FlightBERT++: A Non-autoregressive Multi-Horizon Flight Trajectory Prediction Framework

Dongyue Guo, Zheng Zhang, Zhen Yan, Jianwei Zhang, Yi Lin*

College of Computer Science, Sichuan University, Chengdu 610000, China

{dongyueguo, zhaeng}@stu.scu.edu.cn, tankzhen@163.com, {zhangjianwei, yilin}@scu.edu.cn

Abstract

Flight Trajectory Prediction (FTP) is an essential task in Air Traffic Control (ATC), which can assist air traffic controllers in managing airspace more safely and efficiently. Existing approaches generally perform multi-horizon FTP tasks in an autoregressive manner, thereby suffering from error accumulation and low-efficiency problems. In this paper, a novel framework, called FlightBERT++, is proposed to i) forecast multi-horizon flight trajectories directly in a non-autoregressive way, and ii) improve the limitation of the binary encoding (BE) representation in the FlightBERT. Specifically, the FlightBERT++ is implemented by a generalized encoder-decoder architecture, in which the encoder learns the temporal-spatial patterns from historical observations and the decoder predicts the flight status for the future horizons. Compared with conventional architecture, an innovative horizon-aware contexts generator is dedicatedly designed to consider the prior horizon information, which further enables non-autoregressive multi-horizon prediction. Moreover, a differential prompted decoder is proposed to enhance the capability of the differential predictions by leveraging the stationarity of the differential sequence. The experimental results on a real-world dataset demonstrated that the FlightBERT++ outperformed the competitive baselines in both FTP performance and computational efficiency.

Introduction

Flight Trajectory Prediction (FTP) is essential for Air Traffic Management (ATM), enabling many critical applications to help Air Traffic Controllers (ATCOs) manage airspace more safely and efficiently, such as traffic flow prediction (Lin, wei Zhang, and Liu 2019; Yan et al. 2023), conflict detection (Liu and Hwang 2011; Chen, Guo, and Lin 2020; Li et al. 2019), and arrival time estimation (Wang, Liang, and Delahaye 2018; Ma et al. 2023). In this context, the FTP technique has gathered more attention from researchers and achieved significant progress in the past decade.

In general, FTP in the Air Traffic Control (ATC) domain aims to forecast the flight status in future time steps according to the observed historical flight trajectories. Currently, the short FTP tasks are commonly formulated as a time series forecasting problem, which predicts future flight trajectories based on real-time observations. According to modeling approaches, the FTP methods can be classified into three distinct categories: physical models, filter-based models, and data-driven models. Physical models typically establish a set of mathematical equations based on kinematics and aerodynamics assumptions to estimate the future flight status (Tang et al. 2015; Benavides et al. 2014; FAA 2010), while the filter-based models estimate the flight trajectory iteratively via a predefined system model by considering real-time measurements (Kalman 1960; Yan et al. 2013; Thipphavong et al. 2013). However, limited by the fixed inference rules, these approaches easily suffer from error accumulation problems and not be suitable for multi-horizon FTP tasks. In contrast, data-driven models aim to learn flight transition patterns from historical data, which have emerged as the predominant approach in modern ATC applications due to remarkable performance (Wang, Liang, and Delahaye 2017; Zeng et al. 2020).

Among the data-driven models, the FlightBERT is a pioneering effort that proposes binary encoding (BE) representation to encode the attributes of the trajectory points into a set of binary vectors, further formulating the FTP task as multiple binary classification (MBC) problem (Guo et al. 2023). Benefiting from the innovative idea of BE representation, the FlightBERT framework not only enhances the semantic representation of the trajectory points, but also avoids the vulnerability impacted by the normalization algorithms. However, two primary limitations should be further addressed to enhance the performance of the FTP task.

- A limitation of the BE representation is that the high-bit prediction errors will lead to outliers in the predictions. For instance, given the BE representation "0110 0100" (decimal 100), the absolute error is 128 (decimal) if the misclassification occurs in the 8^{th} bit ("1110 0100") while that is 1 (decimal) if the prediction error occurs in 1^{st} bit ("0110 0101").
- The FlightBERT performs multi-horizon prediction recursively, i.e., predicts the flight status of the next time step based on observation and iteratively applies the predicted values as pseudo-observation to obtain multi-horizon predictions. It is easy to suffer from larger accumulative errors since the prediction errors in the pseudo-observation will be accumulated during the recursive inference process. Additionally, the computational effi-

*Corresponding author

ciency is limited by the prediction horizon due to the step-by-step prediction paradigm. As the prediction horizon increases, the inference speed will experience a severe reduction, leading to unacceptable delays for real-time applications (e.g., conflict detection).

Focusing on the aforementioned challenges, in this paper, a non-autoregressive multi-horizon FTP framework, named FlightBERT++, is proposed to improve the performance of the FTP task. Thanks to the superior trajectory representation ability of the FlightBERT, the proposed framework inherits the BE representation from the FlightBERT, and is also implemented based on the MBC paradigm.

In order to overcome the outlier predictions resulting from the high-bit prediction errors in the BE representation, a differential prediction paradigm is introduced into the FlightBERT++ framework. Specifically, instead of predicting the absolute values, the differential values of the trajectory attributes are formulated as the output objective in the proposed framework, which can be encoded into BE representations in less bits. In this way, the FlightBERT++ framework is able to effectively mitigate the occurrence of extremely unreliable outliers in the predictions.

