A Generalized Multivariate Flight Data Prediction Framework with Dynamic Feature Decomposition

Tao Hong, Qiaoli Zhou*, Zhiqiang Chang, Bin Guo
School of Computer Science, Shenyang Aerospace University
Shenyang, China
E-mail: hongtao1@stu.sau.edu.cn, zhouql@sau.edu.cn*, changzhiqiang@stu.sau.edu.cn, guobin1@stu.sau.edu.cn

***Abstract*—Accurate prediction of multivariate flight data is essential for aviation safety and operational optimization. However, flight sensor signals exhibit strong nonstationarity caused by environmental variations, abrupt maneuvers, and sensor noise, making it difficult for conventional models to capture both long-term trends and short-term fluctuations. To address this challenge, we propose a unified regression framework based on dynamic feature decomposition. The model adaptively separates multivariate time series into low-frequency trends and high-frequency fluctuations using a dual-branch architecture, where an auxiliary compensation branch enhances the representation of transient dynamics. A differencing-based preprocessing strategy improves sensitivity to rapid changes, while dynamic loss weighting balances learning across heterogeneous flight variables. In addition, a sigmoid-based normalization scheme stabilizes training under diverse feature distributions. Experiments on real-world flight datasets demonstrate consistent improvements or competitive performance against strong baselines across multiple flight-related variables, validating the robustness and generalization capability of the proposed method.**

***Index Terms*—Flight Data Prediction, Dynamic Feature Decomposition, Quick Access Recorder**

I. Introduction

Modern aviation operations generate massive amounts of flight data through onboard sensors and monitoring systems. These multivariate flight records contain rich information about aircraft performance, environmental conditions, and operational states. Accurate prediction and analysis of such data are crucial for enhancing operational safety, improving flight efficiency, and enabling proactive maintenance.

Predicting multivariate flight data poses several inherent challenges. First, the data are high-dimensional, heterogeneous, and multi-scale, comprising continuous sensor readings (e.g., speed, altitude) and discrete system states (e.g., flight phase, weight-on-wheels), often sampled at varying frequencies. Prior studies have shown that effectively modeling such data requires capturing both global trends and local temporal variations across multiple time scales [1], [2]. Li et al. [3] further characterized QAR (Quick Access Recorder) data as containing over 2000 variables with diverse temporal resolutions. Second, flight data are typically noisy and strongly nonstationary. Abrupt disturbances caused by flight maneuvers or changing environmental conditions further complicate stable pattern extraction and degrade model robustness [4], [5]. These challenges underscore the necessity of appropriate preprocessing strategies and dynamic temporal modeling.

With the growing demand for safe and efficient air traffic management, data-driven flight status monitoring and prediction have gained increasing importance. Recent studies demonstrate that accurate prediction plays a critical role in both safety assurance and operational optimization, such as fuel efficiency improvement and proactive anomaly detection [6], [7]. Moreover, advances in representation learning and Transformer-based architectures have enabled unified modeling frameworks that support multiple downstream tasks, including long-term trajectory forecasting, flight phase recognition, and safety event prediction [8], [9]. These developments collectively motivate the pursuit of reliable and generalized models for multivariate flight data prediction.

Existing methods often follow a single-task paradigm, designing separate models for specific objectives such as trajectory forecasting, flight phase recognition, or taxiing distance estimation. While effective for individual tasks, such task-specific designs limit the ability to learn shared representations across heterogeneous flight signals. Moreover, many approaches overlook the intrinsic frequency structure of flight data, where low-frequency trends and high-frequency disturbances coexist across variables with different temporal dynamics and scales. This motivates the need for a generalized modeling framework that can jointly handle diverse flight variables within a unified representation.

To overcome these challenges, we propose a unified multivariate regression framework based on a Transformer architecture enhanced with a dynamic feature decomposition mechanism. This mechanism decomposes flight signals into low-frequency trends and high-frequency components. The high-frequency components compensate for rapid fluctuations, enabling joint prediction of diverse flight variables with heterogeneous signal types and frequency characteristics. Our contributions are:

- **Unified Multivariate Regression Framework:** We design a Transformer-based framework that jointly regresses continuous and discrete flight variables within a single architecture.
- **Trend-Disturbance Dual-Branch Design:** A dual-branch structure explicitly models low-frequency trends and high-frequency fluctuations separately, facilitating effective decomposition of mixed-frequency signals.

*Corresponding author.

