ANNALS OF GEOPHYSICS, 69, 2026; doi:10.4401/ag-9518
OPEN ACCESS

Extending the Predictive Horizon of Earth's Polar Motion Using a Hybrid Deep-Learning Framework

Xuyu Hong[1], Tianyi Chen[*,2], Hongtai Xie[2], Jiangping Wang[2], Yumei Guan[1], Yuanhong Yang[1], Dong Hu[2]

(1) Fujian Earthquake Agency, Fujian, China
(2) AVIC Changcheng Institute of Metrology and Measurement, Beijing, China

Article history: received December 15, 2025; accepted March 31, 2026

Abstract

Polar motion (PM) describes the motion of the Earth's rotation axis as it wanders across the Earth's crust and is essential for space geodesy, navigation, and precise geophysical measurements. Real-time PM data are not directly observable, and existing prediction models, such as those in IERS Bulletin A, typically provide forecasts limited to one year. This study proposes a deep learning framework based on a Long Short-Term Memory (LSTM) network augmented with a multi-head attention mechanism for long-term PM prediction. The model jointly predicts PMX and PMY components using differenced input sequences to capture nonlinear temporal dependencies and multi-frequency periodic behaviors. Experiments on the IERS C04 dataset demonstrate that the proposed model substantially outperforms classical linear predictors. Specifically, the proposed model reduces the total prediction error by approximately 61% at a 600-day horizon and 39% at 1100 days compared with a conventional baseline, achieving a mean absolute error of 16 mas at 600 days and 32 mas at 1100 days. This demonstrates the effectiveness of hybrid LSTM-attention architectures in capturing long-range temporal dynamics in geophysical time series and their potential for extended Earth rotation forecasting.

Keywords: Polar Motion Prediction; Earth Orientation Parameters; Long-term Prediction; Deep Learning; LSTM-Attention Model; Time Series Analysis; Earth Rotation

1. Introduction

Polar motion (PM) describes the motion of the Earth's rotation axis relative to the Earth's crust (Gross, 2007), i.e., the wandering of the rotation axis across the Earth's surface. It is defined as the position of the Celestial Intermediate Pole (CIP) in the International Terrestrial Reference Frame (ITRF) and is conventionally expressed by the coordinates x_p and y_p (Petit and Luzum, 2010). As one of the fundamental Earth Orientation Parameters (EOP), PM – together with UT1-UTC, length of day (LOD), nutation (dX/dY), and precession – characterizes the instantaneous rotational state of the Earth.

PM observations are derived from a combination of advanced space geodetic techniques, including the Global Navigation Satellite System (GNSS) (Byram and Hackman, 2012), Very Long Baseline Interferometry (VLBI) (Schuh and

Böhm, 2013), and Satellite Laser Ranging (SLR) (Coulot et al., 2010). However, because the data from these systems differ in timeliness and require extensive post-processing and corrections, the estimated PM values are not immediately available in real time. Consequently, applications that depend on real-time Earth Rotation Parameters (ERP) – such as spacecraft navigation and high-precision geodetic measurements – must rely on PM predictions.

Several organizations provide measurements and forecasts of Polar Motion (PM) data. Among them, one of the most widely used Earth Rotation Parameter (ERP) datasets is the IERS C04 series, published by the International Earth Rotation and Reference Systems Service (IERS). The IERS C04 data are typically released with a delay of several days to weeks to ensure that the final solutions incorporate diverse observational techniques. However, many applications in spaceflight, astronomy, and geophysics require real-time PM information, which the delayed IERS C04 series cannot provide. To address this limitation, IERS also issues ERP forecasts in its Bulletin A, which includes two types of predictions with different time spans and accuracies: short-term forecasts (covering the next 7-10 days) and long-term forecasts (extending up to 365 days). Bulletin A predictions have traditionally served as an important operational reference. However, in recent years, broader community-based evaluation efforts have been conducted to systematically assess the performance of different EOP prediction strategies. In addition to IERS, several other institutions also produce ERP predictions, such as the U.S. Naval Observatory (USNO) (Luzum et al., 2001), NASA's Jet Propulsion Laboratory (JPL) (Gross et al., 1998), and the European Space Agency (ESA) (Bruni et al., 2021; Kehm et al., 2023).

Many scientific disciplines and research institutions have demonstrated strong interest and demand for accurate Earth Orientation Parameter (EOP) prediction. Over the past two decades, this demand has led to the organization of two international ERP prediction campaigns. The first Earth Orientation Parameters Prediction Comparison Campaign (1st EOP PCC) was conducted between 2005 and 2008 (Kalarus et al., 2010).

With the rapid development of new methodologies – particularly machine learning (ML)-based approaches – a second international evaluation initiative, the 2nd Earth Orientation Parameters Prediction Comparison Campaign (2nd EOP PCC), was organized in 2021. This campaign gathered more than 50 research groups worldwide and provided a comprehensive and systematic intercomparison of state-of-the-art EOP prediction algorithms under unified evaluation criteria. The results of the campaign have been reported in several recent studies, including Kur et al. (2024), Sliwinska et al. (2023, 2024), and Winska et al. (2024), covering short-term and medium-term prediction performance and establishing robust community benchmarks for EOP forecasting.

Existing Polar Motion (PM) prediction methods can be broadly divided into two categories: extrapolation-based approaches that rely solely on historical PM time series, and excitation-informed approaches that incorporate geophysical forcing functions such as effective angular momentum (EAM).

Extrapolation-based methods can be further classified into linear and nonlinear models. Linear prediction methods are typically constructed by fitting the periodic components of historical PM data – based on its inherent periodicity – using the least squares method and subsequently forecasting the residuals through various techniques (Akulenko et al., 2002; Kosek et al., 1998; Niedzielski and Kosek, 2008; Wu et al., 2018; Yao et al., 2013). Among these, the method combining least squares and autoregression (LS+AR), along with its variants, represents one of the best-performing linear approaches (L. Wang et al., 2022, 2023; Wu et al., 2018; Xu and Zhou, 2015).

