

# DECOMPOSED ATTENTION FREDFORMER: LARGE TIME-SERIES PREDICTION MODEL FOR SATELLITE ORBIT PREDICTION

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Accurate satellite orbit prediction is critical for collision avoidance and sustainable space operations. However, conventional methods are constrained by coarse update intervals, orbit discontinuities, and other factors. Additionally, building separate prediction models for each satellite is computationally expensive, making large-scale accurate forecasting increasingly impractical. To address the aforementioned challenges, we propose Decomposed Attention FredFormer (DAF), a large time-series prediction model that uses efficient Real Fast Fourier Transform (RFFT)/Inverse RFFT in favor of positional embeddings. Our DAF also integrates Tensorized Multi-Head Attention based on Tensor Train Decomposition for parameter-efficient compression and improved performance. We pre-trained on a large-scale Starlink dataset and evaluated zero-shot performance on seven cross-domain satellite orbit datasets and three real-world datasets. DAF achieves up to 34.85% reduction in mean squared error and 16.01% reduction in mean absolute error over the second-best model, using only 0.05% of its parameters and maintaining inference time as fast as the conventional neural network baselines. These results demonstrate that DAF enables zero-shot, high-precision orbit prediction not only for Starlink satellites, but also for other satellites. The code is available here: <https://anonymous.4open.science/r/DAF-0D75>

## 1 INTRODUCTION

The number of operational spacecraft in Earth orbit has increased dramatically over the last decade, now exceeding 11,700 active payloads by March 2025 (France-Presse, 2025). The dominant growth of large constellations primarily drives the increase in the number of satellites. For example, as of August 1, 2025, the Starlink constellation consists of 8,074 active satellites, with the goal of expanding the network into a ‘megaconstellation’ comprising a total of 42,000 satellites (Pultarova, 2025). The increasing number of satellites imposes a significant mass burden on Earth’s orbital environment, leading to the generation of numerous debris fragments caused by collisions, explosions, and degradation in the harsh conditions of space (Boley & Byers, 2021). The Kessler Syndrome (Kessler & Cour-Palais, 1978) describes a scenario in which the growth of space debris not only poses a serious threat to satellite operations, but also increases the likelihood of cascading collisions among space objects. Over time, this results in an exponential increase in the amount of debris. Such space debris poses critical dangers to satellites and the International Space Station (ISS), threatening the long-term sustainability of space activities. A notable example occurred in February 2009, when the Russian satellite Cosmos 2251 collided with the U.S. Iridium 33 satellite, producing a large amount of debris (Iannotta, 2009). As a result, the ISS had to perform a total of five collision avoidance maneuvers over two years, up to early April 2011 (NASA Orbital Debris Program Office, 2011).

To preserve the sustainability of the space environment, many studies have focused on accurately predicting satellite orbits to prevent such collisions (Choumos et al., 2024; Caldas & Soares, 2024b; Reiland et al., 2021; Grau et al., 2025). Current operational orbit prediction typically relies on the propagation of Two-Line Element (TLE) sets using the analytical model Simplified General Perturbation-4 (SGP4) (Hoots & Roehrich, 1980). Although the SGP4 offers computational efficiency, its kilometer-level prediction error can accumulate over long-term predictions, resulting in differences of hundreds of kilometers from the actual orbit (Vallado et al., 2006). Recent machine

learning (ML)-based approaches for satellite orbit prediction have been proposed to address the accumulated prediction errors of the SGP4 model over long-term prediction. While such methods have demonstrated improved performance over using SGP4 alone, their evaluations were limited to a small number and a narrow variety of satellites. As a result, these approaches exhibit limited generalizability and constrained applicability in real-world operational scenarios (Pihlajasalo et al., 2018; San-Juan et al., 2018; Curzi et al., 2022; Jadala et al., 2022). Furthermore, with the rapid increase in the number of satellites, developing prediction models for each satellite becomes prohibitively time-consuming and computationally expensive. Therefore, large time-series models that are trained on massive datasets, can offer the advantage of zero-shot prediction across various domains without the need for individual model training (Das et al., 2024; Jin et al., 2024). However, large models typically require an increase in inference time. This characteristic makes them unsuitable for satellite orbit prediction, where real-time inference and limited hardware resources are critical constraints.

Therefore, there is a growing need for large models to provide improved zero-shot prediction performance for various satellites, while improving computational efficiency and conserving hardware resources. In this work, we aim to design the first large time-series model for satellite orbit prediction. In particular, we propose Decomposed Attention FredFormer (DAF) model, which is pre-trained on a large-scale Starlink dataset and achieves high prediction performance with a small number of parameters, while yielding fast inference speed. The DAF model leverages FredFormer, which has shown a high prediction performance on time-series datasets. Additionally, we utilize Real Fast Fourier Transform (RFFT) layer, while removing the patching and layer normalization. Furthermore, we develop Tensor Train Decomposition (TTD) to the attention weight matrix of FredFormer, reducing model parameters while simultaneously improving performance. Our DAF model demonstrates its effectiveness compared to existing large time-series models (e.g., TimesFM, Time-MoE) as well as general time-series models (e.g., GRU, LSTM). The main contributions of our work are summarized as follows:

- In this work, we propose the novel Decomposed Attention FredFormer (DAF), which is capable of training on large-scale satellite data and achieving high-performance zero-shot evaluation on various satellites. To the best of our knowledge, this research is the first large time-series model proposed in the field of satellite orbit prediction.
- The DAF model enhances prediction performance, while reducing model parameters by replacing FFT to RFFT, adopting positional embedding instead of patching, and applying TTD to Multi-Head Attention.
- We evaluate the generalizability and deployment potential of the DAF model through zero-shot evaluation on seven constellation datasets and three real-world satellite datasets. In addition, we achieve parameter reduction through TTD and demonstrate its effectiveness by comparing inference time with existing large models and conventional neural networks.

## 2 DATASET PROCESSING PIPELINE

In Figure 1 (a), we first illustrate the data collection process used in our study and explain how Orekit (maisonobe et al., 2024) method was applied for interpolation. Next, in Section 2.1, we provide a detailed description of the data collection process used in our study. Finally, in Section 2.2, we describe the interpolation methods and the corresponding validation procedures.

### 2.1 DATASET DESCRIPTION

Table 1 provides an overview of the nine satellite orbit datasets used in our experiments. We collected all TLE datasets (except for KOMPSAT) from Space-Track via its public API (Space-Track, 2025). The real-world datasets (KOMPSAT-3, 3A, and 5) were collected from analytical computation of orbital elements Jeong et al. (2024). Additionally, all datasets span from Communication and Observation to Multipurpose missions. This diversity in temporal resolution, dataset scale, and mission profile enables a comprehensive evaluation of our prediction models across a wide range of environments.

Satellites used in our research complete an orbit in approximately 90-120 minutes. However, TLEs are typically collected by humans only every eight hours, creating significant temporal gaps in the

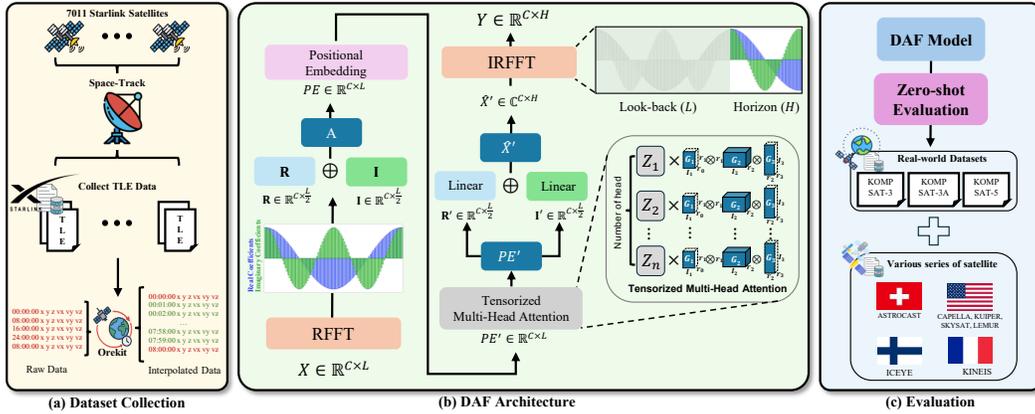


Figure 1: Overall framework of our proposed approach. (a) For data preprocessing and interpolation, we collected TLE data using the Space-Track’s public API and interpolated them using the Orekit method. (b) Input sequences are transformed to the frequency domain via RFFT and concatenated with positional embeddings. The transformed data pass through TMHA, undergo linear projection and IRFFT, and are finally used to predict the next sequence. (c) We evaluate zero-shot performance on seven cross-domain satellite orbit datasets and three real-world datasets.

Table 1: Summary of satellite orbit datasets, including name, mission type, number of rows and satellites, observation period, and average/minimum/maximum revisit frequency.

Dataset	Mission Type	Rows	Satellites	Period	Avg. Freq.	Min. Freq.	Max. Freq.
Starlink	Communication	13,234,406	7,011	19.11.14–25.02.25 (1,930 d)	11 h 21 m	0.864 ms	786 h 36 m
ASTROCAST	Communication	5,864	16	25.02.26–25.07.15 (140 d)	11 h 34 m	2.592 ms	85 h 34 m
CAPELLA	Observation	1,562	6	25.02.26–25.07.15 (140 d)	12 h 20 m	12.096 ms	78 h 50 m
ICEYE	Observation	18,431	42	25.02.26–25.07.15 (140 d)	8 h 47 m	6.048 ms	83 h 10 m
KINEIS	Communication	12,854	25	25.02.26–25.07.15 (140 d)	7 h 46 m	0.864 ms	91 h 14 m
KUIPER	Communication	6,865	55	25.02.26–25.07.15 (140 d)	9 h 26 m	82.080 ms	63 h 34 m
LEMUR	Observation	16,922	63	25.02.26–25.07.15 (140 d)	12 h 52 m	2.592 ms	114 h 11 m
SKYSAT	Observation	4,852	15	25.02.26–25.07.15 (140 d)	12 h 23 m	1 h 32 m	109 h 43 m
KOMPSAT	Multipurpose	7,892,646	3	16.01.01–21.01.01 (1,828 d)	1 m	1 m	1 m
Total		21,194,402	7,236	16.01.01-25.07.15 (3,484 d)		-	

raw data. Such gaps make it impossible to reflect perturbations such as atmospheric drag and solar radiation pressure, rendering the data unsuitable for training continuous prediction models. To address this challenge, we interpolate between gaps using the propagation method, filling in missing positions and velocities to produce a physically consistent orbit.

## 2.2 DATASET PREPROCESSING AND INTERPOLATION METHOD SELECTION

We model satellite orbits as time-series by integrating ground-station observations and TLE data. To determine the most accurate interpolation method among various satellite orbit interpolation techniques, we conducted a validation procedure. First, we collected ground observation data for 7,216 active Starlink satellites as of April 14, 2025, at one-minute resolution, in a modified ephemeris format from the International Telecommunication Center (ITC). After excluding 337 satellites launched after the Starlink period, which were used in our study, we obtained a final dataset of 6,879 satellites for which both ephemeris and TLE data matched.

Using the TLE datasets corresponding to these 6,879 satellites, we generated interpolated epochs and state vectors  $(x, y, z, vx, vy, vz)$  at one-minute intervals using the SGP4 (Hoots & Roehrich, 1980), Skyfield (Rhodes, 2019), and Orekit (maisonobe et al., 2024) methods. The ephemeris dataset served as ground truth (Khairallah & Kassas, 2021) and was used to evaluate interpolation error by comparing interpolated vectors to the ground truth using MSE, MAE, mean difference, and standard deviation. Before conducting the error analysis, we validated the statistical significance between the datasets using the Friedman test (Friedman, 1937) and Nemenyi post-hoc test (Nemenyi, 1963).

We used the Friedman test to examine whether there were statistically significant differences across the three datasets generated by the SGP4, Skyfield, and Orekit methods. When the Friedman test indicated significance, the Nemenyi post-hoc test was applied to perform pairwise comparisons and identify which interpolation pairs showed significant differences. As shown in Table 2, the Friedman

Table 2: Statistical analysis results validating the significance of interpolation performance across SGP4, Skyfield, and Orekit. The table reports results for four metrics across the  $x$ ,  $y$ , and  $z$  features, including Friedman test statistics and  $p$ -values, Nemenyi post-hoc  $p$ -values for pairwise comparisons, and the corresponding mean values for each method. The Nemenyi test was only applied to features with statistically significant Friedman test results. The best result for each metric is highlighted in **bold**.

Metric	Feat.	Friedman test		Nemenyi test ( $p$ -val.)			Mean Values		
		stat.	$p$ -val.	SGP4&Skyfield	SGP4&Orekit	Orekit&Skyfield	SGP4	Skyfield	Orekit
MSE	$x$	8448.9772	< 0.05	< 0.05	< 0.05	< 0.05	10645.732	9816.147	<b>9816.146</b>
	$y$	9174.1494	< 0.05	< 0.05	< 0.05	< 0.05	8133.012	7369.783	<b>7369.780</b>
	$z$	3947.891	< 0.05	< 0.05	< 0.05	0.9	10259.165	<b>10174.8805</b>	10174.8813
MAE	$x$	8497.8197	< 0.05	< 0.05	< 0.05	< 0.05	31.88368	13.587937	<b>13.587854</b>
	$y$	9230.8524	< 0.05	< 0.05	< 0.05	< 0.05	30.527394	13.181023	<b>13.180938</b>
	$z$	4552.0648	< 0.05	< 0.05	< 0.05	0.9	18.179287	13.355324	<b>13.355318</b>
Mean Diff.	$x$	396.8254	< 0.05	0.9	< 0.05	< 0.05	-0.038654	-0.022224	<b>-0.022223</b>
	$y$	30.4268	< 0.05	0.9	< 0.05	< 0.05	-0.002987	<b>-0.000447</b>	<b>-0.000447</b>
	$z$	4.7382	0.093	-	-	-	0.090636	<b>0.091607</b>	<b>0.091607</b>
Mean Std.	$x$	8448.9772	< 0.05	< 0.05	< 0.05	< 0.05	37.83216	18.310809	<b>18.310721</b>
	$y$	9174.1494	< 0.05	< 0.05	< 0.05	< 0.05	36.089137	17.615347	<b>17.615256</b>
	$z$	3947.891	< 0.05	< 0.05	< 0.05	0.9	22.715372	17.910431	<b>17.909425</b>

test revealed statistically significant differences ( $p < 0.05$ ) across all metrics except for the mean difference  $z$  feature. The Nemenyi post-hoc test showed that Orekit generally outperformed the others, whereas comparisons between SGP4 and Skyfield were often not significant. Accordingly, we adopted Orekit with one-minute interpolation for all subsequent experiments. A detailed discussion on the choice of interpolation method is provided in Appendix B.

For pre-training, we used 6,955 satellites. The test set consists of 56 satellites, with one randomly selected per launch month. We split the remaining satellites into training and validation sets in an 8:2 ratio; to maintain distributional balance, the validation set was sampled at regular intervals. For normalization, we computed the mean and standard deviation from the pre-training training split and applied the same z-score normalization to the training, validation, and test sets.

### 3 METHOD

#### 3.1 DECOMPOSED ATTENTION FREDFORMER (DAF)

In our research, based on the Frequency Debaised Transformer model (FredFormer) (Piao et al., 2024) as the backbone, we propose the Decomposed Attention FredFormer (DAF) model with several improvements. First, we replace FredFormer’s DFT layer with a RFFT layer. Next, given the periodic nature of satellite orbits, we remove patching and add positional embeddings so that the attention mechanism can account for the input sequence order. Furthermore, since the inputs are z-scaled, we remove layer normalization. In particular, we replace FredFormer’s attention with Tensorized Multi-Head Attention (TMHA) based on TTD. Lastly, the inverse DFT (IDFT) layer was replaced with an inverse RFFT (IRFFT) layer. The overall architecture of the DAF model is shown in Figure 1 (b).

#### 3.2 RATIONALES OF APPLYING RFFT AND IRFFT

First, FredFormer employed the DFT to convert discrete-time signals from the time domain to the complex frequency domain. However, when the input data is real-valued, as in our study, the Hermitian symmetry property allows the use of the RFFT, which preserves the same information as the full DFT while representing it in  $\frac{L}{2} + 1$  complex components (Zhou et al., 2022; Xu et al., 2023). Therefore, we discard both the DFT and IDFT, and adopt the RFFT–IRFFT pair in our architecture. In particular, we apply an RFFT to the input  $X \in \mathbb{R}^{C \times L}$ , transforming it into the frequency domain:

$$\widehat{X} = RFFT(X) \in \mathbb{C}^{C \times (\frac{L}{2} + 1)}, \quad (1)$$

where  $C$  denotes the number of input channels, and  $L$  denotes the historical look-back window. Subsequently, we separate the real part  $R \in \mathbb{R}^{C \times \frac{L}{2}}$  and the imaginary part  $I \in \mathbb{R}^{C \times \frac{L}{2}}$  of  $\widehat{X}$ , and then stack them to obtain  $A \in \mathbb{R}^{C \times L}$ , which is passed through a positional embedding layer to produce  $PE \in \mathbb{R}^{C \times L}$  and then fed into the TMHA layer.

**Algorithm 1** Algorithm of Decomposed Attention FredFormer (DAF).  $C$  denotes the number of channels,  $L$  the look-back window length,  $H$  the prediction horizon,  $\mathcal{H}$  the number of heads,  $D$  the weight matrix dimension, and  $\mathbf{r}$  the TT ranks.  $\mathcal{T}_h(\cdot)$  refers the Tensorized Multi-Head Attention.

---

```

216 1: Input: Historical data  $X \in \mathbb{R}^{C \times L}$ 
217 2: Output: prediction result  $Y \in \mathbb{R}^{C \times H}$ 
218 3: (i) RFFT-IRFFT Backbone
219 4:  $\widehat{X} \leftarrow \text{RFFT}(X) \in \mathbb{C}^{C \times (L/2)}$ 
220 5:  $R \leftarrow \text{Re}(\widehat{X}) \in \mathbb{R}^{C \times (L/2)}$ ,  $I \leftarrow \text{Im}(\widehat{X}) \in \mathbb{R}^{C \times (L/2)}$ 
221 6:  $A \leftarrow \text{Concat}(R, I) \in \mathbb{R}^{C \times L}$ 
222 7:  $PE \leftarrow A + \text{Positional Embedding}$ 
223 8: (ii) Tensorized Multi-Head Attention
224 9:  $W \leftarrow \text{stack}(W_Q, W_K, W_V) \in \mathbb{R}^{3 \times D \times D}$ 
225 10:  $(G_1, G_2, G_3) \leftarrow \text{TTD}(W, \mathbf{r})$ 
226 11: for  $h = 1$  to  $\mathcal{H}$  do
227 12:    $(Q_h, K_h, V_h) \leftarrow \mathcal{T}_h(Z, G_1, G_2, G_3)$ 
228 13:   Conduct scaled dot-product attention
229 14: end for
230 15: (iii) Frequency Summarization
231 16:  $(R' \in \mathbb{R}^{C \times \frac{L}{2}}, I' \in \mathbb{R}^{C \times \frac{L}{2}}) \leftarrow \text{Split}(PE') \in \mathbb{R}^{C \times L}$ 
232 17:  $\widehat{X}' \leftarrow R'W_R + iI'W_I \in \mathbb{C}^{C \times H}$ ,  $W_R, W_I \in \mathbb{R}^{(L/2) \times H}$ 
233 18:  $Y \leftarrow \text{IRFFT}(\widehat{X}') \in \mathbb{R}^{C \times H}$ 
234 19: return  $Y$ 

```

---

Subsequently, after passing through the TMHA layer, the resulting  $PE' \in \mathbb{R}^{C \times L}$  is again separated into the real component  $R' \in \mathbb{R}^{C \times \frac{L}{2}}$  and imaginary component  $I' \in \mathbb{R}^{C \times \frac{L}{2}}$ . Each component passes through a linear layer  $W_R = W_I \in \mathbb{R}^{\frac{L}{2} \times H}$  to obtain:

$$\widehat{X}' = R'W_R + iI'W_I \in \mathbb{C}^{C \times H}. \quad (2)$$

Here,  $H$  denotes the prediction horizon. Finally, by applying an IRFFT to transform the data back from the frequency domain to the time domain, we obtain the final output  $Y \in \mathbb{R}^{C \times H}$ .