To achieve high-accuracy and -efficiency multi-horizon FTP prediction, a novel sequence-to-sequence (Seq2Seq) based encoder-decoder architecture is proposed to implement the FlightBERT++ framework. Specifically, the FlightBERT++ is constructed by a cascading trajectory encoder, Horizon-aware Context Generator (HACG), and Differential-prompted Decoder (DPD). The trajectory encoder encodes the observations into a trajectory-level representation, while the HACG is designed to produce multi-horizon context representations of the predicted horizons. Benefiting from the proposed HACG module, unlike conventional Seq2Seq architecture, the proposed framework can generate multi-horizon predictions directly (non-autoregressive) rather than perform recursive inference. In succession, the differential-prompted decoder is innovatively proposed to perform the high-confidence predictions based on the multi-horizon context representations, which employ the differential sequence of the observations as the prompting. The Transformer blocks are applied to build the backbone network for both the trajectory encoder and DPD, enabling non-autoregressive inference in the temporal modeling process. In this way, the proposed framework can not only mitigate the error accumulation but also yield a substantial improvement in computational efficiency for multi-horizon FTP tasks.

The proposed framework is evaluated on a real-world flight trajectory dataset from an industrial ATC system. To validate the effectiveness and efficiency of the proposed framework, several competitive baselines are selected to conduct comprehensive comparisons. In addition, extensive ablation studies and insightful analysis are also performed to confirm all proposed technical improvements. The experimental results consistently demonstrated that the proposed framework efficiently addresses the outliers and error accumulation problems, outperforming the comparative baselines in both FTP accuracy and efficiency. In summary, the contributions and novelty of this work are listed as follows:

- A flight trajectory prediction framework, called FlightBERT++, is innovatively proposed to perform high-accuracy and -efficiency multi-horizon forecasting in a non-autoregressive manner.
- The differential prediction paradigm is dedicatedly designed to overcome the limitations of the high-bit prediction errors in the BE representation.
- A HACG is innovatively proposed to generate multi-horizon context representations by leveraging prior horizon knowledge, which is the key to supporting non-autoregressive predictions.
- Considering the stationarity of the differential sequence in flight trajectory, a differential-prompted decoder is proposed to facilitate the learning of transition patterns in trajectory sequences, which further improves the performance of the FlightBERT++.

Methodology

Problem Formulation

In general, the short-term FTP task can be formulated as a multi-variable time-series forecasting problem, i.e., predicting the aircraft status of several future horizons based on historical observations. Let an observed trajectory point p_t in timestamp t, the multi-horizon FTP aims to forecast the $P_{t+1:t+n} = \{p_{t+1}, p_{t+2}, ..., p_{t+n}\}$ based on the observation sequence $O_{t-k+1:t} = \{p_{t-k+1}, ..., p_{t-1}, p_t\}$, as Eq. (1):

$$P_{t+1:t+n} = \{p_{t+1}, p_{t+2}, ..., p_{t+n}\} = \mathcal{F}(O_{t-k+1:t}) \quad (1)$$

where the n, k are the number of prediction horizons and the observed sequence length, respectively. $\mathcal{F}(\cdot)$ denotes the learnable FTP model. Compared to the conventional iterative multi-horizon FTP approaches, $\mathcal{F}(\cdot)$ outputs the multi-horizon predictions directly in the FlightBERT++.

In this work, the trajectory point p_t is formulated as a collection of six key attributes for aircraft status, as Eq. (2):

$$p_t = [Lon_t, Lat_t, Alt_t, Vx_t, Vy_t, Vz_t] \quad (2)$$

where the $Lon_t, Lat_t, Alt_t, Vx_t, Vy_t, Vz_t$ represents the longitude, latitude, altitude, and velocity in x, y, z dimensions (corresponds to the longitudinal, latitudinal, and altitudinal attributes), respectively.

The Proposed Framework

Overview The proposed framework is illustrated in Figure 1. The neural architecture of the proposed framework is cascaded by trajectory encoder, horizon-aware context generator, and differential-prompted decoder. Given the observed trajectory sequence $O_{t-k+1:t}$, the object of the trajectory encoder is to learn the temporal-spatial correlations of the observations and abstract them to a high-dimensional representation $\mathbf{Traj}_{enc}$, as Eq. (3).

$$\mathbf{Traj}_{enc} = \text{TrajectoryEncoder}(O_{t-k+1:t}) \quad (3)$$

In succession, the HACG is designed to generate the multi-horizon context representations $C_{t+1:t+n} =$