Nonlinear prediction methods are primarily based on machine learning techniques. These approaches generally employ network architectures such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and long short-term memory (LSTM) networks to learn from historical PM data and then use the trained models to make predictions from recent PM observations (Jin et al., 2021; Liao et al., 2012; Niu et al., 2025; L. Wang et al., 2025; X. Q. Wang, 2024; Yu et al., 2023, 2024). According to the literature, CNNs can capture latent features in multidimensional data and fit nonlinear relationships, thereby effectively mitigating the uncertainties inherent in the original time series (Kalarus et al., 2010). Among neural network architectures, LSTM and Gated Recurrent Units (GRUs) demonstrate significant advantages in modeling temporal dependencies within time series data (X. Q. Wang, 2024). In recent years, hybrid approaches that integrate LSTM with other network structures have shown excellent performance in PM prediction. The method proposed in this paper also adopts an LSTM-based network architecture enhanced with an attention mechanism.

Excitation-informed approaches explicitly incorporate geophysical excitation functions derived from effective angular momentum (EAM), including atmospheric (AAM), oceanic (OAM), and hydrological (HAM) contributions. The physical relationship between angular momentum exchange within the Earth system and variations in Earth rotation has been comprehensively documented (Gross, 2015). Building upon this physical framework, several studies

have demonstrated that incorporating EAM information can substantially enhance Earth orientation parameter prediction, particularly for short-term forecasts (Dobslaw et al., 2018; Dill et al., 2019). Recent developments have further extended physics-based EAM forecasts to 10 days and shown measurable improvements in short-term ERP prediction accuracy (Dill et al., 2025). Results from the Second Earth Orientation Parameters Prediction Comparison Campaign also confirm that prediction strategies utilizing EAM-based inputs generally achieve superior performance in operational short-term forecasting (Śliwińska et al., 2025). These studies collectively highlight the importance of physically meaningful excitation information in contemporary Earth rotation prediction.

For high-precision measurement instruments such as absolute gravimeters, it is often impractical to frequently synchronize Polar Motion (PM) data during long-term operation. Performing regular maintenance solely to update PM parameters can result in considerable cost and operational complexity. In addition, many geodetic applications inherently require multi-year continuity of stable Earth orientation parameters, where short prediction horizons are insufficient. Typical examples include continuously operating gravity gradiometers and absolute gravity reference-station chains, which cannot be frequently maintained, as well as national and global terrestrial reference frame updates (e.g., ITRF/CGCS), whose revision cycles span multiple years and therefore depend on long-period PM predictions.

However, existing PM prediction models – typically limited to about one year – cannot fully support these long-duration tasks. To address this limitation, this paper proposes a hybrid LSTM-attention neural network model for long-term PM prediction. The proposed model achieves a three-year (approximately 1100-day) PM forecast using deep learning, demonstrating high stability and accuracy. This advancement effectively extends the applicable range of PM prediction and offers significant potential for long-period geodetic and navigation applications.

2. Method

2.1 Overall model architecture

This study proposes a deep learning architecture for polar motion (PM) prediction based on Long Short-Term Memory (LSTM) networks enhanced with a multi-head self-attention mechanism. The overall structure of the model is illustrated in Fig. 1.

The model receives historical PM observations as input and predicts future PM values in a multi-step forecasting framework. The input sequence consists of the two-dimensional PM components (PMX, PMY) with a shape of (*lookback*, 2), where *lookback* denotes the number of historical days used as input. The model output is the predicted PM sequence for the following *steps* days with a shape of (*steps*, 2).

The network architecture consists of three LSTM layers and a multi-head attention module. The input sequence is first processed by two stacked LSTM layers, each containing 64 hidden units and configured to return the full output sequence. These layers mainly extract short-term temporal features from the time series. The output sequence of the second LSTM layer is then fed into a multi-head self-attention module, which models the relationships among different time steps and enhances the contextual representation of the sequence.

The attention output is combined with the original LSTM output through a residual connection, followed by layer normalization to stabilize the training process and maintain the integrity of the sequential information. After this attention-residual block, a third LSTM layer with 32 hidden units performs the final temporal encoding of the enhanced sequence.

Finally, a fully connected layer maps the extracted features to the target prediction space. Specifically, a Dense layer projects the latent representation into a vector of length $steps \times 2$, which is subsequently reshaped into an output tensor of size (*steps*, 2). This design enables the model to perform multi-step prediction of both PM components simultaneously.

To improve model stability and generalization, the input sequence is constructed using the first-order differenced values of the original PM series rather than the raw observations. This preprocessing step reduces the tendency of the model to converge toward trivial mean-value solutions when predicting long-term quasi-periodic signals. The predicted differences are subsequently integrated to recover the final PM values.

In addition, the model is trained jointly on both PMX and PMY components, preserving the intrinsic two-dimensional nature of Earth's polar motion and allowing the model to learn correlated spatio-temporal dynamics between the two components.

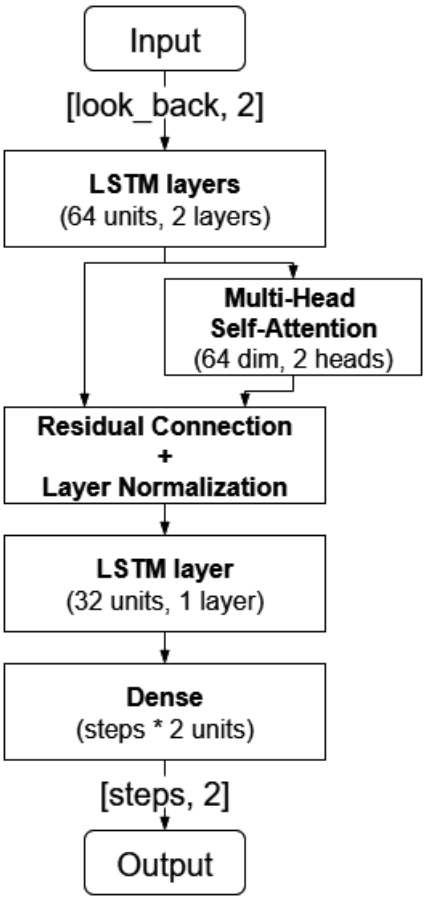

Figure 1. Model structure.