### 3.3 POSITION-AWARE LEARNING

Initially, the FredFormer applies patching to  $A \in \mathbb{R}^{C \times L}$ , performs layer normalization, and then feeds the result into the attention layer. Also, FredFormer adopts patching to enable the model to learn complex temporal patterns at both global and local scales. However, in contrast, satellite orbital data exhibit strong periodicity and a relatively simple structure, thereby reducing the importance of local patterns. Therefore, we remove patching and instead add a learnable positional embedding before passing the input to the attention layer. Additionally, since the validation and test inputs are z-scored using the global mean and standard deviation estimated from 5,564 pre-training series, we omit layer normalization to avoid double normalization. Finally, the input to the attention is defined as follows:

$$Z = A + PE \in \mathbb{R}^{C \times L}. \quad (3)$$

### 3.4 TENSORIZED MULTI-HEAD ATTENTION

Similar to matrix decomposition methods such as Singular Value Decomposition (SVD) (Klema & Laub, 1980), Tensor Decomposition decomposes a high-dimensional tensor into smaller core tensors. There are various types of Tensor Decomposition; in our study, we adopt the Tensor Train Decomposition (TTD) (Oseledets, 2011), which is known to be efficient for model compression, helping reduce the number of parameters while maintaining, or even improving, the model's performance. For an  $d$ -dimensional tensor  $\mathcal{X} \in \mathbb{R}^{i_1 \times i_2 \times \dots \times i_d}$ , where  $i_k$  denotes the size of dimension  $k$ , the formula for TTD is given as follows:

$$\mathcal{X}(i_1, i_2, \dots, i_d) \approx G_1(:, i_1, :) G_2(:, i_2, :) \cdots G_d(:, i_d, :), \quad (4)$$

where  $G_k \in \mathbb{R}^{r_{k-1} \times i_k \times r_k}$ . Here, in order to keep the result of all matrix multiplications as a scalar,  $r_0 = r_d = 1$  and  $\{r_k\}_{k=0}^d$  refers to TT-ranks. In the conventional Attention, given the input  $Z$  and the weight matrices  $W_Q \in \mathbb{R}^{D \times D}$ ,  $W_K \in \mathbb{R}^{D \times D}$ , and  $W_V \in \mathbb{R}^{D \times D}$ , we obtain  $Q = ZW_Q$ ,  $K = ZW_K$ , and  $V = ZW_V$ . To apply TTD to the Attention, which refers to Tensorized Attention (TA), we stack the weight matrices  $W_Q$ ,  $W_K$ , and  $W_V$  to form  $W \in \mathbb{R}^{3 \times D \times D}$ . Then, applying Equation 4 with the TT ranks  $\{r_k\}_{k=0}^4$ , we obtain  $G_1 \in \mathbb{R}^{1 \times 3 \times r_1}$ ,  $G_2 \in \mathbb{R}^{r_1 \times D \times r_2}$ , and  $G_3 \in \mathbb{R}^{r_2 \times D \times 1}$ . In TA, new  $Q'$ ,  $K'$ , and  $V'$  are generated from the input  $Z$  by using  $G_1$ ,  $G_2$ , and  $G_3$  instead of  $W_Q$ ,  $W_K$ , and  $W_V$ , and the final output is obtained by performing scaled-dot product attention. The equations for TA are as follows.

$$\text{TA}(Z, G_1, G_2, G_3) = \text{softmax} \left( \frac{Q'K'}{\sqrt{d_h}} \right) V', \quad s.t.$$

$$[Q', K', V']_{ndj} = \sum_{a=1}^D \sum_{\alpha=1}^{r_1} \sum_{\beta=1}^{r_2} Z_{na} G_1(1, j, \alpha) G_2(\alpha, a, \beta) G_3(\beta, d, 1), \quad (5)$$

where  $a, d \in \{1, \dots, D\}$  are weight dimension index, and  $n \in \{1, \dots, N\}$  is the sequence index,  $j \in \{1, 2, 3\}$  selects  $Q'$ ,  $K'$ ,  $V'$ ,  $\alpha \in \{1, \dots, r_1\}$ ,  $\beta \in \{1, \dots, r_2\}$  are the middle ranks of TT rank  $\{r_0, r_1, r_2, r_3\}$  and  $d_h$  denotes attention key dimension per number of heads.

Extending Tensorized Attention to TMHA, we compose the heads multiplicatively and compute in batched form, resulting in a high-dimensional configuration. We denote this high-dimensional tensor as  $Z \in \mathbb{R}^{3 \times B \times C \times D}$ . For the operations of TMHA, we divide the second dimension of  $Z \in \mathbb{R}^{3 \times B \times C \times D}$  proportionally to the number of heads  $h$ , resulting in a tensor to be restructured to  $Z \in \mathbb{R}^{3 \times \frac{B}{h} \times h \times C \times D}$ , where  $B$  represents the batch size,  $C$  denotes the dimension of the input data. The resulting  $Z$ , along with  $G_1, G_2, G_3$ , is calculated using the following Equation (6):

$$\text{TMHA}(Z, G_1, G_2, G_3) = \text{Concat}_{i=1}^{\text{head}} (\text{TA}_i(Z_i, G_1, G_2, G_3)), \quad (6)$$

where  $\text{TA}_i$  denotes the  $i$ -th head's TA, as defined in Equation 5. The output from the TMHA layer is passed through separate linear layers for the real and imaginary parts to generate the corresponding predictions, which are then fed into the IRFFT layer for decoding.

## 4 EXPERIMENTS

In this section, we describe the experimental setup and results for 14 models, including our model. Section 4.1 outlines the experimental settings for both state-of-the-art large time-series prediction models and the commonly used baseline approach. Section 4.2 presents results on the Starlink dataset for nine fine-tuned large time-series baselines and six pre-trained models (including ours). As shown in Figure 1 (c), Section 4.3 reports zero-shot evaluation on seven cross-domain datasets and three real-world datasets. Section 4.4 provides an ablation study on 11 datasets (including Starlink) for FredFormer and its three variants, along with an analysis of parameter reduction achieved by applying the tensor decompositions. Finally, Section 4.4 investigates the effect of PE in the frequency domain.

### 4.1 EXPERIMENTS SETUP AND BASELINE

To evaluate the practicality of the proposed method, we conduct a comparative analysis against 14 baselines using the Starlink test set and 10 cross-domain datasets. The cross-domain setting is used to measure zero-shot performance, and we also consider real-world applicability using three KOMPSAT datasets (3, 3A, 5). Evaluation is based on MAE and MSE metrics. In the space domain, an error as small as 0.001 can translate to a physical deviation of several to tens of kilometers; therefore, we report our results to the sixth decimal place for a more precise comparison. All tests

are conducted on an NVIDIA GeForce RTX 3090 GPU under identical conditions to measure model inference time for orbit prediction.

We use the five conventional time-series prediction models and nine large time-series models as baselines. The conventional models include GRU (Dey & Salem, 2017), LSTM (Hochreiter & Schmidhuber, 1997), DLinear (Zeng et al., 2023), SparseTSF (Lin et al., 2024), and FredFormer (Piao et al., 2024). We trained these models from scratch on the pre-training dataset. The large time-series models include Time-MoE, Timer, Timer-XL, Time-LLM, TTM, MOMENT, MOIRAI, and TimesFM. For fine-tuning, we used 1/100 of the pre-training dataset. The look-back window is  $L = 512$ , and the prediction horizon is  $H = 90$ . We adopt a sliding-window approach with a stride of  $L + H$  to ensure non-overlapping segments. All models are trained for 100 epochs, starting from the optimal hyperparameters reported in the respective papers, which were further tuned to ensure loss convergence on the Starlink dataset. For univariate models, we report the average performance across the six features, while inference time is measured as the total over all six. We perform all tests under the same settings and conditions. Detailed experimental configurations are provided in the Appendix D.

## 4.2 MAIN RESULT

Table 3 reports the mean error over three runs on the Starlink dataset. The DAF model achieved the best performance across both MSE and MAE. TimesFM ranked second in both metrics. However, our model achieved relative improvements of 34.85% in MSE and 16.01% in MAE compared to TimesFM, while being approximately  $2,221\times$  smaller in parameter count. In contrast, although SparseTSF uses only 94 parameters, it recorded an MSE of 0.004167, which is significantly worse than that of our model.

Table 3: Prediction results on Starlink dataset in MSE and MAE of each model. The rightmost column shows the total parameter count of each model, and the bottom row ‘‘Improvements’’ highlights the performance gain of our model over the best baseline for each metric. The best result is highlighted in **bold**, and the second-best is highlighted with underline.

Metric	MSE	MAE	Parameter
Time-MoE	2.265333 $\pm$ 0.026844	1.263860 $\pm$ 0.009000	113,352,192
Timer	0.327132 $\pm$ 0.004084	0.484511 $\pm$ 0.003660	67,397,728
Timer-XL	0.081020 $\pm$ 0.050731	0.220013 $\pm$ 0.096210	67,398,880
Time-LLM	1.025519 $\pm$ 0.077263	0.876721 $\pm$ 0.046888	105,321,802
TTM	0.499960 $\pm$ 0.378112	0.555809 $\pm$ 0.302664	16,344,891
TimesFM	0.001506 $\pm$ 0.000157	0.028408 $\pm$ 0.001259	498,828,960
MOMENT	0.010566 $\pm$ 0.000856	0.086805 $\pm$ 0.003665	16,085,594
MOIRAI	0.027327 $\pm$ 0.017312	0.118248 $\pm$ 0.045438	15,855,786
GRU	0.002177 $\pm$ 0.000162	0.033817 $\pm$ 0.000262	121,884
LSTM	0.003582 $\pm$ 0.001138	0.043888 $\pm$ 0.007628	139,292
DLinear	0.002516 $\pm$ 0.000032	0.036839 $\pm$ 0.000260	554,040
SparseTSF	0.004167 $\pm$ 0.000014	0.052828 $\pm$ 0.000070	94
FredFormer	0.034392 $\pm$ 0.002362	0.163769 $\pm$ 0.005493	290,194
<b>DAF (Ours)</b>	<b>0.000981 <math>\pm</math> 0.000030</b>	<b>0.023853 <math>\pm</math> 0.000072</b>	224,556
Improvements (%)	34.85 $\pm$ 3.79	16.01 $\pm$ 4.44	-

## 4.3 ZERO-SHOT EVALUATION

Furthermore, Table 4 presents the zero-shot performance of the models trained on the Starlink dataset when evaluated on seven cross-domain satellite datasets. The proposed model delivers the best results on most datasets despite their disparate specifications and operational purposes. It also achieves a 62.81% improvement on the LEMUR dataset. Finally, Table 5 presents a further comparison on three high-precision datasets with analytically computed orbits rather than via interpolation. Despite being real-world datasets, the proposed model achieves the lowest prediction errors in nearly every case.

## 4.4 ABLATION STUDY

In this section, we evaluate the effectiveness of the two core structures of the DAF model. The core structure of the DAF model can be grouped into two components: (1) a layer variant that replaces DFT with RFFT and removes patching and layer normalization, and (2) applying TTD to the Multi-Head Attention. Additionally, we consider a third case that applies Canonical Polyadic Decomposition (CPD) (Hitchcock, 1924) in place of TTD. The explanation of CPD and its application to TMHA is provided in Appendix C. Across 11 datasets, including Starlink, the experimental results are reported in Table 6, where we denote the four configurations as: FredFormer, FredFormer (layer variant), FredFormer (layer variant + CPD), and FredFormer (layer variant + TTD, DAF). Our DAF model achieved the best performance in 6 out of 11

Table 4: Zero-shot evaluation on seven cross-domain datasets in MSE and MAE of each model. The bottom row “Improvements” highlights the performance gain of our model over the best baseline for each dataset and metric. The best result is highlighted in **bold**, and the second-best is highlighted with underline.

Dataset Metric	ASTROCAST		CAPELLA		ICEYE		KINEIS		KUIPER		LEMUR		SKYSAT	
	MSE	MAE												
Time-MoE	2.209372	1.171837	2.295843	1.268301	2.251507	1.195950	2.384254	1.242035	2.213490	1.246838	2.259241	1.187898	2.237188	1.191708
Timer	0.090851	0.217582	0.535211	0.600972	0.285239	0.397307	0.941015	0.808324	0.109496	0.272394	0.295066	0.417342	0.157977	0.303088
Timer-XL	0.036722	0.139096	0.110913	0.252742	0.071666	0.192768	0.177071	0.321643	0.044190	0.159048	0.075421	0.198898	0.051863	0.168050
Time-LLM	0.964805	0.803355	1.061852	0.886094	1.016547	0.830448	1.148139	0.883287	0.963177	0.851150	1.023576	0.826542	0.990677	0.821108
TTM	0.225226	0.347945	0.642272	0.614609	0.430893	0.476646	0.951737	0.752327	0.260675	0.404975	0.454790	0.493409	0.321774	0.428811
TimesFM	0.002258	0.028439	0.006887	0.062037	0.002702	0.036138	0.024786	0.124216	0.004377	0.031227	0.002977	0.033964	0.002506	0.030870
MOMENT	0.012305	0.085201	0.010939	0.085381	0.011772	0.084199	0.016214	0.093539	0.016863	0.093891	0.012093	0.083998	0.012131	0.086694
MOIRAI	0.011396	0.068974	0.042769	0.144607	0.025507	0.103988	0.078260	0.200425	0.014616	0.080221	0.026792	0.107371	0.017238	0.087685
GRU	0.002624	0.038205	0.004105	0.046758	0.003493	0.042391	0.010210	0.074563	0.001242	0.022845	0.003085	0.038363	0.001888	0.032268
LSTM	0.002476	0.036840	0.006965	0.061469	0.004873	0.049932	0.016452	0.096129	0.001765	0.028110	0.004761	0.049704	0.002855	0.038525
DLlinear	0.000448	0.013369	0.007117	0.060827	0.002692	0.031739	0.018088	0.099608	0.000520	<b>0.010687</b>	0.002533	0.030934	0.001126	0.016291
SparseTSF	0.000597	0.017675	0.013467	0.092643	0.004657	0.046133	0.033884	0.152013	0.000553	0.012730	0.004201	0.043359	0.001798	0.020233
FredFormer	0.035184	0.156715	0.031599	0.155111	0.034384	0.155599	0.030640	0.144051	0.035653	0.016794	0.034760	0.155006	0.036094	0.159925
DAF (Ours)	<b>0.000302</b>	<b>0.011767</b>	<b>0.002860</b>	<b>0.037055</b>	<b>0.001004</b>	<b>0.021340</b>	<b>0.009251</b>	<b>0.066928</b>	<b>0.000474</b>	0.012720	<b>0.000942</b>	<b>0.020974</b>	<b>0.000628</b>	<b>0.015339</b>
Improvements (%)	32.59	11.97	59.81	39.08	62.70	32.76	9.40	10.22	8.85	-19.00	62.81	32.20	44.25	5.84

Table 6: Ablation Study in MSE and MAE for four variants of FredFormer on 11 datasets, including Starlink, cross-domain, and real-world datasets. The bottom row shows the number of datasets on which each variant achieved the best performance. The best result is highlighted in **bold**, and the second-best is highlighted with underline.

Model Metric	FredFormer		FredFormer (Layer Variant)		FredFormer (Layer Variant + CPD)		FredFormer (Layer Variant + TTD, DAF)	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Starlink	0.034392	0.163769	0.001061	0.025258	0.001013	0.024449	<b>0.000981</b>	<b>0.023853</b>
K3	0.032467	0.147525	0.014338	0.089235	<b>0.010207</b>	<b>0.071267</b>	0.014248	0.084121
K3A	0.035877	0.158858	0.000748	0.021877	<b>0.000714</b>	0.020978	0.000738	<b>0.020709</b>
K5	0.034810	0.156403	0.001203	0.027767	0.001078	0.025591	<b>0.001042</b>	<b>0.024979</b>
ASTROCAST	0.035184	0.156715	0.000340	0.012795	0.000343	0.012819	<b>0.000302</b>	<b>0.011767</b>
CAPELLA	0.031599	0.155111	0.002685	0.038391	<b>0.002288</b>	<b>0.034798</b>	0.002860	0.037055
ICEYE	0.034384	0.155599	0.001184	0.023629	0.001043	0.022069	<b>0.001004</b>	<b>0.021340</b>
KINEIS	0.030640	0.144051	0.008261	0.068009	<b>0.006156</b>	<b>0.055699</b>	0.009251	0.066928
KUIPER	0.035653	0.016794	0.000487	0.012890	0.000527	0.014434	<b>0.000474</b>	<b>0.012720</b>
LEMUR	0.034760	0.155006	0.001111	0.023197	0.000981	0.021694	<b>0.000942</b>	<b>0.020974</b>
SKYSAT	0.036094	0.159925	0.000633	0.015860	<b>0.000605</b>	0.016552	0.000628	<b>0.015339</b>
1st Count	0	0	0	0	5	3	6	8

datasets based on MSE and in 8 datasets based on MAE. Furthermore, except for the CAPELLA dataset, it consistently ranked second-best even when it did not achieve the top performance.

The FredFormer (Layer Variant + CPD) model was the only model, besides DAF, to achieve the best performance in at least one case. Specifically, it achieved the best performance across 5 datasets based on MSE and across 3 datasets based on MAE. These results demonstrate that modifying the layers of the FredFormer model and applying TTD to the MHA significantly enhances performance. The results in Table 7 show the parameter counts for FredFormer and its three variants. FredFormer has the most parameters, totaling 290,194.

The FredFormer (Layer Variant), which applies RFFT, halves the size of the linear layer after the TMHA layer, then applies IRFFT to R and I, resulting in 224,992 parameters. Our DAF model further reduces the parameter count to 224,556 by applying TTD to FredFormer (Layer Variant), demonstrating improved performance with fewer parameters. Meanwhile, FredFormer (Layer Vari-

Table 5: Zero-shot evaluation on three real-world datasets in MSE and MAE of each model. “Improvements” presents the relative performance of our model compared to the best baseline for each dataset and metric. The best result is highlighted in **bold**, and the second-best is highlighted with underline.

Dataset	KOMPSAT-3		KOMPSAT-3A		KOMPSAT-5	
	MSE	MAE	MSE	MAE	MSE	MAE
Time-MoE	2.418150	1.250692	2.276393	1.198948	2.296285	1.207763
Timer	1.180190	0.905896	0.292992	0.447280	0.390626	0.517673
Timer-XL	0.211756	0.350651	0.078914	0.211551	0.094826	0.232320
Time-LLM	1.179187	0.894232	1.034300	0.835688	1.056063	0.845372
TTM	1.061232	0.797521	0.493810	0.537106	0.582112	0.583277
TimesFM	0.044738	0.169868	0.001129	0.024734	0.001127	<b>0.024651</b>
MOMENT	0.023447	0.116151	0.010463	0.083579	0.010273	0.081888
MOIRAI	0.098327	0.228442	0.026376	0.113609	0.033509	0.128145
GRU	<b>0.014227</b>	0.089150	0.001693	0.030145	0.002543	0.036848
LSTM	0.020632	0.107698	0.003804	0.046117	0.005183	0.053863
DLlinear	0.027910	0.124509	0.001662	0.029771	0.003101	0.040915
SparseTSF	0.053629	0.191637	0.002394	0.040606	0.005025	0.058457
FredFormer	0.032467	0.147525	0.035877	0.158858	0.034810	0.156403
DAF (Ours)	0.014248	<b>0.084121</b>	<b>0.000738</b>	<b>0.020709</b>	<b>0.001042</b>	0.024979
Improvements (%)	-0.15	5.64	34.63	16.27	7.54	-1.33

Table 7: Comparison of parameter counts across FredFormer and its variants.

Model	Parameter
FredFormer	290,194
FredFormer (Layer Variant)	224,992
FredFormer (Layer Variant + CPD)	193,324
FredFormer (Layer Variant + TTD, DAF (Ours))	224,556

Table 9: Total inference time (in seconds) on 11 datasets, including cross-domain and real-world datasets. Additionally, the computational cost is provided in GFLOPs (per sequence) calculated with a batch size of 16 for each model. The lowest values (fastest time and lowest computational cost) are highlighted in **bold**, and the second-lowest are highlighted with underline.

