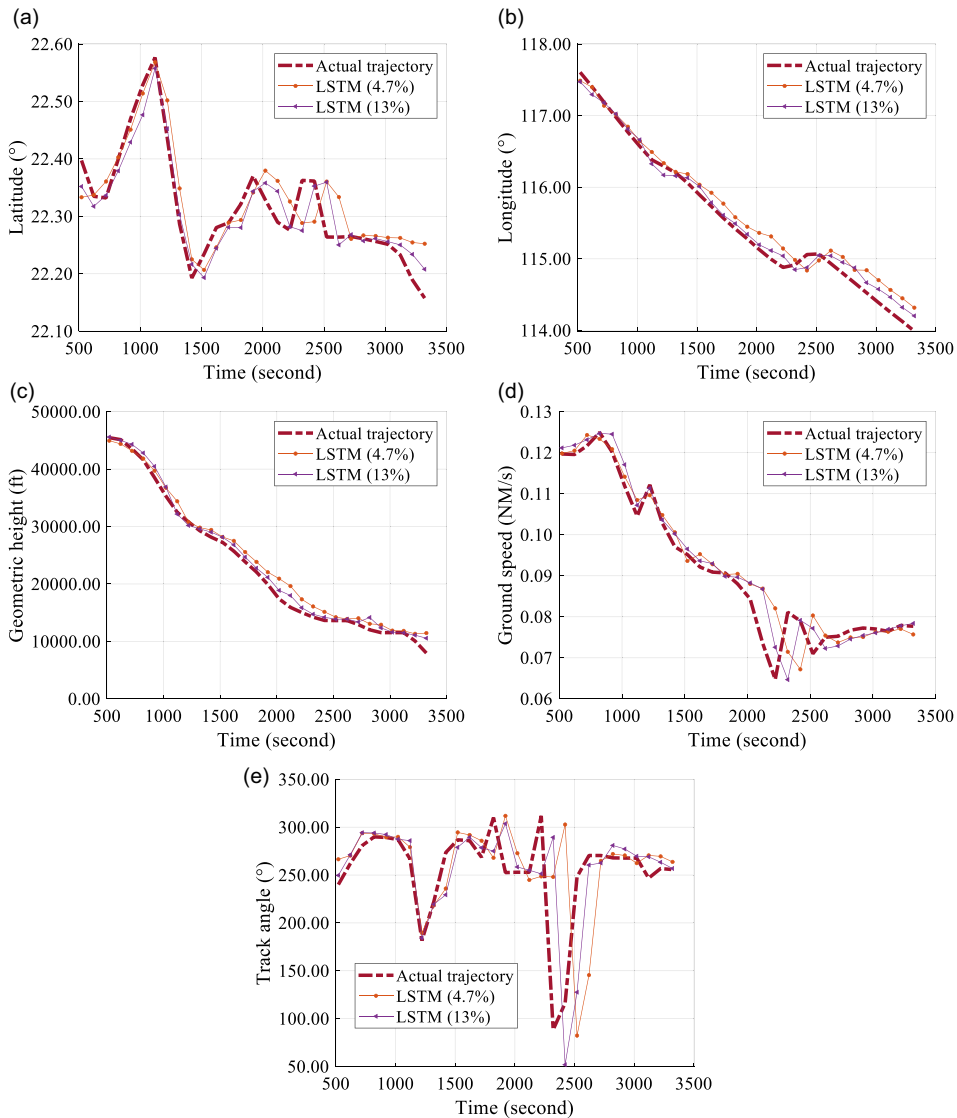


Figure 10. Comparison of the prediction results of the LSTM model for flight CPA475.

as the data percentage increases to 20%, the RMSE decreases again to 0.1255. The RMSE shows a general downward trend as the data percentage increases, but there are significant fluctuations at specific points. These fluctuations may be due to SMOTE's tendency to lose authenticity when handling long time-series data, which in turn affects the performance of the LSTM model.