• **High-Frequency Compensation Module:** A dedicated module captures residual perturbations and abrupt signal variations, improving adaptability to nonstationary flight patterns.

• **Differential Modeling with Sigmoid Normalization:** Applying differencing emphasizes temporal dynamics, followed by Sigmoid normalization to stabilize and bound these representations, yielding a unified responsive feature space.

• **Dynamic Loss Reweighting for Multivariate Joint Optimization:** A two-level dynamic loss reweighting scheme balances heterogeneous flight variables in joint regression and jointly optimizes trend, disturbance, and reconstruction losses across model branches, improving training stability.

II. Related Work

A. Multivariate Flight Data Prediction

Deep learning models, including LSTM, GRU, and Transformers, have advanced multivariate flight data prediction by capturing complex temporal dependencies. LSTMs model sequences for landing and trajectory prediction [10], [11], while GRU and hybrid CNN-GRU architectures improve 4D trajectory and speed forecasts [12], [13]. Transformer-based models such as FlightPatchNet [1] and MDGNN [14] leverage multi-scale patches and dynamic graphs to learn spatiotemporal patterns effectively.

Despite these advances, many models still struggle to capture nonlinear interactions due to limited feature adaptability [15]. Issues such as disparate temporal resolutions and normalization-induced distortions remain insufficiently addressed [1]. These challenges motivate approaches that explicitly model multi-scale dynamics and complex dependencies. Our framework tackles this via dynamic feature decomposition for adaptive temporal modeling across diverse flight variables.

B. Feature Decomposition in Time Series

Feature decomposition is key to capturing multi-scale dynamics in nonstationary flight data. Approaches include ensemble empirical mode decomposition (EEMD) [16], multifractal analysis [17], and deep architectures that separate trend and seasonal components [18], [19]. These methods improve long-term forecasting and highlight the importance of multi-resolution feature extraction.

GAN-based self-attention [20] and multi-step LSTMs [10] have been applied to capture multi-scale QAR patterns and landing distances. SDTAN [15] demonstrates the effectiveness of multi-scale attention and graph-based modeling for flight safety prediction. Our method extends these ideas with a dual-branch architecture supervised by scale-specific losses, enabling end-to-end learning of temporal representations at multiple resolutions.

C. Scale Heterogeneity in Multivariate Flight Time Series

Flight data prediction is further complicated by scale heterogeneity across variables and phases. Variables differ in physical meaning, magnitude, noise, and temporal behavior. Conventional normalization methods (min–max, z-score) are static and may distort high-frequency patterns or bias learning [1], [21]. Liu et al. [22] showed that standard layer normalization can oversmooth temporal dynamics.

Recent deep learning models mitigate scale heterogeneity using attention mechanisms and scale-aware representations. Transformers capture data-driven feature interactions [21], while graph-based methods model inter-variable relations. Scaleformer [23] uses cross-scale normalization, and evolving multi-scale graph networks [24] adapt to changing dynamics. However, most approaches treat scale alignment as auxiliary, leaving room for end-to-end solutions integrated directly into model architectures.

III. Methodology

As illustrated in Fig 1, our framework employs a Transformer-based encoder-decoder architecture to predict future multivariate flight variables from historical sequences. We describe each component below.

A. Differencing-Based Feature Representation

Raw flight data often exhibit strong trends, nonstationarity, and significant inter-channel scale variation. Direct modeling may hinder training and generalization. To mitigate this, we design a differencing-based preprocessing pipeline:

First-order differencing enhances local dynamics and stationarity:

$$\Delta x_t = x_t - x_{t-1} \tag{1}$$

Scale-adaptive Sigmoid normalization constrains values to a smooth, bounded range:

$$x_{\text{norm}} = \sigma\left(-k \cdot \frac{x - \text{start} - n}{n} + \text{offset}\right) \tag{2}$$

where $\sigma(\cdot)$ is the Sigmoid function, $n = |\text{end} - \text{start}|$, and constants k and *offset* control the curve sharpness and horizontal shift. Here, start and end denote the lower and upper bounds of each variable, estimated from its empirical statistical range (e.g., minimum and maximum values in the training set).

Compared with min–max scaling, this scale-adaptive normalization better preserves local variations and gradient stability, particularly after differencing. By adjusting the effective saturation range according to each variable's statistical span, large-magnitude variations are kept within the high-sensitivity region of the Sigmoid mapping, improving robustness for nonstationary multivariate signals.

As illustrated in Fig. 2, the proposed normalization provides more stable and informative inputs under different data ranges.