Dataset	Inference Time (s)											GFLOPs
	Starlink	ASTROCAST	CAPELLA	ICEYE	KINEIS	KUIPER	LEMUR	SKYSAT	KOMPSAT-3	KOMPSAT-3A	KOMPSAT-5	
Time-MoE	9496.95	402.97	123.69	972.14	634.40	458.96	1373.46	389.80	345.77	343.28	338.78	<u>5353.2000</u>
Timer	30.24	1.79	0.89	3.75	2.57	2.00	4.97	1.77	1.65	1.66	1.63	25.1553
Timer-XL	40.13	2.21	1.01	4.78	3.26	2.46	6.39	2.16	1.97	2.00	1.98	25.1557
Time-LLM	671.55	30.03	9.55	73.41	47.07	33.85	100.38	28.79	25.75	25.63	25.79	954.4704
TTM	149.38	6.54	2.14	15.82	10.25	7.43	22.07	6.38	5.65	5.63	5.65	62.1760
TimesFM	756.06	35.76	13.62	83.76	54.40	39.90	116.59	35.04	31.04	31.18	31.04	517.2000
MOMENT	13.64	1.27	0.78	2.05	1.56	1.30	2.70	1.15	1.05	1.12	1.10	7.0992
MOIRAI	16.05	1.27	0.60	2.02	1.52	1.25	2.73	1.08	1.03	1.01	0.98	7.1008
GRU	3.86	0.23	0.04	0.38	<b>0.02</b>	0.17	0.53	0.18	0.12	0.12	0.12	0.4368
LSTM	<b>3.84</b>	<b>0.15</b>	0.04	0.41	0.24	0.16	0.58	0.15	0.12	0.12	0.12	0.5824
DLinear	5.12	0.45	0.04	0.38	0.25	0.17	0.64	0.15	0.14	0.13	0.12	0.0096
SparseTSF	4.62	0.64	<b>0.03</b>	<b>0.31</b>	0.29	<b>0.14</b>	<b>0.42</b>	<b>0.12</b>	<b>0.11</b>	<b>0.11</b>	<b>0.11</b>	<b>0.0016</b>
FredFormer	5.45	0.46	0.05	0.47	0.31	0.23	0.65	0.15	0.14	0.14	0.14	0.0240
<b>DAF (Ours)</b>	3.86	0.16	0.05	0.39	0.24	0.17	0.55	0.14	0.12	0.17	0.13	<b>0.0144</b>

ant + CPD) exhibits even fewer parameters than DAF and achieves better performance than FredFormer (Layer Variant). However, since FredFormer (Layer Variant + CPD) does not outperform the DAF model, the DAF model is more suitable for satellite large time-series prediction tasks.

#### 4.5 EFFECT OF POSITIONAL EMBEDDING IN FREQUENCY DOMAIN

To address the role of PE as a replacement for patching and the effectiveness of frequency-domain representation, we conduct experiments comparing three variants: (1) DAF-timePE with time-domain input and PE (DAF (no RFFT)), (2) DAF-noPE with RFFT and original patching (DAF (no PE)), and (3) DAF with RFFT and PE (DAF (ours)).

The results demonstrate that DAF (ours) reduces MSE by 3.1% compared to DAF (no PE) while using 3,328 fewer parameters, validating that PE is a more parameter-efficient alternative to patching. Furthermore, DAF (ours) outperforms DAF (no RFFT) by 23.5%, confirming that the frequency-domain representation enabled by RFFT is effective for accurate orbit prediction. The slight parameter increase of DAF (ours) over DAF (no RFFT) (256 parameters) arises from the RFFT output dimension: an input sequence of length  $N$  produces  $N/2 + 1$  complex frequency components, which are represented as  $N + 2$  real values after concatenating real and imaginary parts. In summary, DAF (ours) achieves the best performance with comparable parameter count, validating that the combination of RFFT and PE is efficient for satellite orbit prediction.

Table 8: Ablation study on PE and frequency transformation

Model Variant	MSE	MAE	Parameter
DAF (no RFFT)	0.001282	0.027066	224,300
DAF (no PE)	0.001012	0.024515	227,884
DAF (ours)	<b>0.000981</b>	<b>0.023853</b>	224,556

## 5 DISCUSSION AND FUTURE WORK

Our research shows that the proposed model DAF, trained solely on the Starlink dataset, outperforms existing large time-series models. As shown in Table 9, it delivers faster inference through a smaller parameter count and achieves GFLOPs per sequence comparable to lightweight models, making it highly efficient and readily deployable. DAF’s consistent performance across diverse satellites confirms both its robustness and suitability for real mission operation.

However, we acknowledge specific scenarios where the model may face challenges. First, regarding sudden orbital maneuvers, the model may struggle to predict orbit changes resulting from collision avoidance maneuvers if such dynamic events are absent from the training data. Second, extreme environmental variations can impact performance. Severe fluctuations in the space environment, such as rapid changes in atmospheric density caused by intense solar activity, can create out-of-distribution scenarios that may increase prediction errors.

For future work, we aim to address these limitations and move beyond the proposed large model to develop a foundation model for satellite orbit prediction that generalizes across Low, Medium, and Geostationary Earth Orbits. This model can be pre-trained on the large-scale Starlink dataset, followed by transfer learning to accurately predict orbits across satellites with different orbital regimes

486 and mission types, without additional supervision. Furthermore, we plan to extend the model to  
487 support uncertainty quantification. In mission-critical operations such as collision avoidance, es-  
488 timating the reliability of a predicted position is as important as the prediction itself. To address  
489 this, we will explore methods to transition from deterministic to probabilistic forecasting, such as  
490 modifying the output layer to estimate Gaussian distribution parameters (mean and variance) us-  
491 ing negative log-likelihood loss or employing quantile loss to provide confidence intervals. Finally,  
492 we will validate the model’s practical applicability by conducting experiments on a broader set of  
493 high-precision, real-world satellite orbit datasets. Moreover, we aim to expand the domain scope  
494 and evolve the proposed framework into a large-scale model applicable to general time-series pre-  
495 diction tasks beyond orbital dynamics. Given the model’s capability to capture inherent periodicity  
496 and long-range dependencies via frequency-domain analysis, we anticipate its applicability to other  
497 quasi-periodic domains, such as meteorology, energy consumption, and signal processing.

## 498 6 CONCLUSION

500 In this work, we propose the Decomposed Attention FredFormer (DAF) model, a large time-series  
501 prediction model for satellite orbit prediction. DAF incorporates Tensorized Multi-Head Attention  
502 and RFFT for better frequency modeling and removes layer normalization and positional embedding  
503 to replace patching. Our design choices yield improved accuracy and efficiency, while significantly  
504 reducing model size. To validate our design, we train the model solely on the large-scale Starlink  
505 dataset and conduct comparative evaluations, including zero-shot prediction, on eleven satellite orbit  
506 datasets with varying specifications. Our DAF achieves the best performance by a large margin  
507 in most experiments. Despite being pre-trained as a large model on high-volume satellite data,  
508 our method achieves the broad generalizability across diverse orbital regimes with an optimally  
509 compressed parameter space, resulting in a compact and deployable architecture. We believe our  
510 work paves the way for demonstrating the potential of large models in satellite orbit prediction.

## 511 REFERENCES

- 513 Aaron C. Boley and Michael Byers. Satellite mega-constellations create risks in low earth orbit, the  
514 atmosphere and on earth. *Scientific Reports*, 11(1):10642, 2021. ISSN 2045-2322. doi: 10.1038/  
515 s41598-021-89909-7. URL <https://doi.org/10.1038/s41598-021-89909-7>.
- 517 Francisco Caldas and Claudia Soares. Machine learning in orbit estimation: A survey. *Acta Astro-*  
518 *nautica*, 220, 04 2024a. doi: 10.1016/j.actaastro.2024.03.072.
- 519 Francisco Caldas and Cláudia Soares. Precise and efficient orbit prediction in leo with machine  
520 learning using exogenous variables. In *2024 IEEE Congress on Evolutionary Computation (CEC)*,  
521 pp. 1–8, 2024b. doi: 10.1109/CEC60901.2024.10611996.
- 523 Junyu Chen, Baolin Wu, Zhaobo Sun, and Danwei Wang. Distributed safe trajectory optimization for  
524 large-scale spacecraft formation reconfiguration. *Acta Astronautica*, 214:125–136, 2024. ISSN  
525 0094-5765. doi: 10.1016/j.actaastro.2023.10.012.
- 526 George Choumos, Konstantinos Tsaprailis, Vaios Lappas, and Charalampos Kontoes. *Artificial*  
527 *Intelligence for a Safe Space: Data and Model Development Trends in Orbit Prediction and*  
528 *Collision Avoidance*. American Institute of Aeronautics and Astronautics (AIAA), 2024. doi:  
529 10.2514/6.2024-2066. URL [https://arc.aiaa.org/doi/abs/10.2514/6.2024-2](https://arc.aiaa.org/doi/abs/10.2514/6.2024-2066)  
530 [066](https://arc.aiaa.org/doi/abs/10.2514/6.2024-2066).
- 531 Giacomo Curzi, Dario Modenini, and Paolo Tortora. Two-line-element propagation improvement  
532 and uncertainty estimation using recurrent neural networks. *CEAS Space Journal*, 14(1):197–204,  
533 2022.
- 535 Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for  
536 time-series forecasting, 2024. URL <https://arxiv.org/abs/2310.10688>.
- 538 Rahul Dey and Fathi M. Salem. Gate-variants of gated recurrent unit (gru) neural networks. In *2017*  
539 *IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 1597–  
1600, 2017. doi: 10.1109/MWSCAS.2017.8053243.

- 540 Vijay Ekambaram, Arindam Jati, Pankaj Dayama, Sumanta Mukherjee, Nam H Nguyen, Wesley M.  
541 Gifford, Chandra Reddy, and Jayant Kalagnanam. Tiny time mixers (TTMs): Fast pre-trained  
542 models for enhanced zero/few-shot forecasting of multivariate time series. In *The Thirty-eighth*  
543 *Annual Conference on Neural Information Processing Systems*, 2024. URL <https://openreview.net/forum?id=3O5YCEWETq>.  
544
- 545 Agence France-Presse. As satellite numbers soar, experts warn earth’s space may soon hit its limit,  
546 May 2025. URL [https://www.ndtv.com/science/as-satellite-numbers-s](https://www.ndtv.com/science/as-satellite-numbers-soar-experts-warn-earths-space-may-soon-hit-its-limit-8451297)  
547 [oar-experts-warn-earths-space-may-soon-hit-its-limit-8451297](https://www.ndtv.com/science/as-satellite-numbers-soar-experts-warn-earths-space-may-soon-hit-its-limit-8451297).  
548 Accessed: 2025-08-05.  
549
- 550 M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of  
551 variance. *American Statistical Association*, 32:675–701, 1937. doi: 10.1080/01621459.1937.10  
552 503522. URL <http://dx.doi.org/10.1080/01621459.1937.10503522>.
- 553 Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski.  
554 Moment: a family of open time-series foundation models. In *Proceedings of the 41st International*  
555 *Conference on Machine Learning*, ICML’24. JMLR.org, 2024.  
556
- 557 Sebastian Grau, Anton Johann Große Siestrup, Victoria Koßack, Jan Plewe, Vincent Schulz, Nurçah  
558 Solak, Johanna Erna Charlotte Teuchert, Andreas Hornig, and Enrico Stoll. PRecise IN-orbit  
559 Collision prediction and space Environment Surveillance System. In Max Petrozzi-Ilstad (ed.),  
560 *Small Satellites Systems and Services Symposium (4S 2024)*, volume 13546, pp. 1354647. In-  
561 ternational Society for Optics and Photonics, SPIE, 2025. doi: 10.1117/12.3062782. URL  
562 <https://doi.org/10.1117/12.3062782>.
- 563 Jouni Hartikainen and Simo Särkkä. Kalman filtering and smoothing solutions to temporal gaussian  
564 process regression models. In *Proceedings of the 2010 IEEE International Workshop on Machine*  
565 *Learning for Signal Processing*, 2010.
- 566 F. L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathe-*  
567 *matics and Physics*, 6:164–189, 1924.  
568
- 569 Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):  
570 1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL  
571 <https://doi.org/10.1162/neco.1997.9.8.1735>.
- 572 Felix R. Hoots and Ronald L. Roehrich. Models for propagation of norad element sets. Project  
573 Space-Track Report No. 3, 1980.  
574
- 575 Becky Iannotta. U.s. satellite destroyed in space collision, February 2009. URL <https://www.space.com/5542-satellite-destroyed-space-collision.html>. Accessed:  
576 2025-08-01.  
577
- 578 Giridhar Jadala, Gowri Namratha Meedinti, and Radhakrishnan Delhibabu. Satellite orbit prediction  
579 using a machine learning approach. In *ICAI Workshops*, pp. 28–46, 2022.  
580
- 581 SeungWon Jeong, Soyeon Woo, Daewon Chung, Simon S. Woo, and Youjin Shin. Decomposed  
582 attention segment recurrent neural network for orbit prediction. In *Proceedings of the 30th ACM*  
583 *SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD ’24, pp. 5172–5182, New  
584 York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704901. doi: 10.1  
585 145/3637528.3671546. URL <https://doi.org/10.1145/3637528.3671546>.
- 586 Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen,  
587 Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-LLM: Time series forecasting  
588 by reprogramming large language models. In *The Twelfth International Conference on Learning*  
589 *Representations*, 2024. URL <https://openreview.net/forum?id=Unb5CVPtae>.
- 590 Sajjad Kazemi, Nasser L Azad, K Andrea Scott, Haroon B Oqab, and George B Dietrich. Orbit  
591 determination for space situational awareness: A survey. *Acta Astronautica*, 222:272–295, 2024.  
592
- 593 Donald J Kessler and Burton G Cour-Palais. Collision frequency of artificial satellites: The creation  
of a debris belt. *Journal of Geophysical Research: Space Physics*, 83(A6):2637–2646, 1978.

- 594 Nadim Khairallah and Zaher M Kassas. Ephemeris closed-loop tracking of leo satellites with pseu-  
595 dorange and doppler measurements. In *Proceedings of the 34th international technical meeting*  
596 *of the satellite division of the Institute of Navigation (ION GNSS+ 2021)*, pp. 2544–2555, 2021.  
597
- 598 V. Klema and A. Laub. The singular value decomposition: Its computation and some applications.  
599 *IEEE Transactions on Automatic Control*, 25(2):164–176, 1980. doi: 10.1109/TAC.1980.11023  
600 14.
- 601 Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang. Sparsetsf: modeling  
602 long-term time series forecasting with 1k parameters. In *Proceedings of the 41st International*  
603 *Conference on Machine Learning, ICML’24*. JMLR.org, 2024.
- 604 Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Timer-XL: Long-  
605 context transformers for unified time series forecasting. In *The Thirteenth International Confer-*  
606 *ence on Learning Representations*, 2025. URL [https://openreview.net/forum?id=](https://openreview.net/forum?id=MKCJXj1DDR)  
607 [MKCJXj1DDR](https://openreview.net/forum?id=MKCJXj1DDR).
- 608 maisonobe, Bryan Cazabonne, Evan Ward, Maxime Journot, Romain Serra, Vincent Cucchiatti,  
609 Sébastien Dinot, Thomas Neidhart, Guilhem Bonnefille, Clément Jonglez, Luc Maisonobe, yje-  
610 and, Julio Hernanz, lirw1984, Andrew Goetz, pallepop, gaetanpierre0, Guiuux, Lars Næsbye  
611 Christensen, Alberto Fossà, jvalet, Alberto Ferrero, gabb5, liscju, Christopher Schank, Tanner  
612 Mills, plan3d, Vyom Yadav, Shiva Iyer, and Petrus Hyvönen. Cs-si/orekit: 12.1.2, July 2024.  
613 URL <https://doi.org/10.5281/zenodo.12737944>.
- 614
- 615 NASA Orbital Debris Program Office. Orbital debris quarterly news, volume 15, issue 3, july 2011,  
616 July 2011. URL [https://orbitaldebris.jsc.nasa.gov/quarterly-news/pd](https://orbitaldebris.jsc.nasa.gov/quarterly-news/pdfs/odqnv15i3.pdf)  
617 [fs/odqnv15i3.pdf](https://orbitaldebris.jsc.nasa.gov/quarterly-news/pdfs/odqnv15i3.pdf). Accessed: 2025-08-01.
- 618
- 619 Peter Bjorn Nemenyi. *Distribution-Free Multiple Comparisons*. PhD thesis, Princeton University,  
620 1963.
- 621 I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–  
622 2317, 2011.
- 623
- 624 Hao Peng and Xiaoli Bai. Exploring capability of support vector machine for improving satellite  
625 orbit prediction accuracy. *Journal of Aerospace Information Systems*, 15(6):366–381, 2018. doi:  
626 10.2514/1.1010616.
- 627
- 628 Hao Peng and Xiaoli Bai. Gaussian processes for improving orbit prediction accuracy. *Acta Astro-*  
*nautica*, 161:44–56, 2019. ISSN 0094-5765. doi: 10.1016/j.actaastro.2019.05.014.
- 629
- 630 Hao Peng and Xiaoli Bai. Fusion of a machine learning approach and classical orbit predictions.  
631 *Acta Astronautica*, 184:222–240, 2021. ISSN 0094-5765. doi: 10.1016/j.actaastro.2021.04.017.
- 632
- 633 Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer:  
634 Frequency debiased transformer for time series forecasting. In *Proceedings of the 30th ACM*  
*SIGKDD Conference on Knowledge Discovery and Data Mining, KDD ’24*, pp. 2400–2410, New  
635 York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704901. doi: 10.1  
636 145/3637528.3671928. URL <https://doi.org/10.1145/3637528.3671928>.
- 637
- 638 Jaakko Pihlajasalo, Helena Leppäkoski, Simo Ali-Löyty, and Robert Piché. Improvement of gps  
639 and beidou extended orbit predictions with cnns. In *2018 European navigation conference (ENC)*,  
640 pp. 54–59. IEEE, 2018.
- 641
- 642 Tereza Pultarova. Starlink satellites: Facts, tracking and impact on astronomy, 2025. URL <https://www.space.com/spacex-starlink-satellites.html>. Accessed: Accessed:  
643 2025-08-05.
- 644
- 645 Nathan Reiland, Aaron J. Rosengren, Renu Malhotra, and Claudio Bombardelli. Assessing and  
646 minimizing collisions in satellite mega-constellations. *Advances in Space Research*, 67(11):3755–  
647 3774, 2021. ISSN 0273-1177. doi: <https://doi.org/10.1016/j.asr.2021.01.010>. URL <https://www.sciencedirect.com/science/article/pii/S0273117721000326>.  
Satellite Constellations and Formation Flying.