For geometric height, when the data percentage increases from 4.7% to 13%, the RMSE decreases rapidly from 1665.52 to 1314.68. As the data percentage rises to 16.7%, the RMSE slightly increases to 1326.72, and when the data percentage increases to 20%, the RMSE significantly rises to 1489.81. This indicates that in the early stages of growing data volume, the model's prediction accuracy improves significantly, but after reaching a particular data volume, the prediction accuracy decreases. This may be caused by SMOTE's tendency to generate data that converges locally.

For ground speed, the RMSE trend is similar to that of latitude, showing a trend of first increasing, then decreasing, and slightly increasing. Specifically, at 4.7%, 13.0%, and 20% data percentages, the RMSE is very close, being 0.00552, 0.00552 and 0.00553, respectively. When the data percentage is

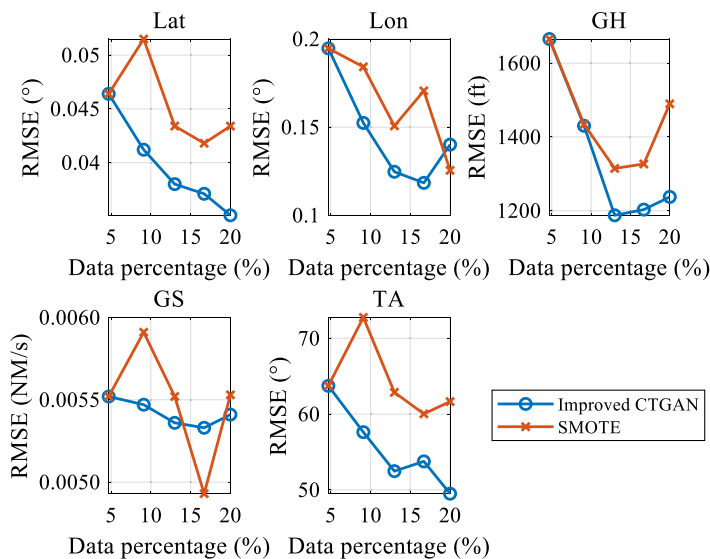


Figure 11. RMSE values of different parameters for various data percentages using improved CTGAN and SMOTE data augmentation techniques.

9.1%, the RMSE increases to 0.00591. At 16.7%, the RMSE decreases to the lowest value of 0.00493. It can be seen that SMOTE augmented data did not significantly improve the LSTM model's prediction performance for GS, especially when the data percentage is 9.1%, the RMSE increases. This indicates that increasing the data volume and using SMOTE have limited effects on the prediction of ground speed and may even cause performance fluctuations and deterioration in some cases.

For track angle (TA), the RMSE trend is the same as that of ground speed. It can be seen that SMOTE augmented data did not significantly improve the LSTM model's prediction performance for track angle.

To visually compare the difference between the improved CTGAN and SMOTE for data augmentation, we plotted the RMSE data from Table 4 and Table 9 for latitude, longitude, geometric height, ground speed and track angle as line graphs, as shown in Fig. 11. It can be observed that, in most cases, the improved CTGAN performs significantly better than SMOTE.

In summary, for minority-class sample trajectory data under severe weather conditions, using the SMOTE data augmentation technique and increasing the data percentage can improve the prediction accuracy of LSTM models for some variables. However, there is inevitable fluctuation in the RMSE changes for certain variables, and increasing the data volume leads to performance degradation at specific data percentages. SMOTE-enhanced data did not significantly improve model performance and even showed performance degradation at some data percentages, particularly for ground speed and track angle. Compared to the improved CTGAN, although SMOTE has higher precision in generating trajectories, it is not a suitable data augmentation model for minority-class sample trajectory data under severe weather conditions. The root cause may be that SMOTE performs data augmentation based on the uniform distribution assumption rather than the actual distribution. At the same time, the improved CTGAN generates synthetic data using the distribution learned from samples.

5.0 Discussions

5.1 Methodological innovations

Our approach utilises an improved CTGAN for data augmentation alongside LSTM models for predictive verification. This hybrid CTGAN-LSTM model effectively modulates the reduced accuracy commonly associated with data imbalance by enabling four-dimensional aircraft TP that accounts for

severe weather impacts. The framework is adaptable, allowing for substitution of CTGAN and LSTM with other augmentation and prediction networks tailored to specific application scenarios, offering a foundation for addressing similar data imbalance issues in aircraft TP.