B. Dual-Branch Dynamic Decomposition

To disentangle temporal signals, we design a dual-branch architecture:

• **Trend Modeling Branch:** This branch models smooth, low-frequency dynamics using the main Transformer encoder-decoder. To enhance temporal receptive fields, we incorporate a multi-scale convolutional embedding module that captures both local and mid-range patterns before input to the encoder.

• **Disturbance Modeling Branch:** To capture transient variations, this auxiliary branch operates on differenced signals

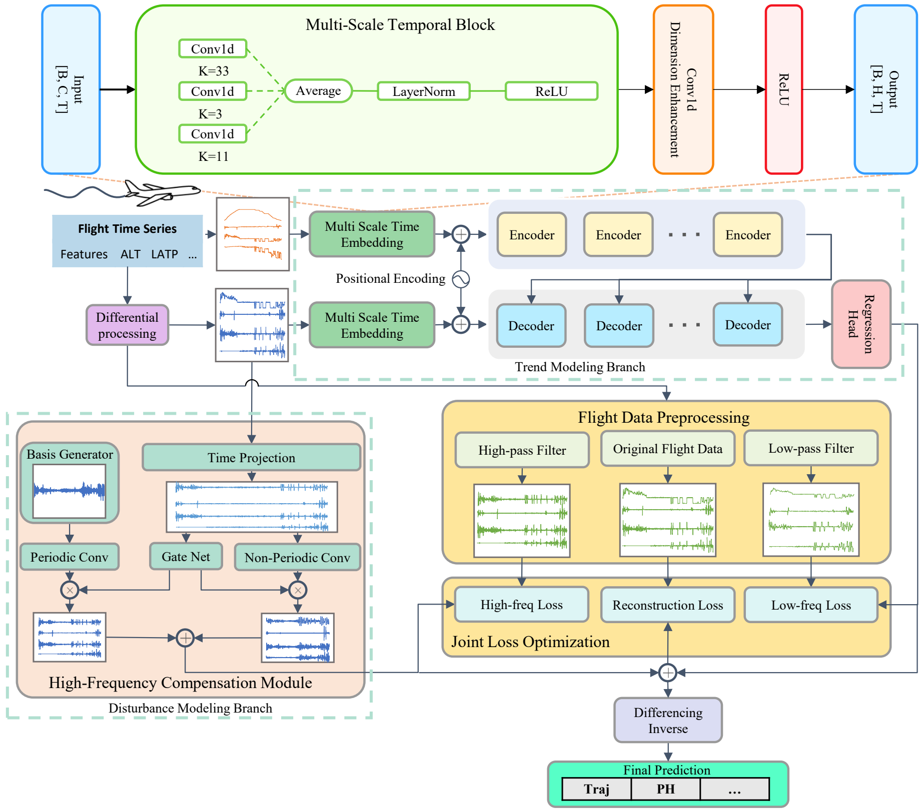

Fig. 1: Overview of the proposed multivariate flight data prediction framework. It uses a dual-branch architecture: the left-bottom branch captures high-frequency variations (with differencing), the right-top branch (multi-scale embedding + Transformer) models low-frequency trends. Branches are supervised independently (high/low-freq losses), fused via a compensation module, and constrained by joint reconstruction loss. This framework enables adaptive decomposition, transient refinement, and multi-variable (trajectory, phase) prediction.

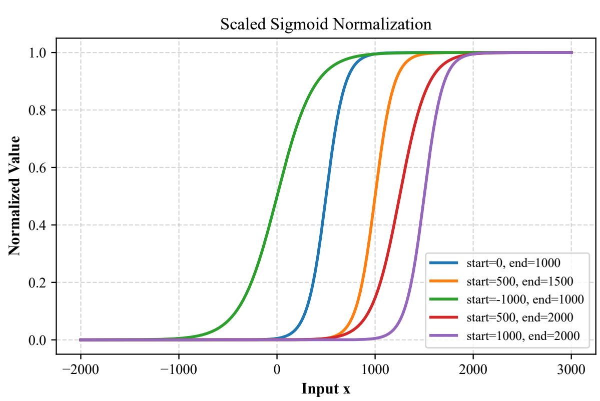

Fig. 2: Scale-adaptive Sigmoid normalization under different statistical ranges, where *start* and *end* denote the lower and upper bounds of the input variable.

and constitutes the *High-Frequency Compensation Module (HFCM)*. It comprises a dual-path convolutional structure: one path convolves a set of predefined basis functions that approximate common high-frequency patterns (e.g., sinusoidal, impulsive, and noisy components), and the other learning residual structures directly via deep convolutions on differenced inputs. A gating mechanism adaptively fuses the two, and a temporal projection layer aligns feature dimensions with the main output.