- 648 Brandon Rhodes. Skyfield: High precision research-grade positions for planets and Earth satellites  
649 generator. Astrophysics Source Code Library, record ascl:1907.024, July 2019.  
650
- 651 Juan F San-Juan, Ivan Perez, Eliseo Vergara, M Montserrat San, R López, Alexander Wittig, and  
652 D Izzo. Hybrid sgp4 propagator based on machine-learning techniques applied to galileo-type  
653 orbits. In *69th international astronomical congress, bremen, Germany*, volume 4, 2018.
- 654 Xiaoming Shi, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen, and Ming Jin. Time-  
655 moe: Billion-scale time series foundation models with mixture of experts. In *The Thirteenth  
656 International Conference on Learning Representations*, 2025. URL [https://openreview  
657 .net/forum?id=e1wDDFm1Vu](https://openreview.net/forum?id=e1wDDFm1Vu).
- 658 Space-Track. Space-track help documentation: Api and tle format, 2025. URL [https://www.  
659 space-track.org/documentation](https://www.space-track.org/documentation). Accessed: 2025-07-23.  
660
- 661 David Vallado, Paul Crawford, Ricahrd Hujsak, and Thomas Sean Kelso. Revisiting spacetrack  
662 report# 3. In *AIAA/AAS astrodynamics specialist conference and exhibit*, pp. 6753, 2006.  
663
- 664 Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo.  
665 Unified training of universal time series forecasting transformers. In *Proceedings of the 41st  
666 International Conference on Machine Learning, ICML'24*. JMLR.org, 2024.
- 667 Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with 10k parameters, 2023.  
668 URL <https://openreview.net/forum?id=m2getD1hpk>.
- 669 Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series  
670 forecasting? In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence  
671 and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth  
672 Symposium on Educational Advances in Artificial Intelligence, AAAI'23/IAAI'23/EAAI'23*.  
673 AAAI Press, 2023. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i9.26317. URL  
674 <https://doi.org/10.1609/aaai.v37i9.26317>.
- 675 Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency  
676 enhanced decomposed transformer for long-term series forecasting. In *Proc. 39th International  
677 Conference on Machine Learning (ICML 2022)*, 2022.  
678
- 679 Xingyu Zhou, Tong Qin, Mingjiang Ji, and Dong Qiao. A lstm assisted orbit determination algorithm  
680 for spacecraft executing continuous maneuver. *Acta Astronautica*, 204:568–582, 2023. ISSN  
681 0094-5765. doi: 10.1016/j.actaastro.2022.09.041.  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701

## GENAI USAGE DISCLOSURE

Text rephrasing and grammar checks were performed using GPT-5 and Grammarly. Code debugging was assisted by GPT-5. All outputs were reviewed and modified by the authors. All ideas and content were originally written and verified by the authors.

## A RELATED WORK

### A.1 ORBIT PREDICTION USING MACHINE LEARNING

Traditional orbit prediction processes are based on physics-based propagators, such as SGP4, which integrate orbital equations under simplifying assumptions. Although effective for short-term predictions, these methods degrade over time due to unmodeled environmental forces and imprecise initial conditions. ML-based approaches have emerged to correct these deficiencies by learning error patterns from historical data (Caldas & Soares, 2024a; Kazemi et al., 2024). One common strategy is hybrid error correction, in which a model learns the deviation of a physics-based prediction from ground truth. Peng and Bai applied SVR to correct SGP4 errors and demonstrated 20–30% reductions in prediction error (Peng & Bai, 2018). They later extended this framework to Gaussian processes and neural networks (Peng & Bai, 2019), and combined Gaussian process error fixes into a Kalman filter (Peng & Bai, 2021).

Temporal models such as RNNs and LSTMs have also been used to predict. Zhou et al. embedded an LSTM within an orbit determination process to maneuver satellites and demonstrated improved robustness in long-horizon predictions (Zhou et al., 2023). Latent force models treat unmodeled accelerations as Gaussian process priors, combining Newtonian mechanics with statistical inference (Hartikainen & Särkkä, 2010). More recently, neural operators and differentiable programming have been proposed to learn orbital dynamics directly from partial differential equations (Chen et al., 2024). Although ML-based approaches to orbit prediction have made progress, open challenges remain in generalizing models across orbital regimes and ensuring robustness with limited training data.

### A.2 LARGE TIME-SERIES MODEL

In time-series, large models have recently been actively proposed to learn general time representations through large-scale pre-training and apply them to various domains in a zero-shot manner.

**TimesFM** (Das et al., 2024) adopts a decoder-only Transformer architecture and is pre-trained on 100 billion time points. It is evaluated across heterogeneous benchmarks such as energy, finance, and weather without fine-tuning. **MOIRAI** (Woo et al., 2024) is a masked encoder-based universal time-series prediction transformer that addresses universal prediction challenges through multiple patch-size projection layers, Any-variate Attention, and a flexible mixture-distribution output layer. **TinyTimeMixers (TTM)** (Ekambaram et al., 2024) is an extremely lightweight model based on the MLP-based TSMixer architecture. It handles multivariate time-series through Adaptive Patching and Multi-Resolution Sampling. **MOMENT** (Goswami et al., 2024) builds a large-scale public time-series dataset called the “Time-Series Pile” and proposes a task-agnostic foundation model applicable to classification, prediction, and anomaly detection. **Timer-XL** (Liu et al., 2025) combines a time-series self-attention mechanism (“TimeAttention”) with a Kronecker-based mask to efficiently process long-context sequences. It supports univariate, multivariate, and 1D/2D inputs within a single architecture. **Time-MoE** (Shi et al., 2025) introduces a sparse Mixture-of-Experts (MoE) structure into the Transformer to enable token-level specialization. It is scalable up to 2.4 billion parameters and is trained on the Time-300B dataset (300 billion time points). **Time-LLM** (Jin et al., 2024) reprograms time-series as text prototypes using existing large language models (LLMs) and applies the Prompt-as-Prefix method without modifying the base architecture.

While all of these models demonstrate strong capabilities in zero-shot learning across various time-series tasks, their large model sizes make fine-tuning computationally demanding. Moreover, it remains challenging to achieve high performance on specific datasets consistently.

## B FULL RESULTS FOR INTERPOLATION METHOD SELECTION

Table 10: Statistical analysis results for validating the statistical significance of the interpolated dataset and mean values for different features. The table includes Friedman test statistics and  $p$ -values for MSE, MAE, mean difference, and mean standard deviation across the  $x$ ,  $y$ ,  $z$ ,  $vx$ ,  $vy$ , and  $vz$  features, Nemenyi post-hoc  $p$ -values for pairwise comparisons among SGP4, Skyfield, and Orekit methods, and the corresponding mean values for each method. Nemenyi test was not performed for features that were not statistically significant in the Friedman test. The best result is highlighted in **bold**.

Metric	Feat.	Friedman test		Nemenyi test ( $p$ -val.)			Mean Values		
		stat.	$p$ -val.	SGP4&Skyfield	SGP4&Orekit	Orekit&Skyfield	SGP4	Skyfield	Orekit
MSE	$x$	8448.9772	< 0.05	< 0.05	< 0.05	< 0.05	10645.732	9816.147	<b>9816.146</b>
	$y$	9174.1494	< 0.05	< 0.05	< 0.05	< 0.05	8133.012	7369.783	<b>7369.780</b>
	$z$	3947.891	< 0.05	< 0.05	< 0.05	0.9	10259.165	<b>10174.8805</b>	10174.8813
	$vx$	8365.3877	< 0.05	< 0.05	< 0.05	< 0.05	0.014363	<b>0.013367</b>	<b>0.013367</b>
	$vy$	9018.8088	< 0.05	< 0.05	< 0.05	< 0.05	0.010629	<b>0.009711</b>	<b>0.009711</b>
	$vz$	4071.345	< 0.05	< 0.05	< 0.05	0.9	0.013902	<b>0.013801</b>	<b>0.013801</b>
MAE	$x$	8497.8197	< 0.05	< 0.05	< 0.05	< 0.05	31.88368	13.587937	<b>13.587854</b>
	$y$	9230.8524	< 0.05	< 0.05	< 0.05	< 0.05	30.527394	13.181023	<b>13.180938</b>
	$z$	4552.0648	< 0.05	< 0.05	< 0.05	0.9	18.179287	13.355324	<b>13.355318</b>
	$vx$	8521.2458	< 0.05	< 0.05	< 0.05	< 0.05	0.035155	<b>0.015061</b>	<b>0.015061</b>
	$vy$	9140.794	< 0.05	< 0.05	< 0.05	< 0.05	0.033603	<b>0.014575</b>	<b>0.014575</b>
	$vz$	4654.464	< 0.05	< 0.05	< 0.05	0.880152	0.020182	<b>0.014809</b>	<b>0.014809</b>
Mean Diff.	$x$	396.8254	< 0.05	0.9	< 0.05	< 0.05	-0.038654	-0.022224	<b>-0.022223</b>
	$y$	30.4268	< 0.05	0.9	< 0.05	< 0.05	-0.002987	<b>-0.000447</b>	<b>-0.000447</b>
	$z$	4.7382	0.093	-	-	-	0.090636	<b>0.091607</b>	<b>0.091607</b>
	$vx$	2.9315	0.231	-	-	-	0.000001	<b>0.000000</b>	<b>0.000000</b>
	$vy$	0.3282	0.849	-	-	-	<b>-0.000005</b>	<b>-0.000005</b>	<b>-0.000005</b>
	$vz$	0.0038	0.998	-	-	-	<b>-0.000003</b>	<b>-0.000003</b>	<b>-0.000003</b>
Mean Std.	$x$	8448.9772	< 0.05	< 0.05	< 0.05	< 0.05	37.83216	18.310809	<b>18.310721</b>
	$y$	9174.1494	< 0.05	< 0.05	< 0.05	< 0.05	36.089137	17.615347	<b>17.615256</b>
	$z$	3947.891	< 0.05	< 0.05	< 0.05	0.9	22.715372	17.910431	<b>17.909425</b>
	$vx$	8365.3877	< 0.05	< 0.05	< 0.05	< 0.05	0.041777	<b>0.020314</b>	<b>0.020314</b>
	$vy$	9019.5578	< 0.05	< 0.05	< 0.05	< 0.05	0.039778	<b>0.019489</b>	<b>0.019489</b>
	$vz$	4071.345	< 0.05	< 0.05	< 0.05	0.9	0.02525	<b>0.019863</b>	<b>0.019863</b>

We present a detailed performance comparison of the three interpolation methods (SGP4, Skyfield, and Orekit) for the six features ( $x$ ,  $y$ ,  $z$ ,  $vx$ ,  $vy$ , and  $vz$ ), as shown in Table 10. For each satellite, we calculated the MSE, MAE, mean difference, and mean standard deviation between the ground truth and the interpolated results. These values were then averaged across all satellites to obtain **Mean Values** for each method. Statistical significance tests follow the same analysis methods described in Section 2.2.

First, in the case of error-based metrics (MSE, MAE), there was a statistically significant difference between the three methods in all features. In the Nemenyi test, SGP4 showed a significant difference with Skyfield and Orekit in all features, and there was no difference between Skyfield and Orekit except for the features  $z$  and  $vz$ . In the mean value, Orekit was the lowest for position features, and Skyfield and Orekit recorded the lowest value for velocity features. These results indicate that Skyfield and Orekit perform better than SGP4.

Second, the mean difference showed statistically significant differences only in the features  $x$  and  $y$ . In these features, Skyfield and Orekit outperform SGP4, and there was no significant difference in other features.

Third, the mean standard deviation significantly outperformed SGP4 in Skyfield and Orekit in all features. In particular, the standard deviation of the two techniques in velocity features is half that of SGP4, indicating that the prediction stability is significantly superior to that of SGP4.

In summary, Skyfield and Orekit overwhelmed SGP4 in all metrics. Although the performance gap between Skyfield and Orekit is insignificant, Orekit was slightly lower in MSE and MAE. Therefore, we selected Orekit as our final interpolation method. Building on this result, the final interpolated Starlink dataset comprises approximately 7.3 billion rows (7,295,420,002). Using this dataset, we created the training, validation, and test splits as described in Section 2.2 and used them to train and evaluate the large time-series prediction models.

## C EFFECTIVENESS OF TENSORIZED MULTI-HEAD ATTENTION

In our study, we adopted tensor decomposition for TMHA. However, as discussed in Section 4.4, we examine the effectiveness of TMHA by applying CPD instead of TTD. Accordingly, we describe CPD in Section C.1, and explain the TMHA method using CPD in Section C.2.

### C.1 CANONICAL POLYADIC DECOMPOSITION (CPD)

CPD is a method for decomposing a tensor into a sum of rank-1 tensors, each represented by the outer product of factor vectors, which are organized as columns in factor matrices. The formulation of the CPD for an N-way tensor is given in Equation 7:

$$\mathcal{X}(i_1, i_2, \dots, i_N) \approx \sum_{r=1}^R A^{(1)}(i_1, r) A^{(2)}(i_2, r) \dots A^{(N)}(i_N, r). \quad (7)$$

For instance, when CPD with rank  $R$  is applied to a 3-way tensor  $\mathcal{X} \in \mathbb{R}^{i_1 \times i_2 \times i_3}$ , obtains factor matrix  $A^{(1)} \in \mathbb{R}^{i_1 \times R}$ ,  $A^{(2)} \in \mathbb{R}^{i_2 \times R}$ , and  $A^{(3)} \in \mathbb{R}^{i_3 \times R}$ . Here,  $R$ , referred to as the CP rank, indicates how many rank-1 tensors are required to approximate the original tensor. In our work, CPD is applied to the weight matrices for the  $Q$ ,  $K$ , and  $V$  of MHA, denoted as  $W_Q \in \mathbb{R}^{D \times D}$ ,  $W_K \in \mathbb{R}^{D \times D}$ , and  $W_V \in \mathbb{R}^{D \times D}$ , respectively. Here,  $D$  denotes the dimension of each weight matrix. These weight matrices are stacked to form a 3-dimensional tensor,  $W \in \mathbb{R}^{3 \times D \times D}$ . We then apply Equation 7 to  $W$ , resulting in the following factor matrices:  $A^{(1)} \in \mathbb{R}^{3 \times R}$ ,  $A^{(2)} \in \mathbb{R}^{D \times R}$ , and  $A^{(3)} \in \mathbb{R}^{D \times R}$ .

### C.2 TENSORIZED MULTI-HEAD ATTENTION USING CPD

Similar to applying TTD to TMHA, we apply CPD to the conventional Attention weight matrices, yielding factor matrices  $A^{(1)}$ ,  $A^{(2)}$ , and  $A^{(3)}$ . Subsequently, we apply Equation 8 to the input  $Z$ , and then compute the final TA value using Equation 8.

$$\begin{aligned} \text{TA}(Z, A^{(1)}, A^{(2)}, A^{(3)}) &= \text{softmax} \left( \frac{Q'K'}{\sqrt{d_h}} \right) V', \quad s.t. \\ [Q', K', V']_{ndj} &= \sum_{a=1}^D Z_{na} W(j, a, d) \\ &\approx \sum_{a=1}^D Z_{na} \sum_{r=1}^R A^{(1)}(j, r) A^{(2)}(a, r) A^{(3)}(d, r) \end{aligned} \quad (8)$$

where  $a, d \in \{1, \dots, D\}$  are weight dimension index, and  $n \in \{1, \dots, N\}$  is the sequence index,  $j \in \{1, 2, 3\}$  selects  $Q', K', V'$ ,  $r \in \{1, \dots, R\}$  is the rank of CPD and  $d_h$  denotes attention key dimension per number of heads.

$$\text{TMHA}(Z, A^{(1)}, A^{(2)}, A^{(3)}) = \text{Concat}_{i=1}^{\text{head}}(\text{TA}_i(Z_i, A^{(1)}, A^{(2)}, A^{(3)})). \quad (9)$$

The process of extending TA with CPD to TMHA is similar to the method used for extending TA with TTD. We first reshape the high-dimensional tensor  $Z \in \mathbb{R}^{3 \times B \times C \times D}$  by splitting the second dimension into multiple heads, resulting in  $Z \in \mathbb{R}^{3 \times \frac{B}{h} \times h \times C \times D}$ . Then, we apply Equation 9 to obtain the final TMHA value. Here,  $\text{TA}_i$  denotes the  $i$ -th head's TA, as defined in Equation 8.

## D EXPLANATION OF DETAIL EXPERIMENTS SETUP

As described in Section 4.1, we fine-tuned large time-series models on 58,488,461 training rows and 15,012,604 validation rows from the Starlink dataset. Unless otherwise noted, models use the

Adam optimizer and MSE loss (TimesFM uses Quantile Loss). A CosineAnnealingLR scheduler with  $T_{\max} = 10$  and  $\eta_{\min} = 10^{-8}$  is applied to MOIRAI, Time-LLM, MOMENT, TTM, Timer, and Timer-XL. The remaining hyperparameters are:

- **DAF (Ours)**: learning rate  $1 \times 10^{-3}$ ; number of layers 2; number of heads 4; model dimension 128; hidden size 128; Tensor Train Rank [1, 3, 32, 1].
- **FredFormer**: learning rate  $1 \times 10^{-3}$ ; number of layers 2; number of heads 4; model dimension 128; patch length 16; hidden size 128.
- **GRU, LSTM**: learning rate  $1 \times 10^{-3}$ ; hidden size 128; number of layers 1.
- **DLinear**: learning rate  $1 \times 10^{-3}$ .
- **SparseTSF**: learning rate  $1 \times 10^{-3}$ ; period length 32, dimension of model 3; model type 'mlp'.
- **Time-MoE**: learning rate  $1 \times 10^{-4}$ ; hidden size 4,096; intermediate size 22,016; 32 transformer layers; 32 attention heads; 32 key-value heads; RoPE  $\theta = 10,000$ ; number of experts 1; experts per token 2; aux loss (router) 0.02; max position embeddings 32,768.
- **TimesFM**: learning rate  $1 \times 10^{-4}$ ; weight decay 0.01; batch size (train/inference) 200/20,000; 50 transformer layers; hidden size 1,024; intermediate size 2,816; 8 attention heads; 8 key-value heads; head dimension 128; patch length 128; quantiles [0.1, 0.5, 0.9]; RMSNorm  $\epsilon = 10^{-6}$ ; tolerance  $10^{-6}$ .
- **MOIRAI**: learning rate  $3 \times 10^{-4}$ ; batch size 256; dropout 0.1; 5 encoder layers; model dimension 512; 8 attention heads; feedforward dimension 2,048; node num 100; node list [23, 37, 40].
- **Time-LLM**: learning rate  $1 \times 10^{-5}$ ; batch size 64; dropout 0.2; model dimension 2; feedforward dimension 8; 4 attention heads; 2 layers; dimension 768; TS vocab size 1,000; patch number 54; head NF 3,456; top- $k$  5.
- **MOMENT**: learning rate  $1 \times 10^{-4}$ ; batch size 128; dropout 0.1; 5 encoder layers; model dimension 512; 8 attention heads; feedforward dimension 2,048; patch length/stride/padding 90; head NF 3,072.
- **TTM**: learning rate  $1 \times 10^{-4}$ ; batch size 128; dropout 0.2; 3 encoder and 2 decoder layers; model dimension 1,024; hidden dimension 64; patch size 64; number of patches 8.
- **Timer** and **Timer-XL**: learning rate  $5 \times 10^{-6}$ ; batch size 2,048; dropout 0.1; 8 encoder layers; model dimension 1,024; 8 attention heads; feedforward dimension 2,048.

## E DETAILED ZERO-SHOT RESULTS WITH STANDARD DEVIATIONS

Table 11 presents the MSE and MAE performance of the baselines and DAF for seven cross-domain datasets and three real-world datasets. We repeated the experiment three times and reported the results **mean  $\pm$  standard deviation**. We describe the experiments on individual satellites for each constellation in Appendix F.

DAF recorded the best performance on 9 MSE and 8 MAE indicators of 10 datasets. In particular, DAF showed the best performance in KOMPSAT-3A, ASTROCAST, CAPELLA, ICEYE, KINEIS, LEMUR, and SKYSAT datasets in both MSE and MAE, and showed overall good performance. For MSE, DAF recorded  $34.63 \pm 11.98\%$  in KOMPSAT-3A,  $7.54 \pm 13.36\%$  in KOMPSAT-5,  $32.59 \pm 1.49\%$  in ASTROCAST,  $59.81 \pm 0.93\%$  in CAPELLA,  $62.70 \pm 1.44\%$  in ICEYE,  $9.40 \pm 29.06\%$  in KINEIS,  $8.85 \pm 0.72\%$  in KUIPER,  $62.81 \pm 1.55\%$  in LEMUR, and  $44.25 \pm 1.63\%$  in SKYSAT for relative improvement. For MAE, DAF shows the improvement  $5.64 \pm 10.14\%$  in KOMPSAT-3,  $16.28 \pm 5.90\%$  in KOMPSAT-3A,  $11.97 \pm 2.01\%$  in ASTROCAST,  $39.08 \pm 0.54\%$  in CAPELLA,  $32.76 \pm 0.81\%$  in ICEYE, and  $10.22 \pm 15.57\%$  in KINEIS,  $32.20 \pm 0.88\%$  in LEMUR, and  $5.84 \pm 1.14\%$  in SKTSAT.

On the KOMPSAT-3, KOMPSAT-5, and KUIPER datasets, other models performed best in some indicators, but DAF also showed the second-best performance in those indicators. Overall, DAF recorded the highest performance in most datasets, and the performance gap is also numerically clear. The standard deviation also showed a very low value in most cases, confirming stable results in terms of prediction consistency.