In summary, the proposed model can be viewed as an LSTM-based framework augmented with a multi-head attention module. This hybrid design enables the model to capture both short-term variations and long-range dependencies, significantly improving its ability to represent the complex dynamics of Earth's polar motion. The following subsections provide a detailed description of the two core components: the LSTM network and the attention mechanism.

2.2 LSTM

This study employs Long Short-Term Memory (LSTM) networks for polar motion prediction. LSTM is a specialized form of recurrent neural network (RNN) designed to overcome the vanishing gradient problem that hampers traditional RNNs (Gers et al., 2000). As an effective tool for time series analysis, LSTM networks can capture complex temporal dependencies where current observations depend on historical states over extended periods.

The standard LSTM architecture without peephole connections is adopted in this study. Its gating architecture – consisting of input, forget, and output gates – allows the network to selectively retain, update, or discard information, thereby enabling effective long-term dependency modeling.

An LSTM cell consists of the following core components:

- **Cell state** (C_t): Serves as the long-term memory of the cell, maintaining information across multiple time steps, akin to a conveyor belt that facilitates gradient flow.
- **Hidden state** (h_t): Represents the short-term memory and serves as the output passed to the next time step or subsequent layers.
- **Gates**: Three gating units regulate the information flow:
  - *Forget gate* (f_t): Determines which information from the previous cell state should be discarded.
  - *Input gate* (i_t): Controls which new information should be incorporated into the cell state.

- *Output gate* (o_t): Determines which part of the cell state is exposed as output.

Each gate uses a sigmoid activation function (σ), producing values in the range [0,1], where 0 blocks and 1 fully allows information flow.

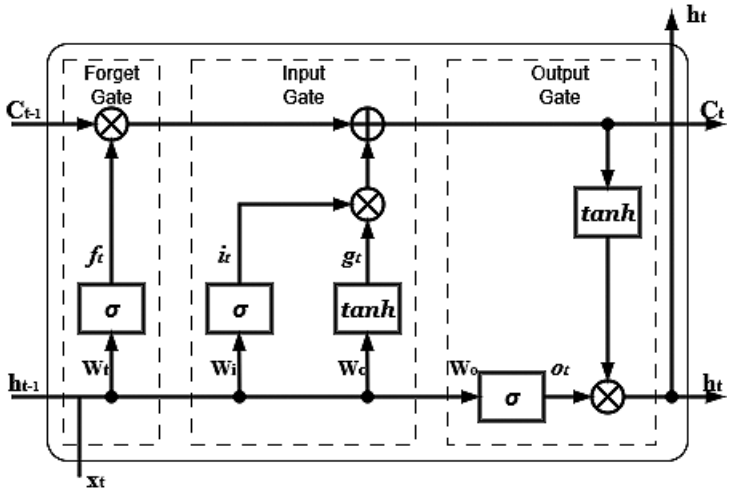

Figure 2. Architecture of an LSTM cell.

As illustrated in Fig. 2, at each time step t, the LSTM cell receives the current input $x_t \in \mathbb{R}^d$ and the previous hidden state $h_{t-1} \in (-1,1)^h$ where d is the feature dimension and h is the number of hidden units. It then updates the cell state $C_t \in \mathbb{R}^h$ and outputs the hidden state h_t. The cell state C_t and hidden state h_t are updated according to (Gers et al., 2000; Yu et al., 2023):

$$\begin{aligned} i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i), \\ f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f), \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o), \\ g_t &= \tanh\left(W_g[h_{t-1}, x_t] + b_g\right), \\ C_t &= f_t \odot C_{t-1} + i_t \odot g_t, \\ h_t &= o_t \odot \tanh(C_t), \end{aligned} \tag{1}$$

where i_t, f_t, and o_t denote the input, forget, and output gates, respectively; W_i, W_f, W_o, W_g denote weight matrices; b_i, b_f, b_o, b_g denote bias vectors; σ represents the sigmoid activation function; and $\odot$ denotes element-wise multiplication.

Through this gating mechanism, LSTM networks effectively learn temporal dependencies across multiple time scales, making them well suited for sequential patterns in geophysical signals such as polar motion. However, LSTM models primarily emphasize local temporal dynamics. To further enhance the ability of the network to capture long-range relationships across the sequence, a self-attention mechanism is incorporated, as described in the following subsection.

2.3 Attention Mechanism

To enhance the model's ability to capture long-range temporal dependencies, a multi-head self-attention mechanism is applied to the sequence output generated by the second LSTM layer. This module allows the model to dynamically evaluate the relevance between different time steps in the sequence, thereby enriching the contextual representation used for prediction.

The self-attention mechanism computes a weighted representation of the input sequence, where each time step selectively attends to others according to learned relevance scores. Unlike traditional sequential models,

self-attention enables flexible, global feature interactions, while multi-head attention extends this capability by learning multiple dependency subspaces simultaneously.

Given an input sequence $X \in \mathbb{R}^{T \times d_{model}}$, where T denotes the sequence length and d_{model} is the feature dimension, the sequence is first linearly projected into query, key, and value matrices using trainable weights W^Q, W^K, W^V. For the i-th attention head, the projections are defined as:

$$Q_i = XW_i^Q, \qquad K_i = XW_i^K, \qquad V_i = XW_i^V. \tag{2}$$

The scaled dot-product attention is then computed by:

$$\text{Attention}_i = \text{softmax}\left(\frac{Q_i K_i^\top}{\sqrt{d_k}}\right) V_i \tag{3}$$

where d_k denotes the dimensionality of the key vectors.