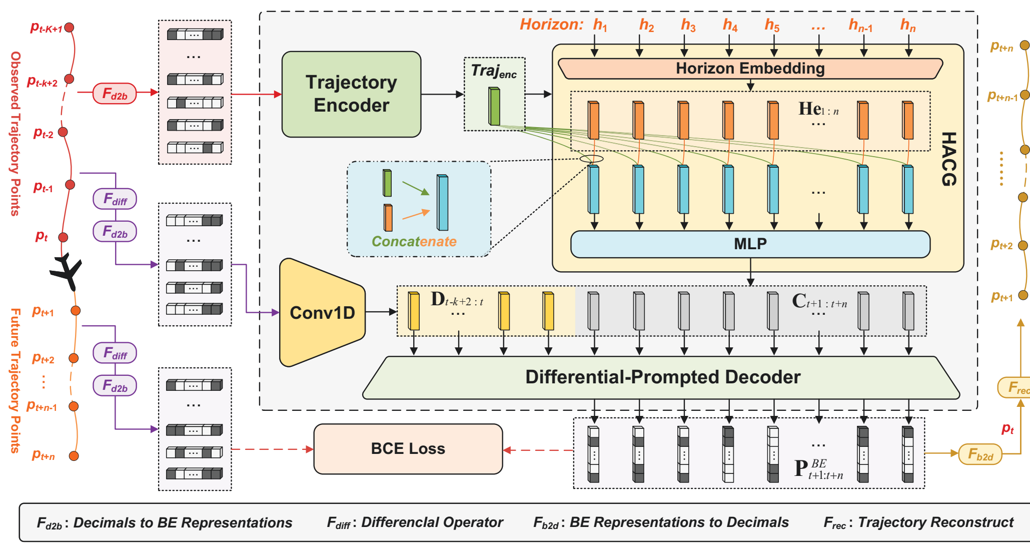


Figure 1: Overview of the proposed FlightBERT++ framework.

$\{c_{t+1}, c_{t+2}, ..., c_{t+n}\}$ by both considering the prior prediction horizon encodings $\mathbf{H} = \{h_1, h_2, ..., h_n,\}$ and the high-level representation $\mathbf{Traj}_{enc}$.

$$\mathbf{C}_{t+1:t+n} = \mathrm{HACG}([\mathbf{Traj}_{enc}, \mathbf{H}]) \tag{4}$$

The DPD receives two input vectors, i.e., the differential embeddings $\mathbf{D}_{t-k+2:t}$ and multi-horizon context representations $\mathbf{C}_{t+1:t+n}$. Specifically, the differential embeddings $\mathbf{D}_{t-k+2:t}$ of the $O_{t-k+1:t}$ are extracted by a Conv1D neural network, which severs as the prompting to learn the transition patterns of the differential sequence. Then, the extracted $\mathbf{D}_{t-k+2:t}$ and multi-horizon context representations $\mathbf{C}_{t+1:t+n}$ are jointly fed into the differential-prompted decoder to generate the final outputs as Eq. (5), where the $\mathbf{P}^{BE}_{t+1:t+n}$ is the BE representation of the predicted differential sequence of future trajectory. Finally, the $\mathbf{P}^{BE}_{t+1:t+n}$ is transformed into decimals to reconstruct the predicted trajectory based on the O_t.

$$\mathbf{P}^{BE}_{t+1:t+n} = \mathrm{DPD}([\mathbf{D}_{t-k+2:t}, \mathbf{C}_{t+1:t+n}]) \tag{5}$$

Note that the inputs and the outputs of the FlightBERT++ are both the BE representation of the trajectory attributes. Therefore, the optimizing objective of the FlightBERT++ is also formulated as the MBC task. The FlightBERT++ framework is trained using the Binary Cross Entropy (BCE) loss function, as that in MBC tasks. More details of the BE representations can be found in (Guo et al. 2023).

Compared with existing multi-horizon FTP approaches, the proposed framework (i) innovatively design a HACG to generate multi-horizon contexts directly, (ii) employ the Transformer-based architecture to conduct the backbone network of the trajectory encoder and differential-prompted decoder, which are the core ideas in enabling non-autoregressive prediction.

Trajectory Encoder As illustrated in Figure 1, the trajectory encoder is composed of three modules, including

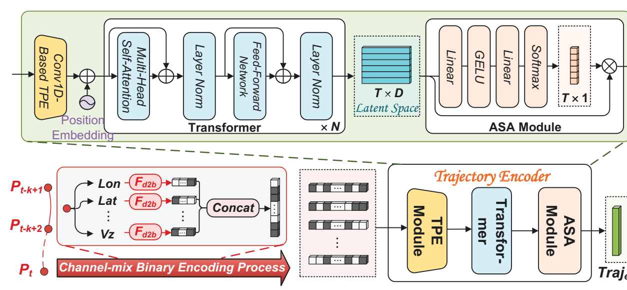


Figure 2: The detailed implementation of the channel-mix trajectory point embedding and trajectory encoder.

the Conv1D-based Trajectory Point Embedding (TPE) module, Transformer-based temporal modeling (TTM) module, and Attention-based Sequence Aggregation (ASA) module. Specifically, the Conv1D-based TPE module projects the BE representation into high-dimensional embedding space to learn discriminative spatial features of the trajectory points, while the Transformer-based module is employed to capture the temporal correlations of the observation sequence. In succession, the outputs of the Transformer module are further fed into the ASA module to generate the trajectory-level embedding and extract the semantic representation over the whole observation sequences.

To learn the informative spatial features of the trajectory points, a Conv1D-based channel-mix trajectory point embedding module is proposed to project the BE representations of trajectory points into embedding space. As illustrated in Figure 2, for a trajectory point, a joint embedding is obtained by concatenating its attribute-wise BE representations, which is further fed into a Conv1D layer to learn trajectory point embedding. The channel-mix strategy not only retains the global correlations of the trajectory attributes, but also is beneficial to learn the local features among the bits of the joint BE representation.