5.2 Limitations

Despite our innovations, some limitations remain:

- (1) The accuracy of TP is highly dependent on the quality of synthetic data generated during augmentation. Our experiments highlight this dependency, yet there is a lack of established methodologies for accurately assessing the quality of synthetic trajectory data. This represents an area for future research.
- (2) Our study focuses on short-term predictions using LSTM-based models, leaving medium to long-term trajectory predictions under severe weather conditions unexplored. Future research should investigate integrating suitable models for medium to long-term predictions within the proposed framework, replacing LSTM for more comprehensive analysis.

These limitations provide avenues for future research and improvement in aircraft trajectory prediction under severe weather conditions.

6.0 Conclusions and Future Work

Trajectory data under severe weather conditions occupies a small proportion in aircraft trajectory sample data, leading to data imbalance. Current deep neural networks are usually trained using multiple rounds of small-batch random sampling of sample data, which cannot effectively solve the data imbalance problem, and aircraft TP under severe weather conditions is challenging. The proposed aircraft TP framework, which includes a minority-class sample data enhancement network and a prediction network, achieves significant improvements in short-term aircraft TP accuracy under severe weather conditions, demonstrating the effectiveness of the improved CTGAN-LSTMs model. Experimental results show that when the proportion of minority-class sample data increases from 4.7% to 13%, the average RMSE of the CTGAN-LSTM model for 5-minute short-term prediction decreases by 20.46%. Based on this architecture, other data augmentation and prediction networks can be selected according to specific application characteristics to replace CTGAN and LSTM-based models, forming new hybrid models. The idea and method proposed in this paper to solve the problem of sample data imbalance through data enhancement can provide a reference for solving other data imbalance problems in aircraft TP tasks. Given the significant impact of severe weather on ATMs, this study is helpful to enhance ATM capabilities.

We compared four commonly used LSTM-based prediction models (LSTM, CNN-LSTM, CNN-LSTM-Attention and CNN-BiLSTM) to evaluate the performance of the proposed improved CTGAN-LSTM hybrid model. Experimental results show that the improved CTGAN-LSTM hybrid model can significantly improve the accuracy of TP under severe weather conditions. Under this architecture, using an LSTM model as the prediction model demonstrates the best performance for minority-class sample data, with significantly higher TP accuracy compared to other benchmark models. Before the proportion of minority-class sample data, the prediction accuracy is 13%. The accuracy indicators increase rapidly with the proportion, and after gaining 16.7%, the prediction accuracy stabilises.

Comparing CTGAN with SMOTE revealed that CTGAN supports better TP of minority-class data under severe weather. Though SMOTE achieves higher synthetic trajectory accuracy, its assumption of uniform data distribution mismatches actual conditions. Thus CTGAN is more suitable. This suggests that the distribution characteristics of synthetic data are more critical than mere accuracy for trajectory augmentation.

In summary, minority-class sample augmentation enhances aircraft TP accuracy under severe weather without requiring adjustments to validated models. Different augmentation models significantly influence TP model training and prediction based on generated and original trajectories. Evaluating generated trajectory quality and augmentation models merits further exploration. Future research will focus on medium to long-term trajectory prediction to address current limitations.

Acknowledgments. The grant supports from the Hong Kong Research Grants Council (RGC) General Research Fund (GRF) (project code: PolyU 15212622/B-Q94L) and from the Otto Poon Research Institute For Climate-Resilient Infrastructure (RICRI) (project code: ZH8Y) are greatly acknowledged.