## F ZERO-SHOT EVALUATION ON INDIVIDUAL SATELLITES

The datasets used in Table 11, excluding KOMPSAT dataset (3, 3A, and 5), are satellite constellations, where multiple satellites are grouped into a single network. Some of the constellations consist of dozens of individual satellites, while others do not. Accordingly, we conducted experiments on each satellite within these constellations and evaluated how many satellites our model performed best on.

Tables 12 to 21 present the experimental results for individual satellites excluding Starlink and KOMPSAT, as listed in Table 1. Among them, one satellite from each of the LEMUR and IC-EYE constellations was excluded because of an insufficient number of data samples, which were not even enough to construct a single input sequence. As a result, we conducted experiments on a total of 220 individual satellites. We repeated the experiment three times and reported the results as **mean  $\pm$  standard deviation**. For “Improvements,” we used a ‘-’ when our model did not achieve the best or second-best performance.

Our DAF model achieved the best performance on 142 satellites in terms of MSE and on 135 satellites in terms of MAE. In contrast, the DLinear model achieved the best performance on 67 satellites for MSE and 79 satellites for MAE. The SparseTSF model showed the best performance on 6 satellites for MSE and 5 satellites for MAE. Lastly, the GRU model achieved the best performance on 5 satellites in terms of MSE, and none in terms of MAE. These results indicate that our model outperformed DLinear, which was the second-best model in terms of the number of satellites where it performed best, by 52.82% more satellites for MSE and 41.48% more satellites for MAE.



1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079

Table 12: Zero-shot evaluation on individual satellites of three cross-domain datasets (ASTROCAST, CAPELLA, ICEYE) with mean and standard deviation in MSE and MAE. “Improvements” presents the relative performance of our model compared to the best baseline for each dataset and metric. The best result is highlighted in **bold**, and the second-best is highlighted with underline.

Dataset	ASTROCAST-F1		ASTROCAST-F2		ASTROCAST-F3		ASTROCAST-F4		ASTROCAST-F5		ASTROCAST-F6	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Time-MoE	2.241446±0.020478	1.933101±0.007028	2.198212±0.019937	1.818087±0.006610	2.182484±0.017097	1.848181±0.007107	2.204848±0.010170	1.848181±0.007107	2.204848±0.010170	1.848181±0.007107	2.204848±0.010170	1.848181±0.007107
Time-Timer	1.672082±0.002808	1.086631±0.001148	2.027767±0.001186	1.204326±0.000832	2.069629±0.000408	1.216346±0.000408	2.036096±0.001940	1.222539±0.001040	2.077396±0.001321	1.245700±0.000320	2.221036±0.001750	1.162630±0.000545
Time-XL	1.924443±0.01312849	1.089099±0.0041676	2.202663±0.027673	1.225345±0.002620	2.263395±0.002620	1.225345±0.002620	2.079138±0.008710	1.225345±0.002620	2.033744±0.001996	1.245700±0.000320	2.194647±0.025171	1.245700±0.000320
Time-LLM	1.055419±0.0160210	0.832568±0.0038114	1.061183±0.009254	0.834053±0.007356	1.062313±0.0116672	0.834053±0.007356	1.062313±0.0116672	0.834053±0.007356	0.960147±0.011859	0.832085±0.0061856	0.935907±0.168856	0.832085±0.0061856
Time-TTM	0.355219±0.0121316	0.457843±0.190561	2.057257±0.118620	0.839362±0.0132313	0.345791±0.0137437	0.839362±0.0132313	0.345791±0.0137437	0.839362±0.0132313	0.202919±0.0115794	0.841751±0.087854	0.841751±0.087854	0.297774±0.115550
TimeFSF	0.001325±0.0000118	0.027995±0.0001172	0.027995±0.0001172	0.027995±0.0001172	0.001385±0.0001030	0.027995±0.0001172	0.001385±0.0001030	0.027995±0.0001172	0.001346±0.0001017	0.027995±0.0001172	0.001346±0.0001017	0.027995±0.0001172
MOMENT	1.16201±0.0000514	0.171668±0.0004921	0.162168±0.0004921	0.162168±0.0004921	0.159620±0.0004008	0.162168±0.0004921	0.159620±0.0004008	0.162168±0.0004921	0.159488±0.0004001	0.159488±0.0004001	0.159488±0.0004001	0.159488±0.0004001
MOIRAI	0.024159±0.0010571	1.113308±0.0026198	0.016075±0.0006324	0.016075±0.0006324	0.016075±0.0006324	0.016075±0.0006324	0.016075±0.0006324	0.016075±0.0006324	0.016075±0.0006324	0.016075±0.0006324	0.016075±0.0006324	0.016075±0.0006324
GRU	0.002129±0.0005534	0.027855±0.0006129	0.002568±0.0002688	0.002568±0.0002688	0.002568±0.0002688	0.002568±0.0002688	0.002568±0.0002688	0.002568±0.0002688	0.002568±0.0002688	0.002568±0.0002688	0.002568±0.0002688	0.002568±0.0002688
LSTM	0.002929±0.0000877	0.040628±0.0007100	0.001669±0.0006062	0.001669±0.0006062	0.001669±0.0006062	0.001669±0.0006062	0.001669±0.0006062	0.001669±0.0006062	0.001669±0.0006062	0.001669±0.0006062	0.001669±0.0006062	0.001669±0.0006062
DL-neur	0.000518±0.0000006	0.014878±0.0001113	0.000223±0.0000001	0.000223±0.0000001	0.000223±0.0000001	0.000223±0.0000001	0.000223±0.0000001	0.000223±0.0000001	0.000223±0.0000001	0.000223±0.0000001	0.000223±0.0000001	0.000223±0.0000001
SparseFSF	0.000453±0.0000004	0.015690±0.0000081	0.000307±0.0000005	0.000307±0.0000005	0.000307±0.0000005	0.000307±0.0000005	0.000307±0.0000005	0.000307±0.0000005	0.000307±0.0000005	0.000307±0.0000005	0.000307±0.0000005	0.000307±0.0000005
FredFormer	0.036831±0.0002427	0.162168±0.0004921	0.035376±0.0001958	0.035376±0.0001958	0.035376±0.0001958	0.035376±0.0001958	0.035376±0.0001958	0.035376±0.0001958	0.035376±0.0001958	0.035376±0.0001958	0.035376±0.0001958	0.035376±0.0001958
DAF (Ours)	<b>0.000422±0.0000017</b>	<b>0.014594±0.000273</b>	<b>0.000227±0.0000005</b>	<b>0.010243±0.0004046</b>	<b>0.000212±0.0000007</b>	<b>0.009834±0.000512</b>	<b>0.000205±0.0000006</b>	<b>0.009944±0.0004046</b>	<b>0.000211±0.0000008</b>	<b>0.009944±0.0004046</b>	<b>0.000164±0.0000007</b>	<b>0.000164±0.0000007</b>
Improvements	6.84±4.38	1.91±2.58	-1.79±2.70	-19.89±5.14	-4.49±3.94	-14.98±6.68	-11.41±3.87	-24.04±5.56	-7.65±4.63	-21.56±5.79	33.33±3.12	16.02±6.64
Time-MoE	2.173325±0.011445	1.159749±0.003943	2.166314±0.011226	1.157171±0.005049	2.139732±0.018210	1.136947±0.008637	2.160046±0.015011	1.148438±0.005583	2.163388±0.015419	1.150390±0.004707	2.172768±0.018157	1.154200±0.006544
Time-Timer	2.227032±0.001974	1.246040±0.000312	2.223626±0.003286	1.245626±0.000611	2.251549±0.004137	1.321783±0.000826	2.419238±0.004368	1.296029±0.000826	2.338733±0.003627	1.274210±0.010024	2.310403±0.001245	1.266656±0.000192
Time-XL	1.162740±0.084262	0.932590±0.020114	1.188664±0.025780	0.932590±0.020114	1.137901±0.051332	0.923016±0.008011	1.144697±0.039559	0.873645±0.003404	1.254077±0.01859	0.862103±0.02014	2.00336±0.041625	1.23782±0.012834
Time-LLM	1.055804±0.08383	0.297790±0.113575	1.155766±0.088576	0.29775±0.113851	0.82656±0.044072	0.28024±0.071755	0.91863±0.049598	0.222497±0.078148	1.011288±0.022434	0.250213±0.090979	1.034295±0.084596	0.82293±0.048957
Time-TTM	0.001485±0.0000012	0.028154±0.0000899	0.001498±0.0000122	0.028214±0.000834	0.001896±0.000402	0.003131±0.000037	0.001698±0.000132	0.029823±0.000850	0.001757±0.000132	0.030308±0.0000251	0.001618±0.000155	0.26206±0.095282
TimeFSF	0.112683±0.0007078	1.595400±0.001965	0.112329±0.001250	1.59446±0.002014	0.156144±0.000838	0.156144±0.000838	0.110596±0.000181	0.155321±0.001625	0.117239±0.0001851	0.157529±0.020051	0.111112±0.0006689	0.156890±0.001128
MOMENT	0.012658±0.0004546	0.082308±0.016228	0.012656±0.0004502	0.082221±0.016049	0.010165±0.0002024	0.073469±0.0011252	0.010699±0.000332	0.075707±0.001259	0.011601±0.0003860	0.078543±0.0013639	0.011787±0.0003860	0.07929±0.0014125
MOIRAI	0.002567±0.0000306	0.039442±0.002078	0.002574±0.0000316	0.039442±0.002078	0.002574±0.0000316	0.039442±0.002078	0.002574±0.0000316	0.039442±0.002078	0.002574±0.0000316	0.039442±0.002078	0.002574±0.0000316	0.039442±0.002078
LSTM	0.001878±0.0001002	0.032777±0.010416	0.001843±0.000986	0.032487±0.010265	0.004198±0.0020290	0.0046818±0.015499	0.002443±0.0001280	0.036842±0.011307	0.002162±0.0001123	0.034674±0.010877	0.002056±0.000982	0.034770±0.009975
DL-neur	0.000225±0.0000001	0.009855±0.000049	0.000225±0.0000001	0.009855±0.000049	0.000305±0.0000002	0.011696±0.000046	0.000414±0.0000022	0.013497±0.000051	0.000456±0.0000022	0.013391±0.000056	0.000377±0.0000001	0.013470±0.000043
SparseFSF	0.000455±0.0000008	0.015992±0.000168	0.000440±0.0000007	0.015749±0.0000340	0.0009687±0.0000006	0.016642±0.000128	0.000687±0.0000008	0.019520±0.000132	0.000743±0.0000033	0.019520±0.000132	0.000663±0.0000008	0.019489±0.000139
FredFormer	0.034499±0.001759	0.151484±0.0033568	0.034470±0.001807	0.151570±0.003430	0.031546±0.001508	0.146756±0.003312	0.033222±0.001429	0.150831±0.001556	0.033914±0.001556	0.152399±0.002847	0.033987±0.001557	0.152597±0.002970
DAF (Ours)	<b>0.001159±0.0000009</b>	<b>0.008533±0.000659</b>	<b>0.000152±0.0000011</b>	<b>0.008467±0.000705</b>	<b>0.000152±0.0000011</b>	<b>0.008467±0.000705</b>	<b>0.000152±0.0000011</b>	<b>0.008467±0.000705</b>	<b>0.000152±0.0000011</b>	<b>0.008467±0.000705</b>	<b>0.000152±0.0000011</b>	<b>0.008467±0.000705</b>
Improvements	29.33±4.31	13.41±7.12	33.33±5.12	16.61±7.51	49.84±3.28	28.30±3.82	44.69±1.21	28.30±3.82	40.83±1.19	24.19±4.49	43.77±1.48	26.49±4.83
Time-MoE	2.25673±0.023297	1.182822±0.008516	2.256871±0.023732	1.182716±0.008277	2.257724±0.027339	1.182831±0.009033	2.258834±0.025273	1.182297±0.008399	2.126943±0.026795	1.121338±0.006886	2.089090±0.021212	1.199660±0.005162
Time-Timer	1.518578±0.002494	1.020033±0.001508	1.520604±0.003576	1.020806±0.001187	1.520663±0.004177	1.020447±0.001539	1.521514±0.003508	1.020953±0.001347	1.521514±0.003508	1.020953±0.001347	1.521514±0.003508	1.020953±0.001347
Time-XL	1.85999±0.132515	1.131265±0.042520	1.859865±0.134854	1.132171±0.042985	1.85988±0.135401	1.132171±0.042985	1.859162±0.134618	1.131928±0.042933	1.859162±0.134618	1.131928±0.042933	2.298659±0.026122	1.34813±0.009351
Time-LLM	1.039033±0.082565	0.868599±0.048336	1.002102±0.127512	0.832030±0.070337	0.999418±0.132103	0.829067±0.072654	1.001837±0.127505	0.82624±0.070487	1.001980±0.127705	0.82624±0.070487	0.990530±0.146404	0.805722±0.078207
Time-TTM	0.432783±0.259315	0.498052±0.209682	0.431410±0.258719	0.497076±0.209727	0.433109±0.259644	0.497076±0.209947	0.431482±0.258786	0.496928±0.209691	0.109979±0.062254	0.250135±0.093606	0.067734±0.033312	0.189266±0.065694
TimeFSF	0.001285±0.0000117	0.026315±0.001152	0.001279±0.0000111	0.026238±0.001080	0.001308±0.0000119	0.026466±0.001148	0.001291±0.0000117	0.026330±0.001101	0.001625±0.0000100	0.030560±0.000763	0.001828±0.0000048	0.001828±0.0000048
MOMENT	0.117157±0.0001961	1.171941±0.002706	0.117234±0.0000712	1.171574±0.002174	0.117803±0.0000286	1.172633±0.001148	0.117557±0.0001030	1.171975±0.000796	0.117557±0.0001030	1.166465±0.000796	0.109311±0.001412	0.165696±0.001768
MOIRAI	0.028261±0.0012787	1.202882±0.031102	0.028194±0.0012583	1.20119±0.030747	0.028306±0.0012561	1.20421±0.030732	0.028137±0.0012656	1.20167±0.030934	0.028137±0.0012656	1.20167±0.030934	0.028137±0.0012656	1.20167±0.030934
GRU	0.001277±0.000933	0.026863±0.009836	0.001295±0.000943	0.026946±0.009885	0.001313±0.000951	0.027065±0.009827	0.001323±0.000944	0.027176±0.009782	0.001323±0.000944	0.027176±0.009782	0.001323±0.000944	0.027176±0.009782
LSTM	0.003591±0.000877	0.045087±0.006886	0.003644±0.000861	0.045250±0.006823	0.003691±0.000884	0.045492±0.006899	0.003692±0.000880	0.045580±0.006876	0.001149±0.000477	0.023554±0.005442	0.000382±0.0000998	0.014517±0.002352
DL-neur	0.000857±0.0000014	0.020617±0.000186	0.000873±0.0000013	0.020670±0.000179	0.000992±0.0000014	0.020926±0.000192	0.000992±0.000001					

1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133

Table 13: Zero-shot evaluation on individual satellites of ICEYE dataset with mean and standard deviation in MSE and MAE. “Improvements” presents the relative performance of our model compared to the best baseline for each dataset and metric. A dash (‘-’) indicates that our model fails to achieve either the best or second-best performance. The best result is highlighted in **bold**, and the second-best is highlighted with underline.