The outputs of both branches are fused via element-wise addition, enabling the model to simultaneously capture global trends and local fluctuations.

C. Joint Loss Optimization and Dynamic Weighting

The model jointly predicts trend-aware signals and high-frequency corrections. We define three MSE-based losses:

$$\mathcal{L}_{lf} = \mathrm{MSE}\big(\hat{y}_{lf}, y_{lf}\big), \tag{3}$$

$$\mathcal{L}_{hf} = \mathrm{MSE}\big(\hat{y}_{hf}, y_{hf}\big), \tag{4}$$

$$\mathcal{L}_{recon} = \mathrm{MSE}\big(\hat{y}_{lf} + \hat{y}_{hf}, y\big), \tag{5}$$

where $\hat{y}_{lf}$ and $\hat{y}_{hf}$ are the predicted trend and residual signals, and y_{lf}, y_{hf}, y are the corresponding ground truths, obtained via moving average decomposition:

$$\hat{y}_{lf} = \mathrm{MA}(y), \quad \hat{y}_{hf} = y - \hat{y}_{lf} \tag{6}$$

Algorithm 1 EMA-Guided Inverse-R^2 Weighting

Require: R^2, previous weights w, EMA factor α, target ratio $target_{\text{ratio}}$, bounds k_{bounds}, tolerance x_{tol}

if first iteration **then**
  $w \leftarrow \frac{1}{N}$ (uniform)
end if
Clip R^2 to $[-1, 1]$
if $|\min(R^2) - \max(R^2)| < 10^{-5}$ **then**
  return $0, w$
end if
Define $w(k) \leftarrow \text{Normalize}\big(\max(1 + k(1 - R^2), 0.1)\big)$
Define objective: $f(k) \leftarrow \frac{\max(w(k))}{\min(w(k))} - target_{\text{ratio}}$
Find k^* via bisection on $f(k)$ over k_{bounds} with tol x_{tol}
$w_{\text{new}} \leftarrow w(k^*)$
$w \leftarrow \alpha \cdot w_{\text{new}} + (1 - \alpha) \cdot w$
return k^*, w

Table I: Selected Variables for Multivariate Flight Modeling

Category	Selected Variables
Binary and Event Indicators	MSQT_1, MSQT_2, WOW
Navigation and Positioning	PH, LATP, LONP, ALT, TRK
Attitude and Motion	PTCH, ROLL, TH, GS, LONG, LATG, VRTG, IVV, TAS, ALTR, DA, RALT
Flight Control Surfaces	FLAP, AIL_1–2, ELEV_1–2, RUDD, CCPC, CWPC
Powerplant and Engine	PLA_1–4, FF_1–4, EGT_1–4, OIP_1–4, N1_1–4
Environmental Sensing	WS, WD
Flight Dynamics and Systems	FPAC, BLAC, BAL_1–2, GLS, LOC

The overall objective is a weighted combination:

$$\mathcal{L} = w_c \left(\lambda_1 \mathcal{L}_{lf} + \lambda_2 \mathcal{L}_{hf} + \lambda_3 \mathcal{L}_{recon} \right) \tag{7}$$

where $\lambda_1, \lambda_2, \lambda_3$ are scalar weights and w_c is a dynamic channel-wise weight vector.

To promote balanced learning across variables with diverse scales and dynamics, we introduce a validation-driven feedback mechanism (Algorithm 1). It adaptively adjusts w_c based on the inverse coefficient of determination (R^2) on the validation set, smoothed via exponential moving average (EMA). This ensures under-optimized variables receive higher attention during training. The weighting strategy is applied only during training and incurs negligible computational overhead.

IV. Experiments

A. Dataset

We conduct experiments on the **DASHlink Sample Flight Data**, a public dataset released by NASA [25]. It contains over 180,000 flight data files, recording multivariate time series from commercial flights and ground operations. Signals include navigation states, flight dynamics, control inputs, engine variables, and environmental conditions, with original sampling rates ranging from 0.25 Hz to 16 Hz.