The TTM module is implemented by the stacked Transformer blocks. Compared with the RNN-based architectures, the Transformer is primarily implemented by the self-attention mechanism, which enables the model to learn the temporal correlations of the observations in a non-autoregressive manner. The ASA module performs a weighted sum operation in the temporal dimension, in which the attention weights of the trajectory points are generated as Figure 2. In this way, the trajectory-level embedding $\mathbf{Traj}_{enc}$ can be obtained from the trajectory encoder, which is expected to capture the temporal-spatial semantic representation of the observations.

Horizon-Aware Context Generator In general, the autoregressive nature of conventional multi-horizon FTP approaches degrades the efficiency in the training and inference process, especially in high temporal resolution conditions, which may be not suitable for real-time forecasting scenarios. Towards this gap, a horizon-aware context generator is innovatively designed to support multi-horizon predictions in a non-autoregressive manner by generating multi-horizon contexts directly.

The architecture of the HACG is illustrated in Figure 1. Specifically, we conduct multi-horizon contexts by considering both the prior temporal specificities of the predicted horizons and trajectory-level embeddings of the observations. The inference rules of the proposed HACG can be shown below. Firstly, the predicted horizons are represented by a set of integer tokens, which are further encoded into corresponding one-hot vectors $\mathbf{H}$. Secondly, as shown in Eq. (6), these one-hot vectors are projected into embedding space to learn the horizon embeddings $\mathbf{He}_{1:n} = \{\mathbf{he}_1, \mathbf{he}_2, ..., \mathbf{he}_n\}$ by horizon embedding module.

$$\mathbf{He}_{1:n} = \mathrm{HorizonEmbedding}(\mathbf{H}) \tag{6}$$

In succession, the horizon embeddings and the trajectory-level embedding $\mathbf{Traj}_{enc}$ are concatenated to generate the context vector $\mathbf{hc}_{t+i}$ of horizon i.

$$\mathbf{hc}_{t+i} = \mathrm{Concat}[\mathbf{Traj}_{enc}, \mathbf{he}_i] \tag{7}$$

Finally, the context vectors are further fed into an MLP block to perform high-dimensional projection and generate the final multi-horizon context representations $\mathbf{C}_{t+1:t+n}$.

$$\mathbf{C}_{t+1:t+n} = \mathrm{MLP}([\mathbf{hc}_{t+1}, ..., \mathbf{hc}_{t+i}, ..., \mathbf{hc}_{t+n}]) \tag{8}$$

The core idea of the HACG is to leverage the trajectory-level embedding $\mathbf{Traj}_{enc}$ and horizon embeddings $\mathbf{He}_{1:n}$ to generate informative multi-horizon context representations. Specifically, the $\mathbf{Traj}_{enc}$ can be regarded as a high-dimensional representation that implies the global features of the observations, while the horizon embeddings $\mathbf{He}_{1:n}$ provide the different semantic representations for each predicted horizon. Based on this assumption, the concatenation operation and MLP block are employed to fuse these vectors and generate the multi-horizon context representations. By integrating the prior horizon knowledge, the HACG can be aware of different horizons and generate the multi-horizon context representations directly via only one-pass inference.

Differential-prompted Decoder To mitigate the high-bits prediction errors of the BE representation, the differential prediction paradigm is introduced into the FlightBERT++ framework, i.e., the objective of the decoder is to predict the differential values instead of the raw absolute values. However, it is challenging to learn the transition patterns of differential sequence from the observations sequence because the differential operation may ignore some geographical and kinematical features of the trajectory attributes. To this end, as illustrated in Figure 1, a DPD is proposed to reduce the learning difficulty of the network by integrating the differential prompted mechanism.

Specifically, the DPD consists of two modules, i.e., masked-Transformer, and predictor. Firstly, the differential sequence is calculated from the observed sequence and encoded into BE representations which can be consistent with the form of the outputs. Similar to the TPE module, these BE representations are further fed into a Conv1D-based embedding layer to learn the high-dimensional differential embeddings $\mathbf{D}_{t-k+2:t}$. Secondly, the $\mathbf{D}_{t-k+2:t}$ is concatenated with the multi-horizon context representations $\mathbf{C}_{t+1:t+n}$ along the temporal dimension as a prompt to learn the transition patterns of the predicted differential sequence. In succession, the concatenated vectors further fed into the masked-Transformer module to build the inter- and intra-temporal correlations across observation and the multi-horizon predictions. The architecture of the masked-Transformer module is similar to the TTM module except it employs the masked self-attention mechanism to ensure the temporal specificities of the sequence. Finally, the predictor is composed of a Fully Connected (FC) layer and the Sigmoid activation, which is applied to output BE representations for the multi-horizons.

Experiments

Experimental Settings

Dataset Preprocessing and Description To validate the proposed framework, a real-world flight trajectory dataset was collected from an ATC system. The dataset contains a total of 9 days of trajectory data with 20-second intervals, in which the range of the interested (ROI) longitude and latitude are $[94.616°, 113.689°]$ and $[19.305°, 37.275°]$, respectively. The key attributes of the flight trajectory are extracted from the raw data to support the experiments, including timestamps, call sign, longitude, latitude, altitude, and velocity in x, y, z directions. After preprocessing, a total of 8643 flight trajectories in the dataset that further split into train, validate, and test subsets. Specifically, the trajectory of the first 7 days is used to train the FTP models while 8^{th} and 9^{th} are for validation and testing, respectively.