Dataset	ICEYE-3		ICEYE-4		ICEYE-5		ICEYE-6		ICEYE-7		ICEYE-8	
	MSE	MAE										
Time-MeE	2.2652±0.024884	1.183607±0.009877	2.26688±0.018425	1.201742±0.006342	2.313715±0.021415	1.229923±0.008306	2.129433±0.013724	1.141823±0.004832	2.168723±0.017825	1.167562±0.006619	2.174873±0.014950	1.163928±0.005529
Time-XL	1.468243±0.003550	0.99385±0.001367	1.496195±0.003360	1.025114±0.001471	1.020000±0.002589	0.844616±0.002606	2.64650±0.003107	2.320858±0.001519	2.290703±0.001034	2.279073±0.001034	2.299073±0.001034	2.266808±0.000210
Time-LLM	1.848283±0.0154663	1.126651±0.004940	1.852675±0.0155699	1.067506±0.002589	1.675601±0.002589	0.919191±0.0069854	0.965259±0.001928	0.955178±0.0010250	2.205982±0.0052687	1.256599±0.016548	2.213715±0.021415	2.5061±0.000708
Time-TTM	0.54057±0.0236245	0.568295±0.242469	0.435884±0.003766	0.443338±0.002161	0.435386±0.002161	0.499931±0.011879	0.452578±0.011327	0.494733±0.011736	0.656160±0.015978	0.82754±0.015282	0.962754±0.015282	0.806595±0.005549
TimeFOMer	0.001365±0.000121	0.026930±0.001167	0.001394±0.000156	0.041131±0.000815	0.001943±0.000156	0.01187±0.000044	0.01647±0.000113	0.01647±0.000113	0.03024±0.000795	0.001663±0.000150	0.001663±0.000150	0.202985±0.000177
MOMENT	0.118179±0.001659	0.173059±0.002869	0.118205±0.000971	0.172522±0.001826	0.121209±0.002081	0.196532±0.002891	0.112057±0.000744	0.157629±0.000642	0.107730±0.000190	0.112988±0.000837	0.112988±0.000837	0.50922±0.000140
MOIRAI	0.037351±0.016831	0.146359±0.036378	0.023360±0.010020	0.109003±0.026968	0.029412±0.013152	0.122656±0.013125	0.029292±0.012425	0.121604±0.003782	0.048290±0.020781	0.159471±0.039918	0.09475±0.020242	0.071659±0.009286
GRU	0.001207±0.000557	0.202639±0.06543	0.001680±0.000324	0.030077±0.003607	0.003180±0.000145	0.035935±0.002965	0.010233±0.000283	0.076580±0.009347	0.003073±0.0009347	0.003073±0.0009347	0.003073±0.0009347	0.042867±0.009286
LSTM	0.003298±0.0001269	0.041736±0.008834	0.002560±0.000398	0.036705±0.011253	0.006822±0.0002164	0.060077±0.011577	0.008330±0.001188	0.038330±0.001046	0.001720±0.000656	0.031558±0.007401	0.002130±0.001117	0.034852±0.010198
DL-neur	0.001218±0.000018	0.024126±0.000210	0.001040±0.000015	0.022581±0.000194	0.005731±0.000099	0.055127±0.000522	0.000416±0.000002	0.012563±0.000048	0.000416±0.000002	0.013693±0.000063	0.000426±0.000002	0.013662±0.000054
SparseTSF	0.001656±0.000010	0.031586±0.000089	0.029007±0.000092	0.030080±0.000092	0.009299±0.000031	0.079111±0.000098	0.000514±0.000004	0.016722±0.000072	0.000754±0.000008	0.021138±0.000132	0.000692±0.000008	0.01986±0.000139
FredFormer	0.036469±0.002701	0.158302±0.005301	0.036515±0.002594	0.166880±0.005432	0.033888±0.000181	0.155578±0.003716	0.031233±0.000151	0.147829±0.0002356	0.033888±0.000150	0.155070±0.002991	0.034201±0.001592	0.155240±0.003111
DAF (Ours)	<b>0.000696±0.00047</b>	<b>0.018802±0.00046</b>	<b>0.000592±0.00038</b>	<b>0.017919±0.00046</b>	<b>0.001966±0.000065</b>	<b>0.032772±0.00040</b>	<b>0.000337±0.000008</b>	<b>0.009303±0.000033</b>	<b>0.000247±0.000004</b>	<b>0.010379±0.000636</b>	<b>0.000255±0.000003</b>	<b>0.010555±0.000589</b>
Improvements	42.86±4.70	22.07±2.33	43.08±4.47	20.65±2.48	37.59±4.70	17.94±2.70	34.86±2.24	25.95±2.77	40.62±1.25	24.20±0.96	63.15±0.86	46.86±3.34
Dataset	ICEYE-14											
Time-MeE	2.126243±0.015284	1.132395±0.005494	2.205098±0.015793	1.173248±0.005623	2.20824±0.015925	1.177897±0.005214	2.06694±0.015141	1.134657±0.004451	2.036338±0.011608	1.128640±0.009787	2.152797±0.014122	1.159395±0.004804
Time-XL	2.676910±0.005728	1.360002±0.001083	2.019034±0.001229	1.191371±0.000662	1.948373±0.000149	1.171870±0.000436	2.97629±0.001063	1.461619±0.002599	3.109202±0.009104	1.493937±0.001718	2.508564±0.004428	1.336683±0.000960
Time-LLM	2.423434±0.0033092	1.296297±0.008269	2.097258±0.045925	2.116619±0.014830	2.03276±0.009806	1.199964±0.030428	2.58086±0.005999	1.363824±0.01646	2.671538±0.051646	1.388339±0.012297	2.325150±0.02668	1.288739±0.001524
Time-TTM	0.959319±0.0175389	0.787199±0.092981	0.998758±0.124095	0.814997±0.068804	1.010388±0.116386	0.843425±0.064523	0.923081±0.167826	0.802383±0.091275	0.957362±0.151667	0.805723±0.082211	0.963546±0.169614	0.801624±0.091555
TimeFOMer	0.105886±0.058200	0.246776±0.089904	0.136679±0.076549	0.279080±0.104376	0.05264±0.026826	0.164378±0.052971	0.254081±0.134293	0.365860±0.146014	0.240194±0.140052	0.374666±0.104304	0.190144±0.070849	0.190144±0.070849
MOMENT	0.001870±0.000145	0.030933±0.000745	0.001590±0.000143	0.029164±0.000195	0.001524±0.000118	0.028850±0.001127	0.00364±0.0000950	0.042467±0.000886	0.001774±0.000141	0.030733±0.000886	0.011765±0.000130	0.030630±0.000783
MOIRAI	0.109623±0.000849	0.151564±0.000785	0.113177±0.000528	0.163914±0.002021	0.113776±0.000494	0.165994±0.001348	0.117103±0.000559	0.174959±0.000548	0.139380±0.000238	0.221762±0.009370	0.111055±0.001176	0.156488±0.000951
GRU	0.006833±0.000778	0.080107±0.014080	0.001207±0.0004105	0.080775±0.014945	0.008722±0.002158	0.068514±0.009021	0.020389±0.0034746	0.109979±0.0073178	0.0187181±0.007181	0.097562±0.020291	0.011115±0.002648	0.078028±0.009978
LSTM	0.004365±0.002054	0.049669±0.013508	0.001890±0.000699	0.03016±0.000719	0.00276±0.000119	0.031965±0.006703	0.00828±0.002330	0.072487±0.010246	0.003035±0.001630	0.041287±0.012921	0.002972±0.001591	0.040224±0.012652
DL-neur	0.000426±0.000002	0.013365±0.000014	0.000323±0.000001	0.010615±0.000623	<b>0.00056±0.000052</b>	<b>0.010774±0.000658</b>	0.000747±0.000039	0.016911±0.000127	0.000440±0.000002	0.04386±0.000051	0.000406±0.000001	0.013329±0.000043
SparseTSF	0.000692±0.000003	0.01344E+0.000050	0.000508±0.000050	0.015480±0.000112	0.002593±0.000080	0.014517±0.000080	0.029341±0.000020	0.037005±0.000174	0.000728±0.000000	0.029435±0.000114	0.000632±0.000006	0.018837±0.000104
FredFormer	0.030772±0.001195	0.148236±0.002470	0.035739±0.002026	0.158615±0.003889	0.056144±0.002102	0.159903±0.004151	0.02593±0.0001715	0.136004±0.004317	0.032402±0.001241	0.151735±0.002520	0.032782±0.001389	0.152736±0.002818
DAF (Ours)	<b>0.000277±0.000007</b>	<b>0.010592±0.000223</b>	<b>0.000300±0.000003</b>	<b>0.012062±0.000451</b>	<b>0.000156±0.000008</b>	<b>0.012859±0.000206</b>	<b>0.000365±0.000010</b>	<b>0.013512±0.000289</b>	<b>0.000239±0.000002</b>	<b>0.009969±0.000357</b>	<b>0.000248±0.000004</b>	<b>0.010160±0.000446</b>
Improvements	20.78±1.95	20.78±1.94	4.95±1.22	18.63±4.81	0.90±2.77	19.35±2.35	10.10±2.31	45.68±0.70	30.70±2.73	38.92±1.14	23.78±3.59	
Dataset	ICEYE-20											
Time-MeE	2.170412±0.016723	1.166078±0.003291	2.210384±0.016392	1.182021±0.005372	2.190308±0.019450	1.173149±0.005498	2.215083±0.014250	1.182011±0.008517	2.206868±0.012731	1.178902±0.008797	2.202963±0.012435	1.177810±0.003705
Time-XL	2.453704±0.003156	1.324353±0.000780	1.934075±0.001768	1.171024±0.000886	2.167147±0.002116	1.240462±0.000731	1.956754±0.001353	1.176688±0.000717	2.020425±0.006718	1.196705±0.001096	2.075047±0.002034	1.215528±0.000888
Time-LLM	2.290651±0.007262	1.281496±0.002828	2.029638±0.097147	2.024688±0.030512	2.129474±0.068646	1.232519±0.021455	2.05438±0.078293	1.207477±0.024325	2.077919±0.062735	1.215410±0.020954	2.105747±0.062607	1.224177±0.019445
Time-TTM	0.974636±0.162804	0.807558±0.087680	1.011795±0.115375	0.834192±0.064309	0.986772±0.140097	0.8212357±0.076937	0.823337±0.0665286	1.003226±0.125856	1.003226±0.125856	0.819712±0.069228	1.009967±0.131074	0.818705±0.071981
TimeFOMer	0.145555±0.077044	0.286606±0.100661	0.073012±0.038346	0.209002±0.067661	0.091699±0.049671	0.222598±0.078262	0.243399±0.142444	0.378398±0.153171	0.157237±0.089320	0.302201±0.16104	0.250887±0.146642	0.38355±0.154282
MOMENT	0.001472±0.000123	0.028442±0.001220	0.001449±0.000114	0.028168±0.000931	0.001375±0.000142	0.027388±0.001187	0.001634±0.000145	0.029676±0.001128	0.001468±0.000133	0.028093±0.001006	0.001623±0.000169	0.029252±0.0001348
MOIRAI	0.11058±0.001513	0.157433±0.001913	0.114777±0.001529	0.166869±0.002539	0.112366±0.001691	0.16362±0.002581	0.113490±0.000992	0.165061±0.001727	0.113591±0.000560	0.164672±0.001628	0.113219±0.001324	0.163492±0.002162
GRU	0.001983±0.0000765	0.034591±0.0005921	0.003343±0.0001376	0.046699±0.008538	0.001551±0.0002385	0.03822±0.0027401	0.001870±0.000353	0.033871±0.0002837	0.002040±0.000330	0.035442±0.003528	0.001284±0.000385	0.027573±0.0006035
LSTM	0.002026±0.000734	0.033572±0.006899	0.001631±0.000551	0.039427±0.006344	0.001834±0.0000727	0.032461±0.007777	0.001907±0.0000775	0.032993±0.008340	0.001788±0.000036	0.031915±0.009068	0.002801±0.000669	0.039058±0.0006572
DL-neur	<b>0.000316±0.000001</b>	<b>0.010160±0.000053</b>	0.000236±0.000001	0.009713±0.000042	0.000177±0.000004	0.007621±0.000038	0.000337±0.000002	<b>0.010810±0.000058</b>	0.000209±0.000001	<b>0.008496±0.000033</b>	0.000665±0.000003	<b>0.015417±0.000090</b>
SparseTSF	0.000480±0.000003	0.013956±0.000081	0.000400±0.000006	0.014882±0.000157	0.000288±0.000004	0.010238±0.000152	0.000533±0.000005	0.015986±0.000109	0.000400±0.000006	0.013832±0.000167	0.000686±0.000003	0.016985±0.000075
FredFormer	0.036771±0.002151	0.160425±0.004306	0.034838±0.001823	0.157240±0.003383	0.035976±0.002062	0.159411±0.004013	0.035808±0.001994	0.158851±0.0003861	0.035367±0.001922	0.158064±0.003719	0.036929±0.002	



Table 15: Zero-shot evaluation on individual satellites of two cross-domain datasets (KINEIS and KUIPER) with mean and standard deviation in MSE and MAE. “Improvements” presents the relative performance of our model compared to the best baseline for each dataset and metric. A dash (‘-’) indicates that our model fails to achieve either the best or second-best performance. The best result is highlighted with **bold**, and the second-best is highlighted with underline.

Dataset	Metric	KINEIS-10			KINEIS-11			KINEIS-12			KINEIS-13			KINEIS-14			KINEIS-15		
		MSE	MAE	MAE	MSE	MAE	MAE	MSE	MAE	MAE	MSE	MAE	MAE	MSE	MAE	MAE	MSE	MAE	MAE
KINEIS-10	Time-MeE	2.3827±0.049719	1.22716±0.001694	1.252228±0.019322	2.398554±0.046914	1.248923±0.015333	2.393893±0.045102	1.248783±0.015797	2.400898±0.050374	1.249923±0.017126	2.384278±0.044790	1.244286±0.003813							
	Time-T	0.520499±0.006960	0.576533±0.004400	0.575882±0.007192	0.511439±0.007160	0.578434±0.007428	0.518093±0.007476	0.582329±0.004567	0.515099±0.007476	0.580355±0.007429	0.576564±0.007134	0.617500±0.004008							
	Time-XL	1.432826±0.231832	0.983685±0.0082454	1.465110±0.220066	1.466242±0.219277	1.009243±0.007844	1.469882±0.218521	1.01020±0.007167	1.469781±0.219124	1.011315±0.007804	1.506608±0.209796	1.021750±0.003738							
	Time-LLM	1.98519±0.116661	0.885791±0.029842	2.044770±0.116720	1.802216±0.115830	0.898409±0.034877	2.000064±0.119653	0.893505±0.031382	1.987765±0.119599	0.897590±0.029769	1.87704±0.103212	0.893249±0.025304							
	Time-F	0.987953±0.587359	0.757562±0.031276	1.035511±0.611662	0.786205±0.032163	0.862526±0.034085	1.031540±0.607708	0.783699±0.320058	1.028940±0.609274	0.784154±0.320955	0.985259±0.587134	0.765550±0.37056							
	Time-MoM	0.024635±0.003341	0.128910±0.002865	0.027792±0.001760	0.133479±0.003326	0.123893±0.003343	0.026768±0.001289	0.131633±0.003485	0.02739±0.001357	0.132314±0.003485	0.021318±0.001081	0.163886±0.006794							
	Time-DAF	0.086991±0.003291	0.121068±0.0048709	0.042195±0.001760	0.143195±0.001760	0.122763±0.003326	0.090045±0.002678	0.142466±0.002880	0.090045±0.002678	0.142466±0.002880	0.082757±0.004446	0.129108±0.006592							
	Time-GRU	0.01480±0.0004708	0.092931±0.0016689	0.011902±0.003285	0.011868±0.002077	0.081311±0.002953	0.009197±0.002681	0.081761±0.004916	0.011942±0.002681	0.081761±0.004916	0.028739±0.007825	0.208473±0.007825							
	Time-LSTM	0.01780±0.0001611	0.099683±0.0050560	0.016900±0.002845	0.016859±0.002872	0.098673±0.007812	0.011698±0.0002818	0.098673±0.007812	0.016967±0.002879	0.098673±0.007812	0.014809±0.002648	0.092146±0.007883							
	Time-DL	0.019096±0.000389	0.101723±0.001039	0.020158±0.000412	0.105422±0.001065	0.105131±0.001065	0.019955±0.000403	0.105131±0.001065	0.020074±0.000408	0.10521±0.001073	0.016781±0.000340	0.096280±0.000974							
Time-SparseSF	0.036365±0.000123	0.156093±0.000177	0.037809±0.000127	0.161556±0.000182	0.160557±0.000182	0.036454±0.000123	0.158611±0.000181	0.037166±0.000120	0.159948±0.000182	0.031240±0.000103	0.146646±0.000161								
Time-FredFormer	0.030666±0.0001180	0.142679±0.0003541	0.030673±0.0001393	0.144370±0.002998	0.145063±0.0001210	0.030903±0.000120	0.145063±0.0001210	0.030768±0.000120	0.144749±0.0003182	0.030484±0.000126	0.146314±0.0003240								
Time-DAF (Ours)	<b>0.009934±0.000688</b>	<b>0.069225±0.003645</b>	<b>0.010652±0.000720</b>	<b>0.072300±0.003825</b>	<b>0.072251±0.003633</b>	<b>0.010569±0.000709</b>	<b>0.072251±0.003633</b>	<b>0.010624±0.000709</b>	<b>0.072348±0.003741</b>	<b>0.08162±0.000326</b>	<b>0.062906±0.002111</b>								
Improvements	13.41±41.50	12.65±23.00	10.4±26.52	11.08±15.13	11.47±14.64	11.62±25.91	11.47±14.64	11.04±25.91	11.12±14.82	20.10±22.76	16.36±13.37								
KINEIS-11	Time-MeE	2.387730±0.052483	1.238183±0.018324	2.379878±0.050885	1.255464±0.018133	2.378396±0.048325	1.255229±0.017597	2.382850±0.046198	1.256892±0.018303	2.373218±0.047256	1.253062±0.016951	2.405618±0.051482							
	Time-T	0.531787±0.007579	0.596496±0.004286	0.545944±0.006635	0.605053±0.004262	0.568936±0.006435	0.618873±0.003690	0.539971±0.006799	0.610100±0.006394	0.529290±0.006018	0.589108±0.004007	0.529290±0.006018							
	Time-XL	1.445647±0.2329502	0.911559±0.0086811	1.451451±0.236548	1.013682±0.0085757	1.462249±0.235503	1.017698±0.085303	1.464332±0.235533	1.011914±0.085431	1.470487±0.232657	1.020677±0.085214	1.455367±0.222246							
	Time-LLM	1.96834±0.115026	0.905283±0.029973	1.995276±0.112909	0.904786±0.029241	1.927142±0.105655	0.903653±0.026982	1.97824±0.114214	0.90523±0.030017	1.878183±0.100653	0.902564±0.025325	1.202747±0.113533							
	Time-F	0.960597±0.573263	0.764356±0.0317300	0.949785±0.567021	0.759748±0.0315918	0.935472±0.558760	0.754067±0.031746	0.954550±0.569180	0.762076±0.031598	0.927351±0.555230	0.750296±0.031364	0.78446±0.320897							
	Time-MoM	0.024274±0.001071	0.128905±0.002998	0.024626±0.000923	0.126081±0.002714	0.025477±0.000882	0.121984±0.002620	0.025477±0.000882	0.121984±0.002620	0.121984±0.002620	0.026050±0.001492	0.121984±0.002620							
	Time-DAF	0.085653±0.003245	0.121984±0.0048856	0.084417±0.0032095	0.125944±0.003216	0.137945±0.003216	0.082331±0.003216	0.141099±0.003216	0.082331±0.003216	0.137945±0.003216	0.082331±0.003216	0.142105±0.003216							
	Time-GRU	0.009953±0.000385	0.073592±0.0013791	0.009940±0.0003774	0.072600±0.0013802	0.009148±0.0003684	0.070841±0.0013856	0.009533±0.000362	0.070841±0.0013856	0.009148±0.0003684	0.070841±0.0013856	0.081721±0.0013856							
	Time-LSTM	0.017312±0.0001153	0.099052±0.003711	0.016958±0.0001230	0.098082±0.003727	0.016538±0.0001121	0.099632±0.003727	0.01729±0.0001339	0.099632±0.003727	0.016538±0.0001121	0.099632±0.003727	0.01729±0.0001339							
	Time-DL	0.018501±0.000372	0.102200±0.001034	0.017924±0.0001156	0.100625±0.001006	0.016868±0.0003338	0.097638±0.000984	0.01828±0.0001365	0.01513±0.001027	0.096258±0.000960	0.019496±0.000398	0.10591±0.001057							
Time-SparseSF	0.030398±0.0001108	0.145157±0.0003134	0.030398±0.0001108	0.145157±0.0003134	0.030398±0.0001108	0.145157±0.0003134	0.030398±0.0001108	0.145157±0.0003134	0.030398±0.0001108	0.030398±0.0001108	0.145157±0.0003134								
Time-FredFormer	<b>0.009594±0.000618</b>	<b>0.069465±0.0003598</b>	<b>0.009594±0.000618</b>																
Time-DAF (Ours)	3.61±43.45	6.44±22.09	5.58±42.32	7.64±21.67	7.90±21.42	8.99±21.25	6.08±21.82	6.08±21.82	6.08±21.82	6.08±21.82	6.08±21.82								
Improvements	13.39±22.09	13.39±22.09	13.39±22.09	13.39±22.09	13.39±22.09	13.39±22.09	13.39±22.09	13.39±22.09	13.39±22.09	13.39±22.09	13.39±22.09								
KINEIS-12	Time-MeE	2.400222±0.046914	1.252228±0.019322	2.398554±0.046914	1.248923±0.015333	2.393893±0.045102	1.248783±0.015797	2.400898±0.050374	1.249923±0.017126	2.384278±0.044790	1.244286±0.003813								
	Time-T	0.520499±0.006960	0.576533±0.004400	0.575882±0.007192	0.511439±0.007160	0.578434±0.007428	0.518093±0.007476	0.582329±0.004567	0.515099±0.007476	0.580355±0.007429	0.576564±0.007134								
	Time-XL	1.432826±0.231832	0.983685±0.0082454	1.465110±0.220066	1.466242±0.219277	1.009243±0.007844	1.469882±0.218521	1.01020±0.007167	1.469781±0.219124	1.011315±0.007804	1.506608±0.209796								
	Time-LLM	1.98519±0.116661	0.885791±0.029842	2.044770±0.116720	1.802216±0.115830	0.898409±0.034877	2.000064±0.119653	0.893505±0.031382	1.987765±0.119599	0.897590±0.029769	1.87704±0.103212								
	Time-F	0.987953±0.587359	0.757562±0.031276	1.035511±0.611662	0.786205±0.032163	0.862526±0.034085	1.031540±0.607708	0.783699±0.320058	1.028940±0.609274	0.784154±0.320955	0.985259±0.587134								
	Time-MoM	0.024635±0.003341	0.128910±0.002865	0.027792±0.001760	0.133479±0.003326	0.123893±0.003343	0.026768±0.001289	0.131633±0.003485	0.02739±0.001357	0.132314±0.003485	0.021318±0.001081								
	Time-DAF	0.086991±0.003291	0.121068±0.0048709	0.042195±0.001760	0.143195±0.001760	0.122763±0.003326	0.090045±0.002678	0.142466±0.002880	0.090045±0.002678	0.142466±0.002880	0.082757±0.004446								
	Time-GRU	0.01480±0.0004708	0.092931±0.0016689	0.011902±0.003285	0.011868±0.002077	0.081311±0.002953	0.009197±0.002681	0.081761±0.004916	0.011942±0.002681	0.081761±0.004916	0.028739±0.007825								
	Time-LSTM	0.01780±0.0001611	0.099683±0.0050560	0.016900±0.002845	0.016859±0.002872	0.098673±0.007812	0.011698±0.0002818	0.098673±0.007812	0.016967±0.002879	0.098673±0.007812	0.014809±0.002648								
	Time-DL	0.019096±0.000389	0.101723±0.001039	0.020158±0.000412	0.105422±0.001065	0.105131±0.001065	0.019955±0.000403	0.105131±0.001065	0.020074±0.000408	0.10521±0.001073	0.016781±0.000340								
Time-SparseSF	0.036365±0.000123	0.156093±0.000177	0.037809±0.000127	0.161556±0.000182	0.160557±0.000182	0.036454±0.000123	0.158611±0.000181	0.037166±0.000120	0.159948±0.000182	0.031240±0.000103									
Time-FredFormer	0.030666±0.0001180	0.142679±0.0003541	0.030673±0.0001393	0.144370±0.002998	0.145063±0.0001210	0.030903±0.000120	0.145063±0.0001210	0.030768±0.000120	0.144749±0.0003182	0.030484±0.000126									
Time-DAF (Ours)	<b>0.009934±0.000688</b>	<b>0.069225±0.003645</b>	<b>0.010652±0.000720</b>	<b>0.072300±0.003825</b>	<b>0.072251±0.003633</b>	<b>0.010569±0.000709</b>	<b>0.072251±0.003633</b>	<b>0.010624±0.000709</b>	<b>0.072348±0.003741</b>	<b>0.08162±0.000326</b>									
Improvements	13.41±41.50	12.65±23.00	10.4±26.52	11.08±15.13	11.47±14.64	11.62±25.91	11.47±14.64	11.04±25.91	11.12±14.82	20.10±22.76									
KINEIS-13	Time-MeE	2.382730±0.052483	1.238183±0.018324	2.379878±0.050885	1.255464±0.018133	2.378396±0.048325	1.255229±0.017597	2.382850±0.046198	1.256892±0.018303	2.373218±0.047256	1.253062±0.016951	2.405618±0.051482							
	Time-T	0.531787±0.007579	0.596496±0.004286	0.545944±0.006635	0.605053±0.004262	0.568936±0.006435	0.618873±0.003690	0.539971±0.006799	0.610100±0.006394	0.529290±0.006018	0.589108±0.004007	0.529290±0.006018							
	Time-XL	1.445647±0.2329502	0.911559±0.0086811	1.451451±0.236548	1.013682±0.0085757	1.462249±0.235503	1.017698±0.085303	1.464332±0.235533	1.011914±0.085431	1.470487±0.232657	1.020677±0.085214	1.455367±0.222246							
	Time-LLM	1.96834±0.115026	0.905283±0.029973	1.995276±0.112909	0.904786±0.029241	1.9271													

1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295

Table 16: Zero-shot evaluation on individual satellites of KUIPER dataset with mean and standard deviation in MSE and MAE. “Improvements” presents the relative performance of our model compared to the best baseline for each dataset and metric. A dash (‘-’) indicates that our model fails to achieve either the best or second-best performance. The best result is highlighted in **bold**, and the second-best is highlighted with underline.