The outputs from all attention heads are concatenated and linearly transformed by W^Q:

$$\text{MultiHead}(X) = \text{Concat}(head_1, \ldots, head_h) W^O \tag{4}$$

where h denotes the number of attention heads.

To stabilize the learning process and preserve the original sequential representation, a residual connection is applied by adding the original input sequence to the attention output, followed by layer normalization:

$$Y = \text{LayerNorm}\big(X + \text{MultiHead}(X)\big). \tag{5}$$

The attention module enriches the representation learned by the LSTM layers by enabling flexible interactions among all time steps in the sequence. In combination with the LSTM backbone, this hybrid structure allows the model to capture both local sequential dynamics and broader temporal dependencies in polar motion signals.

2.4 Implementation details

The proposed model is implemented using the TensorFlow 2.20 deep learning framework (Abadi et al., 2016). Model parameters are optimized using the Adam optimizer with a learning rate of 4×10^{-4}. The batch size is set to 256 and the maximum number of training epochs is 50.

To prevent overfitting and improve training stability, an early stopping strategy is employed. Training is terminated if the monitored validation metric does not improve for five consecutive epochs, and the model weights corresponding to the best validation performance are restored.

Dropout regularization is applied to improve model generalization. Specifically, dropout rates of 0.3 are used after the first two LSTM layers, while a dropout rate of 0.2 is applied in the subsequent stage of the network.

The final Dense layer uses a linear activation function to produce multi-step regression outputs for polar motion forecasting.

All architectural parameters and training configurations necessary for reproducing the proposed model are described in this section, while the detailed descriptions of the dataset, preprocessing procedures, and experimental configurations are provided in Section 3.

3. Experiment

3.1 Model Training

3.1.1 Data Processing

To evaluate the performance of the proposed model, we conducted experiments on the IERS C04 polar motion dataset under multiple hyperparameter configurations. The specific version used in this study is C04 20. The C04 20 series, released by the IERS, is based on a more comprehensive observational dataset and is aligned with ITRF 2020. Compared with earlier releases such as IERS 14 C04, C04 20 benefits from updated reprocessing strategies and improved data handling, providing a more accurate and consistent EOP solution. Considering these advantages, all experiments in this study are conducted using the C04 20 dataset to ensure that the evaluation of the proposed model is based on the most up-to-date and comprehensive EOP series currently available.

The dataset spans 1962-2026 with a 1-day sampling interval. It was chronologically divided into three subsets: the first 44 years were used for training, the following 9 years for validation, and the most recent 10 years for testing. No noise removal, interpolation, or outlier filtering was applied, as the C04 series is generally considered to provide high-quality and reliable polar motion observations without requiring additional preprocessing. After splitting, each subset was normalized independently using the MinMax method to facilitate model convergence.

To construct the model inputs, a sliding-window approach was used to extract fixed-length sequences. Each sample consisted of a window of length(*lookback* + *steps*), where the first *lookback* days served as the input and the subsequent *steps* days served as the prediction targets. The window was then shifted forward by one day to generate the next sample, maximizing the utilization of the continuous daily series. For example, when using 600 historical days to predict the following 400 days, a training segment of 18,993 days yielded 17,994 samples. The same procedure was applied to the validation and test segments. This sliding-window strategy exposed the model to a wide range of history-future pairs and enhanced its ability to learn temporally shift-invariant patterns. During testing, the same mechanism was used to assess multi-step forecasting performance under a fixed historical window.

3.1.2 Input Representation and Normalization

The model takes as input the differenced sequences of the most recent observations, Δ_x and Δ_y, and predicts the incremental differenced sequences relative to the latest observations. The loss function directly quantifies the discrepancy between the predicted differenced values and the true differenced values. Since differencing reduces the magnitude of the data, to accelerate model convergence without compromising data symmetry, Min-Max normalization is applied to the differenced sequences, linearly scaling x_{pole} and y_{pole} to the range $[-1,1]$.

Formally, let $\text{input} \in \mathbb{R}^{\text{lookback}+1\times 2}$ denote the original polar motion series used as input, and $\text{output} \in \mathbb{R}^{\text{steps}\times 2}$ denote the future series to be predicted. The differenced input and output sequences are then defined as:

$$\Delta\text{input}_t = \text{input}_t - \text{input}_{t-1}, \qquad t = 1, \ldots, \text{lookback} \tag{6}$$

$$\text{output}_t = \begin{cases} \text{input}_{\text{lookback}} + \Delta\text{output}_1, & t = 1 \\ \text{output}_{t-1} + \Delta\text{output}_t & t = 2, \ldots, \text{steps} \end{cases} \tag{7}$$

The model is trained to predict Δoutput_t given Δinput_t as input. For instance, in one experimental configuration, the model takes a sequence of 800 days as input; after differencing, this sequence serves as the model input, which is then used to predict the subsequent 400 days of x_{pole} and y_{pole}, yielding 400 predicted values for each component.

3.1.3 Loss Functions

Several loss functions were considered to supervise the training process, including the Mean Absolute Error (MAE), Mean Squared Error (MSE), Pearson Correlation Coefficient (PCC), and Integrated Absolute Error (IAE), as well as their combinations.