Comparison Baselines In this work, a total of 5 competitive approaches serve as baselines to validate the effectiveness of the proposed FlightBERT++. Moreover, the baselines are divided into two groups to conduct comparisons by multi-horizon prediction styles. Group A performs iterative multi-horizon prediction, i.e., the model only predicts the result for the next time step that further serves as a pseudo-observation for multi-horizon inference, including LSTM (Shi et al. 2018), Transformer (Vaswani et al. 2017), Kalman-Filter (Kalman 1960), FlightBERT (Guo et al. 2023). Group B is a direct multi-horizon prediction approach, i.e., forecasts the results in a single inference process for multiple time steps, which is implemented by the LSTM+Attention model (Bahdanau, Cho, and Bengio 2015).

Experimental Configuration Based on ATC work, the quantization precision $0.001°$ (about 110 meters) is selected to adapt the transformation between BE representation and the decimals. In this resolution, the input of the proposed framework is a 78-dimensional vector (BE representation) while the output is a 48-dimensional vector. The embedding size of the trajectory points, horizons, and differentials is set to 128, i.e., 128 channels for the Conv1D networks in the proposed framework. The number of the Transformer blocks in the trajectory encoder and differential-prompted decoder are both set to 4. The number of hidden states of the Transformer blocks is set to 768. An attention operator with 4 heads is applied to all Transformer blocks in the proposed framework. In the experiments, we use the latest 3-minute observations to predict the flight status of the future 5 minutes, i.e., predicting 15 trajectory points based on 9 observed trajectory points. The Adam optimizer with an initial learning rate of 10^{-4} is applied to train all the above deep learning-based models.

Evaluation Metrics

In this work, the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) are applied to evaluate the proposed methods and baselines, which are the common criteria for the FTP tasks. In addition, to intuitively validate the FTP performance in the three-dimensional (3D) airspace, the Mean Distance Error (MDE) is proposed to evaluate the Euclidean distance of the predictions and the actual trajectory points. Specifically, the positional attributes of both predictions and ground truth are projected to the earth-centered and earth-fixed (ECEF) coordinate system, and the distance error is measured to evaluate the model performance as below:

$$MDE = \frac{1}{N}\frac{1}{h}\sum_{i=1}^{N}\sum_{j=1}^{h}\Phi(p_{ij} - p'_{ij}) \tag{9}$$

where p, p' are the transformed values of ground truth and prediction in the ECEF coordinate system. $\Phi(\cdot)$ is the calculation function of Euclidean distance in the 3D airspace.

Furthermore, the Mean Time Costs (MTC) metric is proposed to evaluate the computational performance, which is defined as follows:

$$MTC = \frac{1}{N}\sum_{i=1}^{N} time\_costs_i^h \tag{10}$$

where N is the number of samples in the test process, $time\_costs_i^h$ represents the time cost for h prediction horizons of sample i. In this phase, the batch size of all evaluation models is set to 1 to ensure comparison fairness.

Result and Discussions

Results and Quantitative Analysis

Overall Performance of FTP Table 1 reports the overall performance of the proposed framework and comparison baselines. To investigate the robustness of these models with the horizons increase, the experimental results are divided into four (1, 3, 9, 15) different horizons, corresponding to 20 seconds, 1, 3, and 5 minutes trajectories in the future. It is demonstrated that the proposed FlightBERT++ framework achieves significant performance improvements against FlightBERT and outperforms other baselines in the MAE metric across all predicted attributes. For the RMSE metrics, thanks to the reduction of the BE representation bits, great RMSE reductions of the longitude, latitude, and altitude are obtained in the proposed framework. In addition, benefiting from the dedicated network design of the proposed framework, the error accumulation of the multi-horizon predictions is significantly decreased compared to competitive baselines.

For the iterative multi-horizon prediction approaches, the KF-based model-driven approach suffers from a huge performance degradation with the increasing of prediction horizon due to the lack of real observations. Compared with KF-based approach, the data-driven based approaches achieved better performance across all prediction horizons. However, these approaches also suffer from error accumulation problems significantly in multi-horizon inference procedures. Benefiting from the capacity of the BE representation, the FlightBERT harvests comparable MAE performance in all prediction horizons. However, the RMSE of the FlightBERT is higher than other data-driven baselines due to the outliers caused by high-bit errors in the prediction. The LSTM+Attention achieves superior performance than iterative multi-horizon prediction approaches for long-term horizon predictions (in horizons 9 and 15). However, it also suffers from the performance degradation of the short-term horizons (horizon 1) in both the MAE and RMSE metrics.

In summary, the direct multi-horizon prediction approaches are superior to the iterative approaches in long-term prediction steps since the global dependencies are learned from the training process. Thanks to the robust temporal-spatial dependencies obtained by the trajectory encoder and differential-promoted decoder, the FlightBERT++ framework achieves expected performance in both short- and long-term prediction horizons, which further demonstrates the effectiveness of the network design.