Dataset	Metric	KUIPER-9			KUIPER-10			KUIPER-11			KUIPER-12			KUIPER-13			KUIPER-14		
		MSE	MAE																
Improvements	Time-MoE	2.245668±0.021669	1.262751±0.008367	1.249779±0.002442	2.209790±0.003894	1.243839±0.003894	2.209790±0.003894	1.243839±0.003894	2.209790±0.003894	1.243839±0.003894	2.209790±0.003894	1.243839±0.003894	2.209790±0.003894	1.243839±0.003894	2.209790±0.003894	1.243839±0.003894	2.209790±0.003894	1.243839±0.003894	
	Timer	1.504165±0.004615	1.080581±0.001926	1.172980±0.001014	1.875757±0.002270	1.213867±0.001231	1.881507±0.000951	1.861819±0.001603	2.007355±0.009625	1.249778±0.000574	1.861819±0.001603	2.007355±0.009625	1.249778±0.000574	1.861819±0.001603	2.007355±0.009625	1.249778±0.000574	1.861819±0.001603	2.007355±0.009625	
	Time-XL	1.855342±0.142344	1.083663±0.048576	1.242575±0.108473	2.267977±0.002952	1.283744±0.002952	2.007355±0.009625	1.249778±0.000574	2.267977±0.002952	1.283744±0.002952	2.007355±0.009625	1.249778±0.000574	2.267977±0.002952	1.283744±0.002952	2.007355±0.009625	1.249778±0.000574	2.267977±0.002952	1.283744±0.002952	
	Time-LLM	1.044472±0.003315	0.833323±0.041063	0.869022±0.058308	1.004947±0.110235	0.864196±0.064778	1.004947±0.110235	0.864196±0.064778	1.004947±0.110235	0.864196±0.064778	1.004947±0.110235	0.864196±0.064778	1.004947±0.110235	0.864196±0.064778	1.004947±0.110235	0.864196±0.064778	1.004947±0.110235	0.864196±0.064778	
	TTM	0.416009±0.250759	0.515545±0.219370	0.497884±0.178557	0.298784±0.148044	0.400749±0.165706	0.400749±0.165706	0.400749±0.165706	0.298784±0.148044	0.400749±0.165706	0.400749±0.165706	0.298784±0.148044	0.400749±0.165706	0.400749±0.165706	0.298784±0.148044	0.400749±0.165706	0.400749±0.165706	0.298784±0.148044	
	TimesFM	0.001410±0.000158	0.182418±0.002156	0.028135±0.001529	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	
	MOMENT	0.115508±0.000152	0.182418±0.002156	0.028135±0.001529	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	0.001384±0.000181	0.028583±0.001817	
	MOIRAI	0.027480±0.001248	0.125646±0.032147	0.020042±0.008949	0.108539±0.025857	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	
	GRU	0.001771±0.000422	0.030191±0.002642	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	0.022143±0.002326	0.001058±0.000341	
	LSTM	0.002610±0.001038	0.036192±0.008114	0.001663±0.000809	0.028472±0.007906	0.001526±0.000680	0.027114±0.000734	0.001526±0.000680	0.027114±0.000734	0.001526±0.000680	0.027114±0.000734	0.001526±0.000680	0.027114±0.000734	0.001526±0.000680	0.027114±0.000734	0.001526±0.000680	0.027114±0.000734	0.001526±0.000680	
DL-linear	0.001812±0.000069	0.031151±0.000066	0.001291±0.000066	0.012977±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066	0.000530±0.000066		
SparseTSF	0.035730±0.002514	0.168405±0.005838	0.035965±0.002312	0.169078±0.005354	0.035635±0.002174	0.168330±0.005058	0.035635±0.002174	0.168330±0.005058	0.035635±0.002174	0.168330±0.005058	0.035635±0.002174	0.168330±0.005058	0.035635±0.002174	0.168330±0.005058	0.035635±0.002174	0.168330±0.005058	0.035635±0.002174		
FredFormer	<b>0.006683±0.000039</b>	<b>0.191413±0.000378</b>	<b>0.000291±0.000003</b>	<b>0.012536±0.000238</b>	<b>0.000284±0.000003</b>	<b>0.011649±0.000063</b>	<b>0.000284±0.000003</b>												
DAF (Ours)	46.52±33.64	21.24±12.15	1.69±2.01	18.33±1.74	0.000263±0.000001	0.011306±0.000078	0.000263±0.000001	0.011306±0.000078	0.000263±0.000001	0.011306±0.000078	0.000263±0.000001	0.011306±0.000078	0.000263±0.000001	0.011306±0.000078	0.000263±0.000001	0.011306±0.000078	0.000263±0.000001		
Improvements																			
Dataset	Time-MoE	2.206586±0.001331	1.244876±0.002756	1.241777±0.001319	2.207904±0.003516	1.245192±0.003516	2.207904±0.003516	1.245192±0.003516	2.207904±0.003516	1.245192±0.003516	2.207904±0.003516	1.245192±0.003516	2.207904±0.003516	1.245192±0.003516	2.207904±0.003516	1.245192±0.003516	2.207904±0.003516	1.245192±0.003516	
	Timer	1.842509±0.002412	1.201817±0.001146	1.907691±0.002429	2.223529±0.002429	1.221815±0.001915	1.895577±0.000949	1.895577±0.000949	2.223529±0.002429	1.221815±0.001915	1.895577±0.000949	1.895577±0.000949	2.223529±0.002429	1.221815±0.001915	1.895577±0.000949	1.895577±0.000949	2.223529±0.002429	1.221815±0.001915	
	Time-XL	1.991412±0.100731	1.253429±0.033347	2.020642±0.098785	2.62496±0.033104	2.082329±0.102437	2.28420±0.034912	2.017286±0.101257	2.017286±0.101257	2.017286±0.101257	2.017286±0.101257	2.017286±0.101257	2.017286±0.101257	2.017286±0.101257	2.017286±0.101257	2.017286±0.101257	2.017286±0.101257	2.017286±0.101257	
	Time-LLM	1.003338±0.106084	0.864225±0.063347	0.997125±0.114040	0.86177±0.067101	1.002028±0.110813	0.863277±0.065065	0.99694±0.111745	0.863277±0.065065	0.99694±0.111745	0.863277±0.065065	0.99694±0.111745	0.863277±0.065065	0.99694±0.111745	0.863277±0.065065	0.99694±0.111745	0.863277±0.065065	0.99694±0.111745	
	TTM	0.264644±0.157101	0.412749±0.171141	0.238366±0.140507	0.391605±0.160599	0.250782±0.148132	0.401709±0.165322	0.242744±0.143402	0.250782±0.148132	0.401709±0.165322	0.242744±0.143402	0.250782±0.148132	0.401709±0.165322	0.242744±0.143402	0.250782±0.148132	0.401709±0.165322	0.242744±0.143402	0.250782±0.148132	
	TimesFM	0.001336±0.000071	0.028191±0.001470	0.011384±0.000153	0.028383±0.001254	0.001481±0.000207	0.032912±0.001556	0.001350±0.000181	0.028383±0.001254	0.001481±0.000207	0.032912±0.001556	0.001350±0.000181	0.028383±0.001254	0.001481±0.000207	0.032912±0.001556	0.001350±0.000181	0.028383±0.001254	0.001481±0.000207	
	MOMENT	0.113542±0.0001503	0.174566±0.002275	0.112379±0.000153	0.173483±0.002029	0.114558±0.000201	0.175501±0.002453	0.112699±0.000863	0.114558±0.000201	0.175501±0.002453	0.112699±0.000863	0.114558±0.000201	0.175501±0.002453	0.112699±0.000863	0.114558±0.000201	0.175501±0.002453	0.112699±0.000863	0.114558±0.000201	
	MOIRAI	0.001008±0.000034	0.032745±0.003260	0.001199±0.000683	0.109947±0.022670	0.001080±0.000341	0.022106±0.003361	0.001080±0.000341	0.022106±0.003361	0.001080±0.000341	0.022106±0.003361	0.001080±0.000341	0.022106±0.003361	0.001080±0.000341	0.022106±0.003361	0.001080±0.000341	0.022106±0.003361	0.001080±0.000341	
	GRU	0.001539±0.000071	0.027314±0.007555	0.001480±0.000676	0.026535±0.007157	0.001621±0.000716	0.027612±0.007392	0.001420±0.000663	0.026535±0.007157	0.001621±0.000716	0.027612±0.007392	0.001420±0.000663	0.026535±0.007157	0.001621±0.000716	0.027612±0.007392	0.001420±0.000663	0.026535±0.007157	0.001621±0.000716	
	LSTM	0.000220±0.000054	0.008557±0.000054	0.000202±0.000001	0.007895±0.000049	0.000297±0.000002	0.009103±0.000041	0.000174±0.000001	0.000297±0.000002	0.009103±0.000041	0.000174±0.000001	0.000297±0.000002	0.009103±0.000041	0.000174±0.000001	0.000297±0.000002	0.009103±0.000041	0.000174±0.000001	0.000297±0.000002	
DAF (Ours)	0.035824±0.002220	0.168753±0.005155	0.035500±0.002094	0.167979±0.004888	0.035524±0.002145	0.168049±0.005009	0.035524±0.002145	0.168049±0.005009	0.035524±0.002145	0.168049±0.005009	0.035524±0.002145	0.168049±0.005009	0.035524±0.002145	0.168049±0.005009	0.035524±0.002145	0.168049±0.005009	0.035524±0.002145		
Improvements																			
Dataset	Time-MoE	2.213601±0.001399	1.247784±0.002490	1.240770±0.002270	2.203901±0.003516	1.244178±0.003516	2.203901±0.003516	1.244178±0.003516	2.203901±0.003516	1.244178±0.003516	2.203901±0.003516	1.244178±0.003516	2.203901±0.003516	1.244178±0.003516	2.203901±0.003516	1.244178±0.003516	2.203901±0.003516	1.244178±0.003516	
	Timer	1.799109±0.003348	1.187942±0.001472	1.895177±0.001442	2.219431±0.001442	1.215763±0.000981	1.863698±0.000957	1.863698±0.000957	2.219431±0.001442	1.215763±0.000981	1.863698±0.000957	1.863698±0.000957	2.219431±0.001442	1.215763±0.000981	1.863698±0.000957	1.863698±0.000957	2.219431±0.001442	1.215763±0.000981	
	Time-XL	1.979905±0.113992	1.248772±0.037813	2.011879±0.094944	2.60123±0.031677	2.01226±0.098757	2.29357±0.033180	1.797324±0.156836	1.88209±0.054933	2.01226±0.098757	2.29357±0.033180	1.797324±0.156836	1.88209±0.054933	2.01226±0.098757	2.29357±0.033180	1.797			

Table 17: Zero-shot evaluation on individual satellites of two cross-domain datasets (KUIPER and LEMUR) with mean and standard deviation in MSE and MAE. "Improvements" presents the relative performance of our model compared to the best baseline for each dataset and metric. A dash ('-') indicates that our model fails to achieve either the best or second-best performance. The best result is highlighted with underline, and the second-best is highlighted with underline.

Table with columns for Dataset, Metric, and performance metrics (MSE, MAE) for various models (KUIPER-33 to KUIPER-55, LEMUR-1). Rows are grouped by dataset and metric, showing mean values and standard deviations. Best and second-best results are underlined.

1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403

Table 18: Zero-shot evaluation on individual satellites of LEMUR dataset with mean and standard deviation in MSE and MAE. “Improvements” presents the relative performance of our model compared to the best baseline for each dataset and metric. A dash (‘-’) indicates that our model fails to achieve either the best or second-best performance. The best result is highlighted in **bold**, and the second-best is highlighted with underline.

Dataset	LEMUR-1	LEMUR-2	LEMUR-3	LEMUR-4	LEMUR-5	LEMUR-6	LEMUR-7
Time-MeE	2.270557±0.021837	1.181452±0.008054	1.181452±0.008054	1.177746±0.008657	1.180724±0.001799	2.245731±0.024363	1.171082±0.008491
Time-Timer	1.90327±0.004634	0.69750±0.001738	1.409772±0.004637	0.989095±0.001483	1.84423±0.006092	1.596563±0.003248	1.823798±0.002460
Time-XL	1.852936±0.0115626	1.124566±0.003649	1.85876±0.0108848	1.872772±0.009992	1.127624±0.003528	1.932366±0.003556	1.98179±0.0102448
Time-LLM	0.172625±0.0413632	0.107304±0.003470	0.107304±0.003470	0.107304±0.003470	0.107304±0.003470	0.107304±0.003470	1.48488±0.0100379
Time-TTM	0.517698±0.312095	0.542187±0.230418	0.517698±0.312095	0.526401±0.222597	0.534558±0.226887	0.416300±0.249964	0.824684±0.041809
TimeFOMer	0.001289±0.000143	0.029529±0.001290	0.029529±0.001290	0.001289±0.000143	0.001289±0.000143	0.001289±0.000143	0.281597±0.168388
MOMENT	0.18135±0.000867	0.173576±0.001749	0.18135±0.000867	0.173576±0.001749	0.173576±0.001749	0.173576±0.001749	0.026142±0.001140
MOIRAI	0.032609±0.0014574	0.128383±0.003279	0.128383±0.003279	0.032609±0.0014574	0.032609±0.0014574	0.032609±0.0014574	0.168794±0.000434
GRU	0.001669±0.000605	0.029787±0.003639	0.029787±0.003639	0.001669±0.000605	0.001669±0.000605	0.001669±0.000605	0.019855±0.0008316
LSTM	0.003557±0.000772	0.045564±0.004711	0.045564±0.004711	0.003557±0.000772	0.003557±0.000772	0.003557±0.000772	0.001530±0.000337
DL-neur	0.001531±0.000025	0.027507±0.000239	0.027507±0.000239	0.001531±0.000025	0.001531±0.000025	0.001531±0.000025	0.001764±0.000592
SparseTSF	0.002169±0.000012	0.037118±0.000088	0.037118±0.000088	0.002169±0.000012	0.002169±0.000012	0.002169±0.000012	0.000275±0.000001
FredFormer	0.036228±0.002888	0.157705±0.005664	0.157705±0.005664	0.036228±0.002888	0.036228±0.002888	0.036228±0.002888	0.036526±0.002300
DAF (Ours)	<b>0.00765±0.001051</b>	<b>0.020089±0.000284</b>	<b>0.019358±0.000322</b>	<b>0.00646±0.000142</b>	<b>0.019121±0.000297</b>	<b>0.000447±0.000026</b>	0.000304±0.000005
Improvements	40.65±6.21	22.84±3.22	25.50±2.99	47.12±2.17	46.09±5.21	33.67±3.90	0.012024±0.000058
Dataset	LEMUR-8	LEMUR-9	LEMUR-10	LEMUR-11	LEMUR-12	LEMUR-13	LEMUR-14
Time-MeE	2.133689±0.017159	1.134209±0.005950	2.118540±0.020653	1.152031±0.007701	2.120949±0.0015184	2.124512±0.022187	2.136297±0.015111
Time-Timer	2.60257±0.0005035	1.37182±0.000739	2.179095±0.000705	1.219771±0.000162	2.62818±0.000602	2.608792±0.000541	2.600266±0.000715
Time-XL	2.355040±0.003900	1.275336±0.002049	2.133608±0.006419	1.219771±0.000162	2.394136±0.001951	2.368578±0.0009649	2.363068±0.007367
Time-LLM	0.967904±0.166656	0.789511±0.088922	0.91568±0.141623	0.800104±0.075521	0.96834±0.0148026	0.966455±0.16921	0.966254±0.167289
Time-TTM	0.075748±0.042529	0.189500±0.069417	0.153156±0.088904	0.288336±0.113416	0.172674±0.085924	0.68029±0.037627	0.181330±0.065030
TimeFOMer	0.001936±0.000135	0.30895±0.000625	0.001355±0.000133	0.026780±0.000720	0.00183±0.000198	0.003007±0.000762	0.071625±0.040015
MOMENT	0.111667±0.000682	0.150692±0.000221	0.115777±0.001618	0.165987±0.002465	0.001485±0.000126	0.11387±0.001505	0.11939±0.000094
MOIRAI	0.010085±0.002856	0.027949±0.011058	0.020182±0.008391	0.015798±0.004937	0.009609±0.002470	0.009676±0.002659	0.11939±0.000094
GRU	0.00591±0.0002771	0.082138±0.015740	0.002766±0.00316	0.039083±0.00316	0.003768±0.000317	0.004017±0.019545	0.003886±0.002858
LSTM	0.004992±0.0012192	0.05026±0.010222	0.002014±0.001120	0.034288±0.000947	0.003662±0.000952	0.003909±0.001292	0.044615±0.019092
DL-neur	0.000259±0.000002	0.010525±0.000050	0.000276±0.000001	0.010114±0.000031	0.002628±0.000002	0.010779±0.000081	0.040497±0.0008177
SparseTSF	0.000584±0.000003	0.017722±0.000056	0.000331±0.000000	0.014590±0.000166	0.00061±0.000003	0.000359±0.000002	0.004977±0.000001
FredFormer	0.031324±0.001445	0.146597±0.002848	0.034888±0.0001004	0.153849±0.003271	0.0031289±0.000333	0.0037342±0.000052	0.000483±0.000003
DAF (Ours)	<b>0.00073±0.000006</b>	<b>0.008695±0.000311</b>	<b>0.000210±0.000016</b>	<b>0.009446±0.000522</b>	<b>0.00015±0.000000</b>	<b>0.00020±0.000008</b>	0.014574±0.0001382
Improvements	33.20±2.02	17.37±3.60	6.06±0.73	6.60±3.52	23.31±5.57	44.36±1.41	0.008040±0.000080
Dataset	LEMUR-15	LEMUR-16	LEMUR-17	LEMUR-18	LEMUR-19	LEMUR-20	LEMUR-21
Time-MeE	2.11134±0.001978	1.086436±0.006624	2.274876±0.026810	1.05931±0.004222	2.261404±0.023480	1.198569±0.008118	2.245052±0.018549
Time-Timer	2.676572±0.000649	1.309843±0.001399	2.665756±0.006953	1.330818±0.001415	1.489952±0.003728	1.022097±0.001359	1.998521±0.001309
Time-XL	2.435532±0.035400	1.251803±0.007971	1.798358±0.157926	1.127594±0.051692	1.848669±0.141633	1.42696±0.04826	1.74925±0.032763
Time-LLM	0.961370±0.169408	0.738855±0.086889	0.960600±0.171720	0.744140±0.085833	1.05927±0.054553	0.844310±0.036129	1.042343±0.081873
Time-TTM	0.065451±0.031644	0.172857±0.0661592	0.062004±0.034083	0.162619±0.065287	0.50085±0.302321	0.50852±0.215003	0.344382±0.206961
TimeFOMer	0.001991±0.000232	0.15206±0.001683	0.002332±0.000512	0.031460±0.002392	0.001229±0.000697	0.026050±0.001114	0.001305±0.000136
MOMENT	0.111822±0.002247	0.15206±0.001683	0.112604±0.002988	0.151079±0.002474	0.117557±0.000359	0.17427±0.0001592	0.027126±0.001302
MOIRAI	0.009285±0.002417	0.067767±0.0099305	0.009220±0.002232	0.035558±0.015131	0.125251±0.034443	0.029265±0.013147	0.114891±0.001303
GRU	0.003279±0.001145	0.044785±0.008413	0.004974±0.001614	0.05167±0.0008877	0.004805±0.000921	0.052205±0.006345	0.023176±0.010066
LSTM	0.000357±0.000003	0.011672±0.000061	0.000305±0.000004	0.010137±0.000110	0.002486±0.000217	0.040721±0.000091	0.000431±0.000006
SparseTSF	0.000500±0.000006	0.015707±0.000097	0.000395±0.000004	0.013328±0.000169	0.03959±0.002747	0.16121±0.005606	0.000352±0.000003
FredFormer	0.031641±0.001236	0.139324±0.001990	0.029395±0.000936	0.13328±0.001689	0.036491±0.002744	0.16121±0.005606	0.036741±0.002473
DAF (Ours)	<b>0.000201±0.000006</b>	<b>0.008867±0.000267</b>	<b>0.000172±0.000010</b>	<b>0.008717±0.000092</b>	<b>0.000799±0.000057</b>	<b>0.01158±0.000354</b>	<b>0.000339±0.000008</b>
Improvements	44.34±1.87	24.04±3.04	6.52±0.59	21.72±1.08	40.68±3.70	20.13±3.13	0.015279±0.00213
Dataset	LEMUR-22	LEMUR-23	LEMUR-24	LEMUR-25	LEMUR-26	LEMUR-27	LEMUR-28
Time-MeE	2.259580±0.019395	1.194716±0.0097511	2.232461±0.021311	1.180151±0.006324	2.133343±0.019063	1.149435±0.006154	2.103846±0.012847
Time-Timer	1.492539±0.002652	1.020331±0.000849	1.757170±0.002103	1.098348±0.000768	2.566920±0.001572	1.344317±0.004673	2.785524±0.008022
Time-XL	1.860203±0.142282	1.143329±0.044844	1.94746±0.118771	1.165727±0.037414	2.331246±0.024693	1.82782±0.06763	2.479396±0.050141
Time-LLM	0.057141±0.059496	0.840003±0.037907	0.029420±0.089325	0.83269±0.05102	0.97197±0.165883	0.96849±0.165421	0.955286±0.176233
Time-TTM	0.428488±0.259531	0.100179±0.213881	0.258477±0.195959	0.436801±0.180898	0.07905±0.040530	0.197843±0.086678	0.046676±0.024475
TimeFOMer	0.001330±0.000138	0.026849±0.001258	0.001310±0.000124	0.026657±0.001924	0.001895±0.000662	0.001902±0.000690	0.001944±0.000317
MOMENT	0.117078±0.000514	0.173609±0.001354	0.115242±0.001294	0.168573±0.001718	0.131161±0.000469	0.163955±0.000582	0.119251±0.000959
MOIRAI	0.02849±0.012802	0.122244±0.036938	0.022224±0.009684	0.199383±0.026765	0.017494±0.011418	0.017494±0.011418	0.09119±0.002249
GRU	0.001826±0.000253	0.031955±0.002745	0.001304±0.000045	0.027938±0.005224	0.00828±0.002768	0.048347±0.002377	0.01812±0.003537
LSTM	0.00272±0.000939	0.03853±0.008368	0.004352±0.000621	0.031972±0.00732	0.00265±0.000916	0.03923±0.007671	0.02846±0.001401
DL-neur	0.001030±0.000006	0.030769±0.000029	0.000375±0.000002	0.01748±0.000044	0.000383±0.000001	0.000383±0.000001	0.000383±0.000001
SparseTSF	0.034558±0.002700	0.160330±0.004543	0.034603±0.003306	0.076518±0.013590	0.034603±0.003306	0.034603±0.003306	0.034603±0.003306
FredFormer	0.04658±0.002700	0.160330±0.004543	0.04658±0.002700	0.04658±0.002700	0.04658±0.002700	0.04658±0.002700	0.04658±0.002700
DAF (Ours)	<b>0.000566±0.000039</b>	<b>0.017838±0.000381</b>	<b>0.000327±0.000004</b>	<b>0.012939±0.000157</b>	<b>0.000167±0.000004</b>	<b>0.008387±0.000432</b>	<b>0.00016±0.000009</b>
Improvements	43.42±3.65	21.08±2.21	2.91±0.43	21.09±2.43	37.99±1.00	37.45±1.33	37.04±1.39