Formally, let $\mathbf{y}_t = (x_t, y_t)$ denote the ground truth differenced polar motion at time step t, and $\hat{\mathbf{y}}_t = (\hat{x}_t, \hat{y}_t)$ the corresponding model prediction. The commonly used loss functions are defined as follows:

- **Mean Absolute Error (MAE)**:

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T} \| \mathbf{y}_t - \hat{\mathbf{y}}_t \|_1 = \frac{1}{T}\sum_{t=1}^{T} (|x_t - \hat{x}_t| + |y_t - \hat{y}_t|) \tag{8}$$

- **Mean Squared Error (MSE)**:

$$\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T} \| \mathbf{y}_t - \hat{\mathbf{y}}_t \|_2^2 = \frac{1}{T}\sum_{t=1}^{T} ((x_t - \hat{x}_t)^2 + (y_t - \hat{y}_t)^2) \tag{9}$$

- **Pearson Correlation Coefficient (PCC) Loss**:

$$\mathrm{PCC} = 1 - \frac{\sum_{t=1}^{T}(x_t - \bar{x})(\hat{x}_t - \bar{\hat{x}}) + (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 + (y_t - \bar{y})^2}\sqrt{\sum_{t=1}^{T}(\hat{x}_t - \bar{\hat{x}})^2 + (\hat{y}_t - \bar{\hat{y}})^2}} \tag{10}$$

  where $\bar{x}, \bar{y}, \bar{\hat{x}}, \bar{\hat{y}}$ denote the temporal means of the respective sequences.

- **Normalized Integrated Absolute Error (NIAE)**: This metric evaluates the cumulative error over time, particularly useful for sequences:

$$\mathrm{NIAE} = \frac{1}{T}\sum_{t=1}^{T} \left\| \sum_{k=1}^{t} (\mathbf{y}_k - \hat{\mathbf{y}}_k) \right\|_1 \nparallel \frac{1}{T}\sum_{t=1}^{T} \left(\left| \sum_{k=1}^{t} \| (x_k - \hat{x}_k) \right| + \left| \sum_{k=1}^{t} (y_k - \hat{y}_k) \right| \right) \tag{11}$$

In this work, a total of 4 loss functions were employed, with their naming conventions and specific definitions summarized as follows:

- **mae**: Mean Absolute Error (MAE).
- **mse**: Mean Squared Error (MSE).
- **mse-corr**: A composite loss defined as

$$\mathcal{L}_{\text{mse-corr}} = \mathrm{MSE} + 0.001 \times \mathrm{PCC}.$$

- **mse-int**: A composite loss defined as

$$\mathcal{L}_{\text{mse-int}} = \mathrm{MSE} + 0.002 \times \mathrm{NIAE}.$$

The weighting factors (0.001 for PCC and 0.002 for NIAE) were selected based on the typical numerical scale of the MSE term. Their magnitudes are chosen to ensure that MSE remains the dominant optimization target during training, while the auxiliary terms provide meaningful but non-overwhelming regularization effects. This design allows the composite losses to improve structural correlations (PCC) or interval consistency (NIAE) without destabilizing the primary error minimization process.

For all 4 loss functions, a learning rate of 4×10^{-4} was adopted during training. The performance of the resulting models is reported in Section 3.2.

3.1.4 Input-Output Configuration and Training

To comprehensively evaluate the polar motion (PM) prediction performance of the proposed model, we trained and tested the model under different configurations of historical input length (*lookback*) and forecasting horizon (*steps*). The value of *lookback* ranged from 600 to 1400 days, while *steps* was set to no more than the corresponding *lookback* length, with values ranging from 100 to 1100 days (about 3 years). The detailed parameter settings and the corresponding experimental results are presented in Section 3.2.

Model training was performed using the TensorFlow framework (Abadi et al., 2016), with the Adam optimizer configured at a default learning rate of 4e-4 and a batch size of 256. Model weights were randomly initialized prior to training. During training, a maximum of 50 epochs was allowed for each configuration, however the actual training converged much earlier, with an average of approximately 26 effective epochs due to the early stopping strategy. Early stopping monitored validation performance and terminated training if no improvement was observed for 5 consecutive epochs, after which the weights were restored to those corresponding to the best epoch.

To improve generalization, we applied dropout regularization throughout the model architecture: a dropout rate of 0.3 was added to each of the two LSTM layers in the front stage, and a dropout rate of 0.2 was applied to the attention module. The final Dense output layer adopted a linear activation, producing direct multi-step regression outputs for polar motion forecasting.

3.2 Model Testing

3.2.1 Evaluation Metrics

To quantitatively evaluate the accuracy of the predicted PM given by the model, three commonly used statistical metrics are employed: *Mean Absolute Error (MAE)*, the *Pearson Correlation Coefficient (Corrcoef)*, and an additional metric named *Tolerance Accuracy (TA)*. Suppose the ground truth sequence is denoted as $\{y_i\}_{i=1}^N$ and the corresponding prediction as $\{\hat{y}_i\}_{i=1}^N$, where N is the sequence length.

1) **Mean Absolute Error (MAE)**. MAE measures the average magnitude of prediction errors without considering their direction. It is defined as

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|. \tag{12}$$

This metric provides an intuitive interpretation of the average deviation between the predicted and observed values. Since the original data are expressed in milliarcseconds (mas), the unit of MAE is also mas.

2) **Pearson Correlation Coefficient (Corrcoef)**. The correlation coefficient quantifies the linear relationship between the predicted and observed sequences:

$$r = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}, \tag{13}$$

where $\bar{\hat{y}}$ is the mean of the predicted values. The coefficient ranges from –1 to 1, with values close to 1 indicating strong positive correlation. Corrcoef is also dimensionless.

3) **Tolerance Accuracy (TA)**. To further assess the practical reliability of predictions, the Tolerance Accuracy quantifies the proportion of predicted values whose absolute errors fall within a predefined tolerance range. Let $\Delta = \max(y_i) - \min(y_i)$ denote the overall range of all historical PM data and set the tolerance threshold as $\tau = 0.1\Delta$ (i.e., 10% of the total variation). Then the metric is computed as

$$TA = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(|y_i - \hat{y}_i| \le \tau\right), \tag{14}$$

where $\mathbb{I}(\cdot)$ is the indicator function that equals 1 if the condition is true and 0 otherwise. This metric represents the ratio of time steps where the prediction error is acceptably small, thus providing an intuitive measure of the

model's stability and practical usefulness. The value of TA ranges from 0% to 100%, and higher values indicate better performance within the tolerance band.