Computational Performance Evaluation The computational performance and the size of the model parameters are presented in Table 2. In this experiment, the prediction horizon is set to 15 for all models. As can be seen from the results, it is evident that the KF-based model achieves faster prediction speed due to its lower computational complexity and few parameters. Among the deep learning-based models, the direct multi-horizon approaches demonstrate substantial improvements in computational performance compared to the iterative multi-horizon approaches. In addition, it is worth noting that the model size of FlightBERT and FlightBERT++ is larger than comparison models. It can be

Style	Methods	Hor.	MAE ↓			MAPE (%) ↓			RMSE ↓			MDE ↓
			Lon	Lat	Alt	Lon	Lat	Alt	Lon	Lat	Alt	
Iterative	LSTM	1	0.0054	0.0055	1.71	0.0050	0.0202	0.29	0.0093	0.0124	6.04	0.91
		3	0.0065	0.0065	3.01	0.0061	0.0238	0.50	0.0110	0.0151	7.63	1.08
		9	0.0127	0.0121	10.33	0.0118	0.0443	1.48	0.0239	0.0270	18.48	2.05
		15	0.0214	0.0194	20.74	0.0199	0.0714	2.74	0.0436	0.0427	33.06	3.38
	Transformer	1	0.0038	0.0040	1.59	0.0036	0.0146	0.26	0.0071	0.0133	8.81	0.65
		3	0.0073	0.0073	3.24	0.0068	0.0268	0.54	0.0127	0.0172	9.12	1.21
		9	0.0173	0.0172	8.68	0.0161	0.0628	1.37	0.0305	0.0325	17.19	2.85
		15	0.0274	0.0270	13.87	0.0256	0.0986	2.08	0.0496	0.0501	26.25	4.50
	Kalman-Filter	1	0.0067	0.0032	1.50	0.0062	0.0120	0.28	0.0054	0.0171	8.50	0.82
		3	0.0112	0.0059	2.97	0.0104	0.0220	0.54	0.0861	0.0272	11.48	1.41
		9	0.0281	0.0168	8.73	0.0261	0.0624	25.83	0.1847	0.0619	25.83	3.69
		15	0.0494	0.0313	15.63	0.0459	0.1162	2.53	0.2896	0.1016	42.39	6.62
	FlightBERT	1	0.0024	0.0021	1.20	0.0023	0.0077	0.23	0.0387	0.0325	12.04	0.44
		3	0.0039	0.0036	2.19	0.0035	0.0133	0.41	0.0486	0.0506	13.65	0.71
		9	0.0091	0.0086	6.20	0.0085	0.0317	1.09	0.0608	0.0679	22.45	1.60
		15	0.0159	0.0148	10.80	0.0148	0.0549	1.84	0.0742	0.0794	32.76	2.71
Direct	LSTM+Attention	1	0.0064	0.0068	1.98	0.0059	0.0249	0.30	0.0109	0.0138	**5.96**	1.09
		3	0.0058	0.0060	2.75	0.0054	0.0221	0.43	0.0101	0.0144	**7.36**	0.98
		9	0.0082	0.0083	6.07	0.0076	0.0305	**0.94**	0.0178	**0.0221**	**14.03**	1.37
		15	0.0125	0.0122	9.05	0.0116	0.0450	1.38	0.0299	0.0327	**20.01**	2.04
	FlightBERT++	1	**0.0017**	**0.0017**	**1.15**	**0.0016**	**0.0066**	**0.20**	**0.0037**	**0.0115**	12.07	**0.31**
		3	**0.0031**	**0.0031**	**2.23**	**0.0029**	**0.0117**	**0.41**	**0.0067**	**0.0131**	12.46	**0.55**
		9	**0.0076**	**0.0074**	**5.30**	**0.0070**	**0.0277**	0.96	**0.0172**	0.0232	17.92	**1.29**
		15	**0.0124**	**0.0117**	**7.43**	**0.0109**	**0.0425**	**1.37**	**0.0265**	**0.0326**	22.89	**1.97**

Table 1: The experimental results of the proposed framework and baselines.

Methods	Parameters (M)	MTC (ms)
LSTM	0.55	48.96
Transformer	0.42	63.18
Kalman-Filter	–	0.64
FlightBERT	56.85	201.41
LSTM+Attention	0.90	14.43
FlightBERT++	44.13	6.81

Table 2: Comparison of the computational performance.

attributed that the BE representations extend the dimension of the inputs and enable us to dedicatedly design the sophisticated neural architecture to capture the flight transition patterns. Moreover, the proposed FlightBERT++ harvests the fastest computational speed among deep learning based models, even with a tens larger model size than other baselines. This significant improvement in computational efficiency is primarily attributed to the design of the HACG module, enabling the FlightBERT++ to perform multi-horizon prediction in a non-autoregressive manner.