1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511

Table 20: Zero-shot evaluation on individual satellites of two cross-domain datasets (LEMUR and SKYSAT) with mean and standard deviation in MSE and MAE. “Improvements” presents the relative performance of our model compared to the best baseline for each dataset and metric. A dash (‘-’) indicates that our model fails to achieve either the best or second-best performance. The best result is highlighted in bold, and the second-best is highlighted with underline.

Dataset	LEMUR-50			LEMUR-52			LEMUR-53			LEMUR-54			LEMUR-55					
	MSE	MAE	MAE	MSE	MAE	MAE	MSE	MAE	MAE	MSE	MAE	MAE	MSE	MAE	MAE			
Time-MeE	2.262186±0.0021645	1.177649±0.0007740	1.181452±0.0080854	2.270776±0.024840	1.181442±0.008499	1.184420±0.006764	2.263670±0.0068657	1.177746±0.0008657	2.269115±0.028305	1.180724±0.010799	1.180243±0.0024363	2.245717±0.024363	1.171082±0.008491					
Time-XL	1.456705±0.002565	0.993270±0.001740	1.245660±0.001738	1.499327±0.004634	1.185760±0.016070	1.499399±0.001670	1.449399±0.001670	1.177746±0.0008657	1.484232±0.006092	0.979079±0.002248	1.180243±0.0024363	1.596365±0.0024363	1.039972±0.001411					
Time-LLM	1.874711±0.013555	1.131665±0.0032705	1.296750±0.0036349	1.858761±0.108848	1.126660±0.010317	1.872722±0.109992	1.126660±0.010317	1.130132±0.0034541	1.864888±0.121155	1.127624±0.003598	1.180243±0.0024363	1.932366±0.003535	1.148484±0.002956					
Time-TTM	0.369845±0.022085	0.463222±0.193351	0.835038±0.027098	0.671192±0.044603	0.835159±0.026793	0.667441±0.049129	0.835159±0.026793	0.623280±0.038274	1.068855±0.043839	0.833304±0.030766	0.833304±0.030766	1.030314±0.009734	0.284684±0.041809					
TimeFM	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047	0.005239±0.000047					
MOMENT	0.118409±0.000509	0.172925±0.0001321	0.118409±0.000509	0.118409±0.000509	0.118409±0.000509	0.118409±0.000509	0.118409±0.000509	0.118409±0.000509	0.118409±0.000509	0.118409±0.000509	0.118409±0.000509	0.118409±0.000509	0.118409±0.000509					
MOIRAI	0.030570±0.0013680	0.124132±0.001924	0.032605±0.0014574	0.032605±0.0014574	0.032605±0.0014574	0.032605±0.0014574	0.032605±0.0014574	0.032605±0.0014574	0.032605±0.0014574	0.032605±0.0014574	0.032605±0.0014574	0.032605±0.0014574	0.032605±0.0014574					
GRU	0.005783±0.001049	0.059985±0.0035357	0.005799±0.001056	0.005799±0.001056	0.005799±0.001056	0.005799±0.001056	0.005799±0.001056	0.005799±0.001056	0.005799±0.001056	0.005799±0.001056	0.005799±0.001056	0.005799±0.001056	0.005799±0.001056					
LSTM	0.007353±0.0002806	0.062792±0.0113355	0.007348±0.002818	0.007348±0.002818	0.007348±0.002818	0.007348±0.002818	0.007348±0.002818	0.007348±0.002818	0.007348±0.002818	0.007348±0.002818	0.007348±0.002818	0.007348±0.002818	0.007348±0.002818					
DL-neur	0.006956±0.000129	0.061157±0.000587	0.006868±0.000125	0.006868±0.000125	0.006868±0.000125	0.006868±0.000125	0.006868±0.000125	0.006868±0.000125	0.006868±0.000125	0.006868±0.000125	0.006868±0.000125	0.006868±0.000125	0.006868±0.000125					
SparseTSF	0.012238±0.000042	0.090675±0.000102	0.012020±0.000040	0.012020±0.000040	0.012020±0.000040	0.012020±0.000040	0.012020±0.000040	0.012020±0.000040	0.012020±0.000040	0.012020±0.000040	0.012020±0.000040	0.012020±0.000040	0.012020±0.000040					
FredFormer	0.032808±0.0001347	1.50009±0.002330	0.032880±0.001384	0.032880±0.001384	0.032880±0.001384	0.032880±0.001384	0.032880±0.001384	0.032880±0.001384	0.032880±0.001384	0.032880±0.001384	0.032880±0.001384	0.032880±0.001384	0.032880±0.001384					
DAF (Ours)	<b>0.002218±0.000091</b>	<b>0.035246±0.000736</b>	<b>0.002179±0.000076</b>															
Improvements	57.68±0.56	36.27±0.32	56.42±0.40	57.39±0.36	35.33±0.07	57.36±0.39	35.92±0.40	57.36±0.39	35.92±0.40	57.36±0.39	35.92±0.40	57.36±0.39	35.92±0.40					
Dataset	LEMUR-56			LEMUR-57			LEMUR-58			LEMUR-59			LEMUR-60			LEMUR-61		
Time-MeE	2.226021±0.023306	1.167437±0.007895	2.133689±0.017159	2.226061±0.0101691	1.167599±0.006040	2.185401±0.0202653	1.152031±0.0077701	2.120949±0.015184	1.107181±0.004279	2.124512±0.022187	1.123984±0.007318	2.245717±0.024363	1.171082±0.008491					
Time-XL	1.823798±0.002460	1.116722±0.000926	2.602057±0.005035	1.737182±0.000739	1.806608±0.001634	1.117995±0.000705	1.219771±0.001921	2.628155±0.006602	1.316072±0.001955	2.608792±0.005341	1.33044±0.001177	2.608792±0.005341	1.33044±0.001177					
Time-LLM	1.981719±0.104828	1.167001±0.031754	2.355040±0.003900	1.275336±0.002049	1.975695±0.108719	1.165562±0.032977	1.211971±0.001921	2.394136±0.014862	1.259023±0.002931	2.36878±0.009649	1.27057±0.0089321	2.36878±0.009649	1.27057±0.0089321					
Time-TTM	0.020443±0.0100379	0.814670±0.056371	0.967904±0.166656	0.789511±0.088232	1.02458±0.098862	0.816201±0.055487	0.800104±0.075521	0.968342±0.168026	0.773080±0.085769	0.966455±0.169121	0.78237±0.089321	0.966455±0.169121	0.78237±0.089321					
TimeFM	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000					
MOMENT	0.011429±0.000134	0.156970±0.001280	0.011257±0.000132	0.011257±0.000132	0.011257±0.000132	0.011257±0.000132	0.011257±0.000132	0.011257±0.000132	0.011257±0.000132	0.011257±0.000132	0.011257±0.000132	0.011257±0.000132	0.011257±0.000132					
MOIRAI	0.019855±0.0008316	0.164970±0.001543	0.011685±0.002856	0.011685±0.002856	0.011685±0.002856	0.011685±0.002856	0.011685±0.002856	0.011685±0.002856	0.011685±0.002856	0.011685±0.002856	0.011685±0.002856	0.011685±0.002856	0.011685±0.002856					
GRU	0.001162±0.000527	0.025675±0.023908	0.001162±0.000527	0.001162±0.000527	0.001162±0.000527	0.001162±0.000527	0.001162±0.000527	0.001162±0.000527	0.001162±0.000527	0.001162±0.000527	0.001162±0.000527	0.001162±0.000527	0.001162±0.000527					
LSTM	0.004117±0.000165	0.047210±0.0099841	0.004195±0.001659	0.004195±0.001659	0.004195±0.001659	0.004195±0.001659	0.004195±0.001659	0.004195±0.001659	0.004195±0.001659	0.004195±0.001659	0.004195±0.001659	0.004195±0.001659	0.004195±0.001659					
DL-neur	0.002365±0.000039	0.035145±0.000311	0.002340±0.000039	0.002340±0.000039	0.002340±0.000039	0.002340±0.000039	0.002340±0.000039	0.002340±0.000039	0.002340±0.000039	0.002340±0.000039	0.002340±0.000039	0.002340±0.000039	0.002340±0.000039					
SparseTSF	0.003443±0.000017	0.048161±0.005069	0.003509±0.000016	0.003509±0.000016	0.003509±0.000016	0.003509±0.000016	0.003509±0.000016	0.003509±0.000016	0.003509±0.000016	0.003509±0.000016	0.003509±0.000016	0.003509±0.000016	0.003509±0.000016					
FredFormer	0.035521±0.002261	1.57894±0.005566	0.035536±0.002268	0.035536±0.002268	0.035536±0.002268	0.035536±0.002268	0.035536±0.002268	0.035536±0.002268	0.035536±0.002268	0.035536±0.002268	0.035536±0.002268	0.035536±0.002268	0.035536±0.002268					
DAF (Ours)	<b>0.000941±0.000056</b>	<b>0.0234069±0.000291</b>	<b>0.000983±0.000054</b>															
Improvements	18.99±4.83	10.15±23.92	16.83±4.66	16.83±4.66	16.83±4.66	16.83±4.66	16.83±4.66	16.83±4.66	16.83±4.66	16.83±4.66	16.83±4.66	16.83±4.66	16.83±4.66					
Dataset	LEMUR-62			SKYSAT-1			SKYSAT-2			SKYSAT-3			SKYSAT-4			SKYSAT-5		
Time-MeE	2.136297±0.015111	1.142362±0.006351	2.111134±0.019578	2.136297±0.015111	1.142362±0.006351	2.080526±0.021200	1.058333±0.004776	2.286254±0.026995	1.217689±0.010161	2.340616±0.0305415	1.220237±0.013082	2.233311±0.024398	1.195440±0.008174					
Time-XL	2.602666±0.0004903	1.348083±0.000715	2.676572±0.006479	1.309843±0.001399	2.65756±0.006953	1.330818±0.001415	2.67273±0.006282	2.482562±0.072837	2.32523±0.017217	1.765544±0.168296	1.23196±0.055624	1.732426±0.020258	1.10940±0.000999					
Time-LLM	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000					
Time-TTM	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000	0.000000±0.000000					
TimeFM	0.004652±0.000025	0.025094±0.000023	0.00123±0.000090	0.026190±0.001019	0.009429±0.000411	0.075920±0.001804	0.001311±0.000021	0.026451±0.0027916	0.570121±0.243046	0.797171±0.483974	0.680548±0.291897	0.321564±0.198893	0.375250±0.052013					
MOMENT	0.111939±0.000867	0.158019±0.000471	0.111822±0.002237	0.152506±0.001683	0.112604±0.002988	0.151079±0.002476	0.120214±0.000681	0.112604±0.002988	0.182154±0.001790	0.001295±0.0001249	0.026674±0.001249	0.001271±0.000101	0.026694±0.01158					
MOIRAI	0.098930±0.002762	0.072845±0.0110957	0.099282±0.002417	0.067767±0.009595	0.099220±0.02232	0.066258±0.008451	0.057525±0.016833	0.140818±0.0364629	0.062578±0.005607	0.19428±0.043562	0.022394±0.009745	0.022394±0.009745	0.19428±0.043562					
GRU	0.005524±0.001026	0.054680±0.003559	0.001947±0.001504	0.031954±0.012274	0.006250±0.000653	0.058182±0.001780	0.001536±0.000573	0.039375±0.006629	0.039375±0.006629	0.035136±0.006507	0.001978±0.000772	0.001978±0.000772	0.034376±0.006244					
LSTM	0.007530±0.002751	0.06159±0.013338	0.005685±0.000929	0.056436±0.008514	0.009211±0.003178	0.0770850±0.0128												

1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565

Table 21: Zero-shot evaluation on individual satellites of SKYSAT dataset with mean and standard deviation in MSE and MAE. “Improvements” presents the relative performance of our model compared to the best baseline for each dataset and metric. A dash (‘-’) indicates that our model fails to achieve either the best or second-best performance. The best result is highlighted in **bold**, and the second-best is highlighted with underline.

Dataset Metric	SKYSAT-12		SKYSAT-13		SKYSAT-14		SKYSAT-15	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Time-MoE	2.236464±0.017642	1.190623±0.005391	2.240123±0.017810	1.193063±0.005769	2.222521±0.019208	1.183334±0.006306	2.187511±0.016989	1.176568±0.005243
Timer	1.773883±0.002530	1.120311±0.000622	1.738289±0.001117	1.109440±0.000759	1.870675±0.001904	1.147982±0.000934	2.157144±0.001933	1.241508±0.000774
Timer-XL	1.979871±0.097502	1.185919±0.030516	1.961218±0.097415	1.180946±0.030776	2.020076±0.108422	1.195464±0.033309	2.133858±0.074213	1.237540±0.022912
TTM	0.329865±0.195286	0.440593±0.181416	0.346014±0.205206	0.451804±0.186507	0.262402±0.154509	0.393302±0.160666	0.158082±0.090164	0.305651±0.118242
TimesFM	0.001435±0.000168	0.027763±0.001373	0.001407±0.000161	0.027686±0.001365	0.001317±0.000122	0.026855±0.001296	0.001418±0.000391	0.028006±0.000827
MOMENT	0.115225±0.000494	0.169201±0.001709	0.116160±0.000915	0.170774±0.002496	0.114590±0.001089	0.167235±0.002416	0.113804±0.000451	0.163318±0.001895
MOIRAI	0.021122±0.009072	0.106200±0.025822	0.022382±0.009738	0.109233±0.026976	0.019134±0.008025	0.101078±0.023960	0.013851±0.005079	0.087202±0.017715
GRU	0.001125±0.000394	0.026258±0.004584	0.001105±0.000493	0.025847±0.005708	0.002014±0.000767	0.034824±0.006402	0.003570±0.001394	0.045992±0.008385
LSTM	0.002181±0.000615	0.035355±0.006311	0.002395±0.000616	0.036996±0.006034	0.001860±0.000630	0.032204±0.006240	0.001476±0.000497	0.029140±0.006224
DLinear	<b>0.000306±0.000003</b>	<b>0.010729±0.000078</b>	0.000399±0.000003	0.012524±0.000086	<b>0.000228±0.000002</b>	0.008665±0.000052	0.000224±0.000001	0.009644±0.000039
SparseTSF	0.000408±0.000001	0.012135±0.000031	0.000351±0.000001	<b>0.012471±0.000046</b>	0.000233±0.000001	<b>0.008454±0.000088</b>	0.000419±0.000007	0.015087±0.000172
FredFormer	0.036736±0.002332	0.161403±0.004648	0.036799±0.002329	0.161707±0.004668	0.036341±0.002247	0.159876±0.004398	0.034907±0.001832	0.157590±0.003574
DAF (Ours)	0.000314±0.000002	0.012669±0.000110	<b>0.000366±0.000008</b>	0.013558±0.000193	0.000269±0.000007	0.011493±0.000078	<b>0.000179±0.000011</b>	<b>0.009352±0.000686</b>
Improvements	-2.55±0.76	-	8.26±2.22	-8.23±1.80	-	-	20.09±6.43	3.03±7.13