Together, these metrics provide complementary perspectives on model performance: MAE reflects the absolute magnitude of prediction errors, Corrcoef measures the consistency and linear association between predicted and observed time series, while TA evaluates how frequently the model's predictions remain within a practically acceptable error margin.

3.2.2 Evaluation Baseline

To comprehensively evaluate the performance of the proposed long-term PM prediction model, two representative baseline methods were employed for comparative analysis. These baselines reflect both a classical geodetic prediction paradigm and an operational reference routinely used in practical PM-related applications.

(1) Least Squares + Autoregressive (LS+AR) Model

The first baseline is a hybrid Least Squares + Autoregressive (LS+AR) model, which has long been used in geodetic time-series prediction. This model includes both a deterministic component and a stochastic component.

The LS component fits long-term variations of PM using a combination of linear trend terms and harmonic signals, including the annual term, semiannual term, and the Chandler wobble. The model parameters were independently estimated using a 2000-day PM history immediately preceding the test set. This ensures consistent historical context between the baseline and the proposed model and avoids any potential data leakage.

The LS component is defined as:

$$PM = a_0 + a_1 \cdot t + a_2 \cdot \cos\left(\frac{2\pi t}{P_A} + \phi_A\right) + a_3 \cdot \cos\left(\frac{2\pi t}{P_C} + \phi_C\right) + x_t \tag{15}$$

Here:

- a_0, a_1 describe the bias and linear trend of PM.
- P_A denotes the annual period, and P_C the Chandler period.
- ϕ_A, ϕ_C represent the phase parameters of the respective harmonic components.
- x_t denotes residuals not captured by the deterministic model.

This formulation reflects the well-established assumption that PM can be decomposed into long-term smooth behavior (trend + periodic terms) plus short-term stochastic variations.

Residuals from the LS fit are further modeled using an autoregressive (AR) process that captures short-term temporal dependencies:

$$x_t = \sum_{i=1}^{p} b_i \cdot x_{t-i} + \varepsilon_t \tag{16}$$

where the AR order p is selected by minimizing the Akaike Information Criterion (AIC) (Akaike, 1974). The AR component effectively models the stochastic variability in PM that cannot be explained by deterministic harmonic and trend terms.

Together, the LS + AR baseline represents traditional statistical prediction approaches and serves as a strong classical benchmark for evaluating the long-horizon forecasting capabilities of the proposed deep learning model.

(2) IERS Bulletin A Operational Predictions

The second baseline utilizes the IERS Bulletin A PM prediction series. Bulletin A provides weekly operational Earth orientation parameter (EOP) forecasts issued by the IERS Rapid Service and Prediction Center and is widely used in navigation, satellite tracking, precise orbit determination, and maintenance of global reference frames.

For this study, historical Bulletin A files corresponding exactly to the test dates were collected. Only the predictions available at their original time of issuance were used, ensuring that comparisons reflect real operational performance rather than retrospectively improved solutions.

Bulletin A routinely publishes polar motion predictions with horizons of up to 1 year (≈365 days). Although this prediction span is shorter than the longest forecasting horizon considered in this study, several models in our framework operate within a comparable range. Therefore, Bulletin A remains an appropriate and meaningful baseline for short- to mid-term comparison, enabling a fair assessment of how the proposed method performs relative to an authoritative operational product.

(3) LSTM-only Model

To evaluate the contribution of the attention mechanism introduced in the proposed model, an ablation baseline denoted as LSTM-only was constructed. This baseline employs the same network architecture as the proposed model but removes the attention module. Consequently, the model consists solely of stacked Long Short-Term Memory (LSTM) layers followed by the output layer.

To ensure a fair comparison, the LSTM-only model was trained using exactly the same training dataset, preprocessing procedure, hyperparameter settings, and optimization strategy as the proposed LSTM + attention model. The only modification is the removal of the attention mechanism that reweights temporal features in the original architecture.

This design allows the experiment to isolate the impact of the attention module on long-horizon PM prediction. By comparing the predictive performance of the LSTM-only model with that of the full model, the effectiveness of the attention mechanism in capturing long-term dependencies and improving forecasting accuracy can be quantitatively assessed.

(4) 2nd EOP PCC Benchmark Predictions

To further position the proposed model within the broader research landscape, results from the Second Earth Orientation Parameters Prediction Comparison Campaign (2nd EOP PCC) were incorporated as an additional benchmark. The 2nd EOP PCC was an international community effort that systematically compared the performance of numerous EOP prediction algorithms submitted by different research groups (Sliwinska et al., 2023).

The campaign provides an extensive collection of prediction series produced by various institutions using diverse modeling strategies. Most prediction horizons in the campaign focus on short- and medium-term forecasting, with more than two-thirds of the submissions covering prediction spans shorter than 100 days and the maximum horizon reaching 366 days.

In this study, the original prediction series released by the campaign were used. For each prediction submission, the predicted PM values were compared with the corresponding IERS C04 20 reference series, and prediction errors were calculated. The mean absolute error (MAE) was then computed to maintain consistency with the evaluation metrics used in this study.

Based on the statistical analysis of the campaign results, two prediction series demonstrating the best performance in representative prediction ranges were selected as reference baselines: prediction C_116, which achieved strong performance in short-term forecasts (10-30 days), and prediction C_138, which performed best for medium-term forecasts up to 365 days. The prediction errors of these two submissions are therefore used as benchmark baselines in the evaluation experiments.

3.2.3 Model Performance

To ensure a fair and unbiased evaluation across different prediction methods, careful consideration was given to the selection of test data. As described in Section 3.1, the most recent portion of the data was reserved exclusively for testing to ensure a strict temporal separation between the training and testing phases, thereby preventing any potential data leakage. Although the full test set spans approximately 10 years, the latter portion of this interval is too distant from the model's training period and from the known values available to classical predictors. Using such far-into-the-future segments would disproportionately penalize methods that rely on shorter, data-proximal histories, especially the linear baselines and Bulletin A forecasts. Therefore, for cross-method comparison, we use only the earliest portion of the test set, where no information leakage occurs and all predictors operate under comparable conditions.