References

- [1] International Civil Aviation Organization. Why Global TBO Concept, ICAO. [Online]. <https://www.icao.int/airnavigation/tbo/Pages/Why-Global-TBO-Concept.aspx> (accessed: July 11, 2024).
- [2] Huang, J. and Ding, W. Aircraft trajectory prediction based on Bayesian optimised temporal convolutional network–bidirectional gated recurrent unit hybrid neural network, *Int. J. Aerosp. Eng.*, 2022, **2022**, p 2086904.
- [3] Zeng, W., Chu, X., Xu, Z., Liu, Y. and Quan, Z. Aircraft 4D trajectory prediction in civil aviation: a review, *Aerospace*, 2022, **9**, (2), p 91.
- [4] Shi, Z., Xu, M., Pan, Q., Yan, B. and Zhang, H. LSTM-based flight trajectory prediction, in *Presented at the 2018 International joint conference on neural networks (IJCNN)*, IEEE, 2018, New York City, USA, pp 1–8.
- [5] Degas, A. et al., A survey on artificial intelligence (AI) and explainable AI in air traffic management: current trends and development with future research trajectory, *Appl. Sci.*, 2022, **12**, (3), p 1295.
- [6] Schimpf, N., Wang, Z., Li, S., Knoblock, E.J., Li, H. and Apaza, R.D. A generalized approach to aircraft trajectory prediction via supervised deep learning, *IEEE Access*, 2023, **11**, pp 116183–116195.
- [7] International Civil Aviation Organization, Meteorological Service for International Air Navigation Annex 3, ICAO, 2010. [Online]. <https://www.icao.int/airnavigation/IMP/Documents/Annex%203%20-%202075.pdf>
- [8] Porretta, M., Dupuy, M.-D., Schuster, W., Majumdar, A. and Ochieng, W. Performance evaluation of a novel 4D trajectory prediction model for civil aircraft, *J. Navig.*, 2008, **61**, (3), pp 393–420.
- [9] Tang, X., Chen, P. and Zhang, Y. 4D trajectory estimation based on nominal flight profile extraction and airway meteorological forecast revision, *Aerosp. Sci. Technol.*, 2015, **45**, pp 387–397.
- [10] Ma, L. and Tian, S. A hybrid CNN-LSTM model for aircraft 4D trajectory prediction, *IEEE Access*, 2020, **8**, pp 134668–134680.
- [11] Liu, Y. and Hansen, M. Predicting aircraft trajectories: a deep generative convolutional recurrent neural networks approach,” *ArXiv Prepr. ArXiv:1812.11670*, 2018.
- [12] Ayhan, S. and Samet, H. Aircraft trajectory prediction made easy with predictive analytics, in *Presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, The Association for Computing Machinery (ACM), 2016, New York City, NY, USA, pp 21–30.
- [13] Pang, Y., Xu, N. and Liu, Y. Aircraft trajectory prediction using LSTM neural network with embedded convolutional layer, in *Presented at the Proceedings of the Annual Conference of the PHM Society*, PHM Society, Scottsdale, AZ, USA, 2019, pp 1–8.
- [14] Pang, Y. and Liu, Y. Conditional generative adversarial networks (CGAN) for aircraft trajectory prediction considering weather effects, in *Presented at the AIAA Scitech 2020 Forum*, The American Institute of Aeronautics and Astronautics (AIAA), 2020, Reston, Virginia, p 1853.
- [15] Pang, Y., Zhao, X., Yan, H. and Liu, Y. Data-driven trajectory prediction with weather uncertainties: a Bayesian deep learning approach, *Transp. Res. Part C Emerg. Technol.*, 2021, **130**, p 103326.
- [16] Majeed, A. and Hwang, S.O. CTGAN-MOS: conditional generative adversarial network based minority-class-augmented oversampling scheme for imbalanced problems,” *IEEE Access*, 2023, **11**, pp 85878–85899.
- [17] Spelman, V.S. and Porkodi, R. A review on handling imbalanced data, in *Presented at the 2018 international conference on current trends towards converging technologies (ICCTCT)*, IEEE, 2018, New York City, USA, pp 1–11.
- [18] Cheng, K., Zhang, C., Yu, H., Yang, X., Zou, H. and Gao, S. Grouped SMOTE with noise filtering mechanism for classifying imbalanced data, *IEEE Access*, 2019, **7**, pp 170668–170681.
- [19] Pei, W., Xue, B., Zhang, M., Shang, L., Yao, X. and Zhang, Q. A survey on unbalanced classification: how can evolutionary computation help?, *IEEE Trans. Evol. Comput.*, 2023, **28**, pp 353–373.
- [20] Vondrick, C., Pirsivash, H. and Torralba, A. Generating videos with scene dynamics, *Adv. Neural Inf. Process. Syst.*, 2016, **29**, pp 613–621.
- [21] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B. and Lee, H. Generative adversarial text to image synthesis,” in *Presented at the International Conference on Machine Learning*, New York, NY, PMLR, 2016, pp 1060–1069.
- [22] Brock, A., Donahue, J. and Simonyan, K. Large scale GAN training for high fidelity natural image synthesis, *ArXiv Prepr. ArXiv:1809.11096*, 2018.
- [23] Ding, W., Wang, W. and Zhao, D. Multi-vehicle trajectories generation for vehicle-to-vehicle encounters, in *Presented at the 2019 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 4255–4261.