For evaluation, 4,463 flight records were selected. Data segments corresponding to PH values 3–7 were extracted, covering major phases such as taxiing, takeoff, cruise, and landing. All signals were resampled to 1 Hz to unify temporal resolution. From the original 186 channels, 57 representative variables were chosen, spanning navigation, attitude and motion, control surfaces, powerplant, environmental sensing, and flight dynamics (Table I). Variables labeled "1–n" indicate multiple instances of the same type (e.g., PLA_1-4).

Geodetic positions (LATP, LONP) were converted to relative distances (LATP_DIST, LONP_DIST), and Euler angles were transformed into quaternions (QW, QX, QY, QZ). The dataset was split at the flight level into training, validation, and test sets to prevent leakage. Within each subset, a sliding-window strategy was applied to generate observation–prediction pairs.

B. Experiment Settings

All models are trained using Mean Squared Error (MSE) loss on an NVIDIA RTX 4090 GPU with 24 GB memory. The input sequence length is $L = 200$, with a prediction horizon of $H = 50$. Training is conducted for a total of 20 epochs with a learning rate of 2×10^{-4} and a batch size of 128, and early stopping is applied to prevent overfitting during training.

Model architectures are configured to ensure comparable capacity while accommodating design differences. The hidden dimension is set to $d = 512$ for most models, whereas *iTransformer* adopts $d = 1024$ to support its channel-wise attention. Encoder-decoder models use 3 encoder layers and 6 decoder layers, while decoder-only models have 6 layers. All Transformer-based models employ 8 attention heads and a dropout rate of 0.1.

Table II: Performance comparison on multivariate flight data prediction measured by RMSE.

Variable (Unit)	iTransformer	Fredformer	TDformer	xPatch	TimeMixer	Ours
LATP_DIST (m)	798.569	**287.060**	581.535	506.003	394.883	322.096
LONP_DIST (m)	804.873	**293.688**	629.209	472.476	394.932	350.024
ALT (ft)	278.349	118.119	182.221	146.581	105.351	**98.662**
TRK (°)	47.327	27.037	37.298	36.156	31.237	**27.441**
QW (-)	0.128	0.045	0.062	0.075	0.052	**0.043**
QX (-)	0.052	0.028	0.036	0.030	0.032	**0.024**
QY (-)	0.050	0.029	0.039	0.032	0.033	**0.023**
QZ (-)	0.238	0.157	0.205	0.204	0.187	**0.147**
DA (°)	3.014	0.968	1.185	1.424	0.999	**0.893**
GS (knots)	15.779	6.825	9.142	10.054	6.272	**6.098**
LONG (g)	0.120	0.047	0.066	0.063	0.058	**0.037**
VRTG (g)	0.506	0.116	0.408	0.232	0.167	**0.101**
LATG (g)	0.083	0.036	0.062	0.053	0.047	**0.026**
IVV (ft/min)	608.570	384.632	434.337	422.686	345.401	**313.641**
TAS (knots)	20.657	8.422	10.718	12.379	7.260	**7.682**
ALTR (ft/min)	697.602	390.886	433.557	429.563	341.951	**312.398**
RALT (ft)	244.535	116.604	165.036	139.565	109.980	**88.613**
FLAP (count)	338.150	174.830	225.616	306.759	174.735	**149.759**
AIL_1–2 (°)	3.145	1.392	2.384	2.707	1.628	**1.354**
ELEV_1–2 (°)	4.395	1.557	2.523	2.179	1.866	**1.489**
RUDD (°)	7.160	**4.573**	6.038	5.187	5.399	5.138
CCPC (count)	359.685	**109.190**	163.276	175.856	122.390	121.260
CWPC (count)	201.803	85.889	147.508	105.412	90.250	**80.606**
PLA_1–4 (°)	6.649	4.569	5.210	5.017	**4.317**	4.668
FF_1–4 (lb/h)	199.011	141.461	158.541	153.538	130.637	**130.398**
EGT_1–4 (°C)	27.837	17.578	21.328	19.784	16.953	**15.769**
OIP_1–4 (psi)	9.176	**4.258**	5.762	5.089	4.502	5.153
N1_1–4 (%RPM)	7.495	5.317	5.901	5.838	4.912	**4.666**
BAL_1–2 (ft)	279.209	120.475	184.235	151.076	108.576	**100.638**
WS (knots)	6.741	3.838	4.450	6.774	3.264	**3.563**
WD (°)	46.275	28.368	38.335	34.740	30.717	**26.946**
BLAC (g)	0.034	0.022	0.024	0.023	**0.019**	**0.019**
FPAC (g)	0.033	0.023	0.023	0.026	0.018	**0.017**
GLS (DDM)	0.109	**0.050**	0.067	0.063	0.060	0.053
LOC (DDM)	0.115	**0.068**	0.095	0.088	0.081	0.070