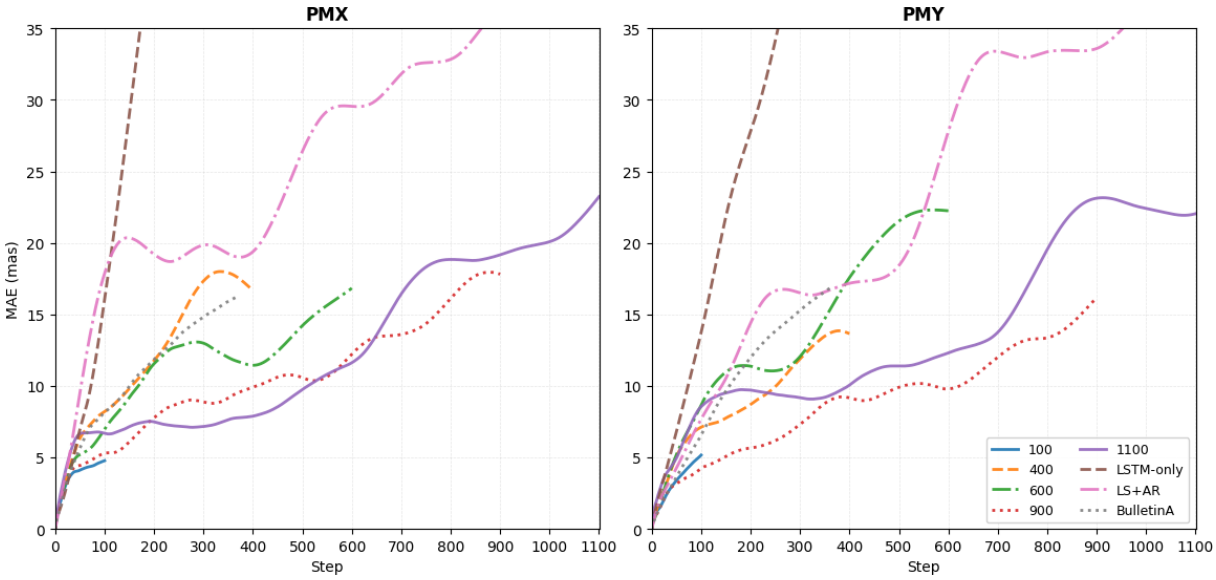

Figure 3. MAE by days of PMX and PMY predictions by different approaches (mas).

To obtain more statistically reliable conclusions, the evaluation is additionally averaged over multiple prediction windows within the first 100 days of the test period. This strategy reduces the influence of individual-window fluctuations and yields a more robust assessment of long-term prediction stability.

Figure 3 illustrates the mean absolute error (MAE) as a function of prediction horizon, separately for PMX and PMY components. As the forecast length increases, all methods exhibit a gradual rise in MAE, while the rate of degradation varies substantially among models. Table 1 provides the numerical MAE values corresponding to seven representative forecast lengths – 10, 30, 100, 365, 600, 800, and 1100 days – allowing direct comparison among linear models, Bulletin A, and several variants of the proposed LSTM-attention framework. Table 2 summarizes the total polar motion magnitude error, defined as $\sqrt{MAE_{PMX}^2 + MAE_{PMY}^2}$, for the same forecast horizons. For clarity of visualization, the plotting range in Fig. 3 is limited to better highlight the differences among the main methods. As a result, parts of the curves for the LSTM-only ablation baseline and the LS+AR model extend beyond the displayed range at longer prediction horizons. The corresponding numerical values can be found in Tables. In addition, the two baselines derived from the 2nd EOP PCC benchmark predictions are presented only in the tables. Because the prediction submissions in the campaign cover different and partially non-overlapping time intervals, their prediction horizons cannot be aligned with a common horizontal axis, making them unsuitable for direct visualization in Fig. 3.

To facilitate clear comparison among different configurations of the proposed deep learning framework, the models listed in the tables are denoted by concise numeric labels such as *100*, *400*, *600*, *900*, and *1100*. Each number corresponds to a specific configuration of the LSTM-attention architecture, primarily distinguished by the length of the internal integration window used to encode past polar motion sequences before forecasting. Within each forecast horizon, multiple candidate configurations were trained and evaluated, and the model reported in the tables represents the one achieving the lowest mean absolute error (MAE) over the full evaluation windows used for cross-method comparison. This selection strategy ensures that every model included in the analysis reflects its best-performing variant under identical training-testing conditions, thereby enabling a fair and interpretable examination of how different architectural scales influence long-term prediction capability. With this convention established, the following analysis discusses the performance patterns that emerge across the various forecast horizons.

From the tables, several observations can be drawn when examining the prediction results from different perspectives:

Before comparing the proposed models with external baselines, it is instructive to first examine the behavior of the LSTM-only ablation model. Although this model exhibits competitive performance at very short horizons, its prediction error increases sharply once the forecast length exceeds roughly 100 days. This rapid degradation highlights the limited capacity of a plain LSTM to preserve long-range temporal dependencies when extrapolating

Table 1. MAE of PMX and PMY predictions by different approaches.