Visualization and Qualitative Analysis

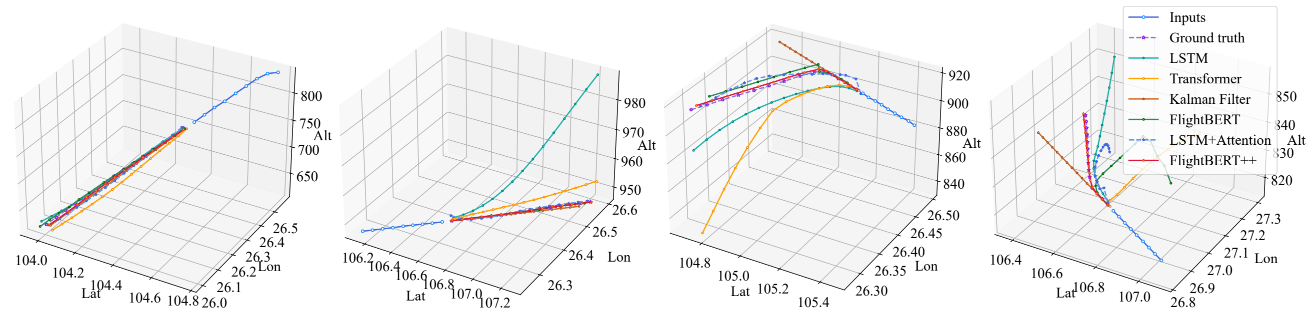
In this section, to better understand the learned flight transition patterns and qualitatively analyze the performance of different approaches, a total of 4 typical flight scenarios in the test set are selected to visualize the prediction results in Figure 3, including common flight scenarios (descending, en-route) and complex flight patterns (turn, climbing and turn right). Each sample is visualized with 9 observation trajectory points (inputs), ground truth, and 15 predicted trajectory points generated by different methods.

As shown in Figure 3, the FlightBERT++ achieves superior performance over comparison baselines for both common flight scenarios and complex flight patterns. It is observed that the proposed model also exhibits the ability to estimate the flight intents in future horizons (Figure 3c, 3d). This indicates that the FlightBERT++ not only captures the flight dynamics from the observations but also learns typical flight patterns, such as fixed waypoints of turns or descents, from a substantial amount of historical trajectories in training samples. Moreover, the FlightBERT++ shows lower error accumulation during the multi-horizon prediction process, indicating significant reliability over comparison approaches. Considering the enhancement of downstream tasks, such as conflict detection and airspace planning, FlightBERT++ can be regarded as a powerful tool to improve the overall efficiency and safety of ATC operations.

Insights

Ablation study In this section, to better understand the contributions of the designed components in the FlightBERT++ framework, an additional group C experiment is presented to conduct the ablation study. In experiment C1, a naive sum operator is applied to replace the ASA module in the trajectory encoder, while the differential prompt mechanism is removed from the DPD in experiment C2.

As can be seen in Table 3, all the designed network components make expected contributions to the FlightBERT++ framework. In experiment C1, the model suffers from more considerable error accumulation than the original FlightBERT++ with the increasing of prediction horizon. The weighted sum operation of the ASA module is believed to



(a) Descending (b) En-route (c) Turn (d) Climbing and turn right

Figure 3: Visualization of the FTP results in selective flight scenarios, in which the altitude is measured in 10 m.

Exp.	Models	Hor.	MAE ↓			MAPE (%) ↓			RMSE ↓			MDE ↓
			Lon	Lat	Alt	Lon	Lat	Alt	Lon	Lat	Alt	
C1	FlightBERT++ (without ASA module)	1	0.0018	0.0018	1.18	0.0016	0.0067	0.21	0.0039	0.0119	12.10	0.32
		3	0.0033	0.0032	2.28	0.0031	0.0119	0.42	0.0071	0.0138	12.56	0.57
		9	0.0080	0.0077	5.25	0.0074	0.0285	0.94	0.0180	0.0243	17.89	1.34
		15	0.0123	0.0118	7.16	0.0114	0.0439	1.31	0.0274	0.0339	22.38	2.05
C2	FlightBERT++ (without differential prompted mechanism)	1	0.0019	0.0018	1.27	0.0018	0.0071	0.23	0.0044	0.0116	12.18	0.34
		3	0.0035	0.0034	2.56	0.0033	0.0126	0.46	0.0084	0.0138	12.97	0.60
		9	0.0091	0.0086	6.71	0.0085	0.0318	1.15	0.0227	0.0267	20.56	1.51
		15	0.0153	0.0144	10.66	0.0143	0.0534	1.81	0.0386	0.0409	28.59	2.53

Table 3: The experimental results of the ablation study.

be more effective in capturing informative flight dynamics and temporal correlations from the observation trajectory sequence. Furthermore, in experiment C2, it can be found that the differential prompted mechanism is critical to the proposed FlightBERT++, especially in the long horizon prediction settings. It can be attributed that the differential prediction paradigm may lose some geographical and kinematical features of the flight trajectory, which makes it difficult to learn the differential transition patterns implicitly (without prompting) for the long-horizon prediction conditions. Therefore, the proposed differential prompted mechanism significantly contributes to the overall performance of the proposed FlightBERT++.

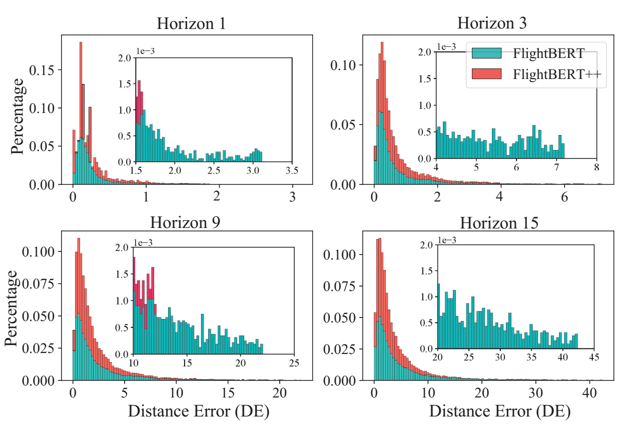


Figure 4: The histogram of the distance error.