Discrete flight attributes (e.g., flight phase indicators and binary events) are incorporated within a unified regression

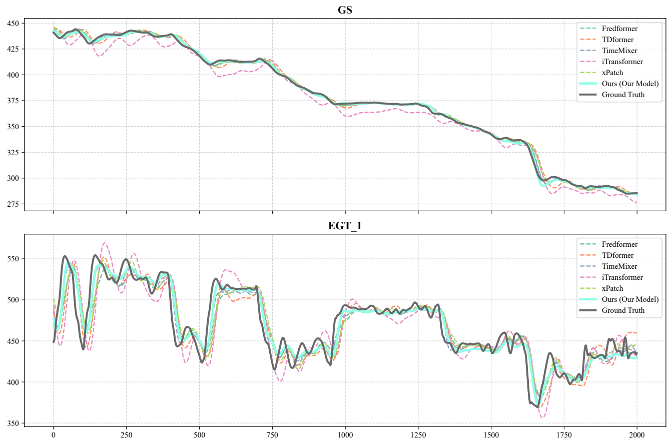

Fig. 3: Model prediction performance on GS and EGT_1 variables.

Table III: Comparison of Model Performance on Discrete Flight Attributes

Variable	iTransformer		Fredformer		TDformer		xPatch		TimeMixer		Ours	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
PH	0.464	0.538	0.963	0.961	0.935	0.931	0.943	0.940	**0.971**	**0.970**	**0.971**	0.969
MSQT_1	0.976	0.718	0.990	0.868	0.988	0.835	0.988	0.843	0.993	0.903	**0.994**	**0.920**
MSQT_2	0.976	0.716	0.990	0.869	0.988	0.834	0.988	0.843	0.993	0.903	**0.994**	**0.921**
WOW	0.882	0.937	0.996	0.998	0.993	0.997	0.993	0.996	0.998	**0.999**	**0.999**	**0.999**

framework. During evaluation, predictions are thresholded to compute classification metrics, allowing joint learning without separate task-specific heads or losses.

C. Overall Performance on Multivariate Prediction

We evaluate the proposed model from two complementary perspectives: (1) continuous flight variable forecasting and (2) discrete and event-related attribute prediction. Comparisons are made against representative state-of-the-art baselines for multivariate time-series modeling, including iTransformer [22], Fredformer [18], TDformer [26], xPatch [19], and TimeMixer [27]. These baselines span diverse paradigms such as feature-wise attention, frequency-domain enhancement, temporal-difference modeling, dual-stream decomposition, and trend-seasonality mixing.

All models are trained and evaluated under the same experimental settings, including sequence length, prediction horizon, learning rate, batch size, and early stopping, to ensure a fair comparison. Minor structural adjustments are applied to maintain comparable model capacity across different architectures.

Table II reports RMSE results over a broad set of multivariate flight variables. The proposed model achieves the lowest RMSE on most attitude, motion, and engine-related variables, while remaining competitive on navigation-related ones. For altitude and vertical-motion indicators (ALT, RALT, IVV), it consistently outperforms strong baselines, yielding relative RMSE reductions of approximately 6–20%. Clear improvements are also observed on control and engine variables such as FLAP and EGT_1-4, with reductions of about 7–15%. For FF_1-4, performance is comparable to the best baseline, with only marginal differences. On navigation-related variables (LATP_DIST, LONP_DIST), although not achieving the lowest error, the proposed method consistently outperforms several competitive baselines, indicating robust spatial modeling capability. For low-amplitude or noise-sensitive variables (LATG, GLS), performance remains comparable to the strongest baseline, demonstrating stable and non-degrading behavior.

Fig. 3 illustrates representative predictions for GS, a trend-dominant variable, and EGT_1, a fluctuation-dominant variable. The proposed model closely matches the ground truth, accurately capturing both long-term trends and short-term variations. For GS, most baselines gradually deviate during plateau and terminal stages, while for EGT_1, several methods overly smooth high-frequency fluctuations or exhibit temporal misalignment. In contrast, the proposed model preserves sharp variations without introducing phase shifts, demonstrating robustness to heterogeneous temporal dynamics.