- [24] Zhang, L., Chen, H., Jia, P., Tian, Z. and Du, X. WGAN-GP and LSTM based prediction model for aircraft 4-D trajectory, in *Presented at the 2022 International Wireless Communications and Mobile Computing (IWCMC)*, IEEE, 2022, New York City, USA, pp 937–942.
- [25] Xu, L., Skoularidou, M., Cuesta-Infante, A. and Veeramachaneni, K. Modeling tabular data using conditional GAN, *Adv. Neural Inf. Process. Syst.*, **32**, 2019, pp 7335–7345.
- [26] An, C., Sun, J., Wang, Y. and Wei, Q. A K-means improved CTGAN oversampling method for data imbalance problem, in *Presented at the 2021 IEEE 21st International Conference on Software Quality, Reliability and Security (QRS)*, IEEE, 2021, New York City, USA, pp 883–887.
- [27] Alqarni, A.A. and El-Alfy, E.-S.M. Improving intrusion detection for imbalanced network traffic using generative deep learning, *Int. J. Adv. Comput. Sci. Appl.*, 2022, **13**, (4), pp 959–967.
- [28] Xu, L. and Veeramachaneni, K. Synthesizing tabular data using generative adversarial networks, *ArXiv Prepr. ArXiv:1811.11264*, 2018.
- [29] Cote, M.-P., Hartman, B., Mercier, O., Meyers, J., Cummings, J. and Harmon, E. Synthesizing property & casualty ratemaking datasets using generative adversarial networks, *ArXiv Prepr. ArXiv:2008.06110*, 2020.
- [30] Zhang, H. and Liu, Z. Four-dimensional aircraft trajectory prediction based on generative deep learning, *J. Aerosp. Inf. Syst.*, 2024, **21**, (7), pp 554–567.
- [31] Ali, B.S., Ochieng, W.Y. and Zainudin, R. An analysis and model for automatic Dependent surveillance broadcast (ADS-B) continuity, *GPS Solut.*, 2017, **21**, pp 1841–1854.
- [32] DeLaura, R., Robinson, M., Pawlak, M. and Evans, J. Modeling convective weather avoidance in enroute airspace, in *Presented at the 13th Conference on Aviation, Range, and Aerospace Meteorology*, AMS, New Orleans, LA, 2008, pp 1–12.
- [33] Nugraha, R.A., Pardede, H.F. and Subekti, A. Oversampling based on generative adversarial networks to overcome imbalance data in predicting fraud insurance claim: 10.48129/kjs.splml.19119, *Kuwait J. Sci.*, 2022.
- [34] Borji, A. Pros and cons of GAN evaluation measures, *Comput. Vis. Image Underst.*, 2019, **179**, pp 41–65.
- [35] Lin, L., Li, W., Bi, H. and Qin, L. Vehicle trajectory prediction using LSTMs with spatial–temporal attention mechanisms, *IEEE Intell. Transp. Syst. Mag.*, 2021, **14**, (2), pp 197–208.