Error analysis (FlightBERT++ v.s. FlightBERT) To investigate the outliers caused by high-bit errors of BE representation, the distribution of the distance errors between the FlightBERT and FlightBERT++ is visualized by a histogram in Figure 4. It is evident that the FlightBERT framework still outputs larger distance error values in a certain number of samples. In contrast, thanks to the design of the differential prediction paradigm, the predictions of the FlightBERT++ show a significant reduction in outliers, which further supports our motivation to overcome the high-bit errors.

Conclusion

In this paper, we present a novel FlightBERT++ framework to perform the multi-horizon FTP task in a non-autoregressive manner. The proposed framework not only inherits the superior representation ability of the BE in the FlightBERT framework but also develops an innovative multi-horizon FTP model. Benefiting from the proposed differential prediction paradigm, the FlightBERT++ can mitigate the high-bit errors of the BE representation and achieve a significant reduction in outliers than that of the FlightBERT. In addition, the proposed differential prompt mechanism is also confirmed to contribute to the performance improvements of the FTP task. In summary, FlightBERT++ achieves significant performance improvements and outperforms the comparison baselines across all evaluation metrics, especially in multi-horizon prediction scenarios.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under grants No. 62371323, U20A20161, U2333209, and 62001315.

References

Bahdanau, D.; Cho, K.; and Bengio, Y. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Benavides, J. V.; Kaneshige, J.; Sharma, S.; Panda, R.; and Steglinski, M. 2014. *Implementation of a Trajectory Prediction Function for Trajectory Based Operations*.

Chen, Z.; Guo, D.; and Lin, Y. 2020. A Deep Gaussian Process-Based Flight Trajectory Prediction Approach and Its Application on Conflict Detection. *Algorithms*, 13(11): 293.

FAA, E. 2010. Common trajectory prediction structure and terminology in support of SESAR & NextGen. Technical report, Technical Report. USA: FAA/EUROCONTROL AP 16, Brussels.

Guo, D.; Wu, E. Q.; Wu, Y.; Zhang, J.; Law, R.; and Lin, Y. 2023. FlightBERT: Binary Encoding Representation for Flight Trajectory Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 24(2): 1828–1842.

Kalman, R. E. 1960. A New Approach To Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82D: 35–45.

Li, Y.; Du, W.; Yang, P.; Wu, T.; Zhang, J.; Wu, D.; and Perc, M. 2019. A Satisficing Conflict Resolution Approach for Multiple UAVs. *IEEE Internet of Things Journal*, 6(2): 1866–1878.

Lin, Y.; wei Zhang, J.; and Liu, H. 2019. Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation. *Aerospace Science and Technology*, 93: 105113.

Liu, W.; and Hwang, I. 2011. Probabilistic Trajectory Prediction and Conflict Detection for Air Traffic Control. *Journal of Guidance, Control, and Dynamics*, 34(6): 1779–1789.

Ma, Y.; Du, W.; Chen, J.; Zhang, Y.; Lv, Y.; and Cao, X. 2023. A Spatiotemporal Neural Network Model for Estimated-Time-of-Arrival Prediction of Flights in a Terminal Maneuvering Area. *IEEE Intelligent Transportation Systems Magazine*, 15(1): 285–299.

Shi, Z.; Xu, M.; Pan, Q.; Yan, B.; and Zhang, H. 2018. LSTM-based Flight Trajectory Prediction. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, 1–8. IEEE.

Tang, X.; Zhou, L.; Shen, Z.; and Tang, M. 2015. *4D Trajectory Prediction of Aircraft Taxiing Based on Fitting Velocity Profile*, 1–12.

Thipphavong, D. P.; Schultz, C. A.; Lee, A. G.; and Chan, S. H. 2013. Adaptive Algorithm to Improve Trajectory Prediction Accuracy of Climbing Aircraft. *Journal of Guidance, Control, and Dynamics*, 36(1): 15–24.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 5998–6008.

Wang, Z.; Liang, M.; and Delahaye, D. 2017. Short-term 4D Trajectory Prediction Using Machine Learning Methods. In *SID 2017, 7th SESAR Innovation Days*. Belgrade, Serbia.

Wang, Z.; Liang, M.; and Delahaye, D. 2018. A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area. *Transportation Research Part C: Emerging Technologies*, 95: 280–294.

Yan, H.; Huang, G.; Wang, H.; and Shu, R. 2013. Application of Unscented Kalman Filter for Flying Target Tracking. In *2013 International Conference on Information Science and Cloud Computing*, 61–66.

Yan, Z.; Yang, H.; Guo, D.; and Lin, Y. 2023. Improving airport arrival flow prediction considering heterogeneous and dynamic network dependencies. *Information Fusion*, 100: 101924.

Zeng, W.; Quan, Z.; Zhao, Z.; Xie, C.; and Lu, X. 2020. A Deep Learning Approach for Aircraft Trajectory Prediction in Terminal Airspace. *IEEE Access*, 8: 151250–151266.