Table III summarizes classification results on discrete flight attributes. The proposed model delivers strong and consistent performance across all variables and achieves the best overall results on most tasks. It yields accuracy improvements of about 0.1% on MSQT_1 and MSQT_2, together with F1-score gains of approximately 1.7% and 1.8%. On the WOW event, it achieves a modest accuracy advantage and attains the top F1-score jointly with the strongest baseline. For flight phase recognition (PH), the proposed approach reaches the highest accuracy while maintaining a comparable F1-score. Overall, these results demonstrate stable and reliable advantages in discrete flight attribute prediction, particularly for mode-switching and event-related variables.

D. Ablation Study

Fig. 4 presents a removal-based ablation study, where key components are progressively disabled to evaluate their contributions. Prediction errors are shown on a logarithmic scale to emphasize large performance variations. Removing dynamic loss weighting (R2) causes only a marginal degradation, suggesting its auxiliary role in optimization. In contrast, eliminating the decomposition module (R3) leads to a noticeable performance drop, highlighting its importance for separating trend and residual dynamics. Further disabling Sigmoid normalization (R4) results in a substantial error increase, indicating its critical role in amplitude stabilization. Removing first-order differencing (R5) causes a severe performance collapse, confirming that handling nonstationarity is

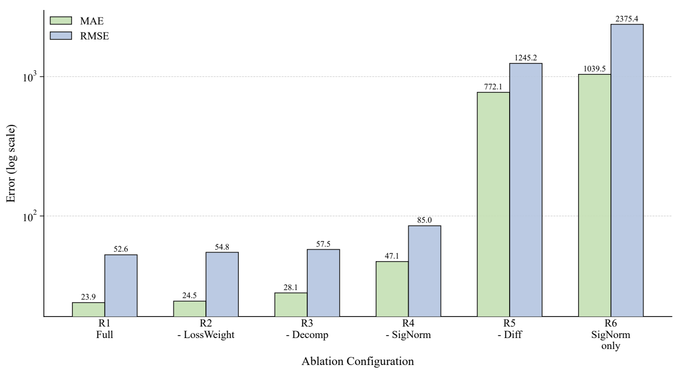

Fig. 4: **Removal-based ablation study.** Error degradation under progressive removal of model components (logarithmic scale).

essential. The "SigNorm only" configuration without differencing (R6) yields the worst performance, revealing a strong interdependence between temporal stabilization and amplitude normalization.

V. Conclusion

We presented a unified regression framework for multivariate flight data prediction, addressing nonstationarity and scale heterogeneity. The model uses dynamic feature decomposition to separate low-frequency trends and high-frequency fluctuations via a dual-branch architecture. Complementary strategies—including differencing-based preprocessing, scale-adaptive Sigmoid normalization, and dynamic loss reweighting—stabilize learning across heterogeneous continuous and discrete variables. Experiments on real-world flight datasets show consistent improvements for both continuous and discrete variables, capturing mixed-frequency dynamics and complex inter-variable dependencies. These results confirm the framework's robustness and generalization potential for reliable operational monitoring. Future work will focus on adaptation to rare or imbalanced flight states and optimization for computational efficiency to enable real-time deployment.

References

[1] L. Wu, X. Wang, R. Chu, G. Liu, Y. Chen, J. Zhang, and L. Wang, "FlightPatchNet: Multi-Scale Patch Network with Differential Coding for Flight Trajectory Prediction," May 2024.

[2] Y. Fan, Y. Tan, L. Wu, H. Ye, and Z. Lyu, "Global and Local Inter-attribute Relationships-Based Graph Convolutional Network for Flight Trajectory Prediction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 60, no. 3, pp. 2642–2657, Jun. 2024.

[3] X. Li, J. Shang, L. Zheng, Q. Wang, D. Liu, X. Liu, F. Li, W. Cao, and H. Sun, "IMTCN: An Interpretable Flight Safety Analysis and Prediction Model Based on Multi-Scale Temporal Convolutional Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 1, pp. 289–302, Jan. 2024.

[4] J. Wang, Y. Song, and T. He, "A novel adaptive monitoring framework for detecting the abnormal states of aero-engines with maneuvering flight data," *Reliability Engineering & System Safety*, vol. 258, p. 110910, Jun. 2025.

[5] J. Zhong, H. Zhang, and Q. Miao, "Flight anomaly detection and localization based on flight data fusion and random channel masking," *Applied Soft Computing*, vol. 174, p. 113023, Apr. 2025.

[6] Z. Wang, "A reliable predict-then-optimize approach for minimizing aircraft fuel consumption," 2025.

[7] B. Ayhan, A. Imanian, E. P. Vargo, M. Ravi, H. Tang, and J. P. Ice, "Forecasting-Based Anomaly Detection with Temporal Fusion Transformers Using Flight Operational Quality Assurance Data," in *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference (DASC)*. San Diego, CA, USA: IEEE, Sep. 2024, pp. 1–9.

[8] S. Liu, W. Li, D. Yao, and J. Bi, "Effective and Efficient Representation Learning for Flight Trajectories," in *AAAI*, 2025.

[9] S. Luo, M. Zhao, Z. Zhao, L. Li, S. Zhang, and X. Zhang, "FT-TF: A 4D Long-Term Flight Trajectory Prediction Method Based on Transformer," in *2023 42nd Chinese Control Conference (CCC)*. Tianjin, China: IEEE, Jul. 2023, pp. 4616–4621.

[10] Y. Kong and S. Mahadevan, "Aircraft Landing Distance Prediction: A Multistep Long Short-Term Memory Approach," *Journal of Aerospace Information Systems*, vol. 19, no. 5, pp. 344–354, May 2022.

[11] Z. Shi, M. Xu, Q. Pan, B. Yan, and H. Zhang, "Lstm-based flight trajectory prediction," in *2018 International joint conference on neural networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[12] Z. Kang, J. Shang, Y. Feng, L. Zheng, Q. Wang, H. Sun, B. Qiang, and Z. Liu, "A deep sequence-to-sequence method for accurate long landing prediction based on flight data," *IET Intelligent Transport Systems*, vol. 15, no. 8, pp. 1028–1042, 2021.

[13] L. Ma and S. Tian, "A hybrid cnn-lstm model for aircraft 4d trajectory prediction," *IEEE access*, vol. 8, pp. 134 668–134 680, 2020.

[14] L. Song, X. Li, H. Liu, L. Wu, H. Sun, L. Zheng, and J. Shang, "MDGNN: Multiple Flight Safety Incidents Prediction Model Based on Dynamic Graph Neural Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 4, pp. 5598–5612, Apr. 2025.

[15] H. Chen, J. Shang, L. Zheng, X. Li, X. Liu, H. Sun, X. Zhao, and L. Yu, "SDTAN: Scalable Deep Time-Aware Attention Network for Interpretable Hard Landing Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 9, pp. 10 211–10 223, Sep. 2023.

[16] G. Lin, "Multidimensional KNN algorithm based on EEMD and complexity measures in financial time series forecasting," *Expert Systems With Applications*, 2021.

[17] Y. Shen, "Based on hypernetworks and multifractals: Deep distribution feature fusion for multidimensional nonstationary time series prediction," 2024.

[18] X. Piao, Z. Chen, T. Murayama, Y. Matsubara, and Y. Sakurai, "Fredformer: Frequency Debiased Transformer for Time Series Forecasting," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Aug. 2024, pp. 2400–2410.

[19] A. Stitsyuk and J. Choi, "xpatch: Dual-stream time series forecasting with exponential seasonal-trend decomposition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, pp. 20 601–20 609.

[20] J. Zhao, C. Rong, X. Dang, and H. Sun, "QAR Data Imputation Using Generative Adversarial Network with Self-Attention Mechanism," *Big Data Mining and Analytics*, vol. 7, no. 1, pp. 12–28, Mar. 2024.

[21] D. Guo, E. Q. Wu, Y. Wu, J. Zhang, R. Law, and Y. Lin, "Flightbert: Binary encoding representation for flight trajectory prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 2, pp. 1828–1842, 2022.

[22] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "iTransformer: Inverted Transformers Are Effective for Time Series Forecasting," Mar. 2024.

[23] A. Shabani, A. Abdi, L. Meng, and T. Sylvain, "Scaleformer: Iterative Multi-scale Refining Transformers for Time Series Forecasting," Feb. 2023.

[24] J. Ye, "Learning the Evolutionary and Multi-scale Graph Structure for Multivariate Time Series Forecasting," 2022.

[25] NASA, "Discovery in aeronautics systems health (dashlink): Sample flight data," https://c3.nasa.gov/dashlink/projects/85/, 2012, accessed: 2025-01-10.

[26] Z. Zhu, Y. Yu, Z. Zheng, Z. Yu, and Y. Jin, "TDFormer: A Top-Down Attention-Controlled Spiking Transformer," May 2025.

[27] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, "TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting," May 2024.