MAE (mas)			@10d	@30d	@100d	@365d	@600d	@800d	@1100d
PMX	baseline	LS+AR	1.8	5.3	17.9	19.0	29.5	32.8	35.4
		Bulletin A	1.5	3.9	8.1	16.2	—	—	—
		LSTM-only	1.4	3.9	16.1	52.4	61.5	—	—
		2ndEOPPCC-116	0.8	3.1	7.7	—	—	—	—
		2ndEOPPCC-138	0.8	2.8	8.2	14.9	—	—	—
	ours	100	1.8	3.7	4.8	—	—	—	—
		400	1.9	4.4	8.2	17.7	—	—	—
		600	1.8	4.3	7.0	11.8	16.8	—	—
		900	1.7	3.8	5.3	9.5	12.2	16.1	—
		1100	2.3	5.3	6.7	7.8	11.6	18.8	23.2
PMY	baseline	LS+AR	1.3	2.9	7.7	16.8	27.8	33.4	38.2
		Bulletin A	1.0	2.4	6.6	17.0	—	—	—
		LSTM-only	1.7	4.2	13.7	49.1	58.0	—	—
		2ndEOPPCC-116	0.5	1.6	6.1	—	—	—	—
		2ndEOPPCC-138	0.7	2.0	8.8	17.1	—	—	—
	ours	100	1.2	2.4	5.2	—	—	—	—
		400	1.4	3.2	7.1	13.8	—	—	—
		600	1.1	3.3	8.7	15.6	22.2	—	—
		900	1.0	2.5	4.2	9.1	9.8	13.4	—
		1100	2.0	3.9	8.6	9.4	12.3	19.5	22.0

far beyond the observed sequence. In contrast, incorporating the multi-head attention mechanism substantially mitigates this limitation by enabling the model to selectively emphasize informative temporal patterns across extended time spans. The marked performance gap between the LSTM-only baseline and the full LSTM-attention framework thus underscores the critical role of the attention module in achieving stable and accurate long-term PM predictions.

Table 2. Total polar motion prediction magnitude error by different approaches.

Total PM Error (mas)		@10d	@30d	@100d	@365d	@600d	@800d	@1100d
baseline	LS+AR	2.2	6.1	19.5	25.4	40.6	46.8	52.1
	Bulletin A	1.8	4.6	10.5	23.4	—	—	—
	LSTM-only	2.2	5.8	21.1	71.9	84.5	—	—
	2ndEOPPCC-116	**0.9**	3.5	9.8	—	—	—	—
	2ndEOPPCC-138	1.1	**3.4**	12.0	22.7	—	—	—
ours	100	2.2	4.4	7.0	—	—	—	—
	400	2.3	5.5	10.9	22.4	—	—	—
	600	2.1	5.4	11.1	19.5	27.9	—	—
	900	2.0	4.6	**6.8**	13.1	**15.7**	**20.9**	—
	1100	3.1	6.6	10.9	**12.2**	16.9	27.1	**32.0**

At short forecast horizons, particularly within the first 10-30 days, the benchmark results from the 2nd EOP PCC campaign show very strong performance. In this range, the best-performing submissions in the campaign (C_116 and C_138) achieve slightly lower errors than the proposed model. This outcome is expected, as many of the algorithms developed for the campaign were specifically optimized for short-term prediction and are highly tuned to capture immediate temporal dynamics. Nevertheless, the proposed LSTM-attention models remain competitive in this regime and already outperform the classical LS+AR baseline and the operational Bulletin A predictions.

As the forecast horizon increases, the advantage of the proposed method becomes more evident. Around the 100-day horizon, the LSTM-attention models begin to surpass the best results reported in the 2nd EOP PCC benchmark. This improvement becomes even more pronounced near the one-year prediction range. Although Bulletin A reaches its maximum official horizon at approximately 365 days and provides a strong operational reference, the deep learning approaches still exhibit a clear advantage, with the 900 configuration achieving the smallest total error. These results suggest that the proposed architecture not only maintains competitive short-term accuracy but also learns the multi-frequency structure of polar motion more effectively, enabling improved prediction performance at medium and extended forecast horizons.

As the forecast horizon extends into the medium- and long-term range (600-800 days), the performance of classical methods deteriorates more rapidly. In contrast, several variants of the proposed model maintain relatively stable error growth, avoiding the divergence typically observed in linear regressors. Around 800 days, the *900* configuration yields the best performance among all candidates, balancing long-term stability with the preservation of meaningful temporal correlations. Although the error increase is inevitable over such extended horizons, the deep learning approach continues to demonstrate a smoother and more consistent trajectory than *LS+AR*, whose errors accelerate substantially with time.

At the extreme limit of 1100 days, only a subset of the proposed models can produce predictions, as Bulletin A and some intermediate configurations do not extend to this horizon. The *1100* variant is the only model capable of sustaining forecasts at this length, albeit with noticeably larger errors compared to its own performance at shorter horizons. This behavior reflects the inherent difficulty of ultra-long-term polar motion forecasting and underscores the challenge of maintaining accuracy when extrapolating far beyond the temporal region covered by observed data.

Nevertheless, the model remains numerically stable over the entire three-year range, illustrating the robustness of the LSTM-attention architecture.

In summary, the performance across different horizons reveals distinct strengths among the model variants. At the one-year horizon, the configuration with a 400-length intermediate integration window (*400*) achieves the best results. Around the 800-day horizon, the model with a 900-day window (*900*) provides the strongest performance, maintaining comparatively low errors under long forecast conditions. At the full 1100-day extent, only the model explicitly trained for this extreme horizon (*1100*) capable of producing predictions, albeit with higher error levels. These findings collectively confirm that the proposed methodology delivers state-of-the-art accuracy at short and medium ranges, while also enabling the longest stable PM predictions currently achievable with data-driven approaches.

3.2.4 Model Analysis

For the evaluation between models, we consistently used 10 years of polar motion observations spanning from January 2016 to January 2026 as the test dataset. During testing, the same sliding-window preprocessing procedure employed during training was applied. To ensure reliable evaluation, each model's performance was assessed by averaging the metrics across multiple test windows, and the final reported values represent the mean performance over the entire test dataset.

The experimental results are illustrated in Figs. 4, 5 and 6. In these figures, the horizontal axis "steps" denotes the prediction horizon, while the vertical axis "lookback" represents the length of historical input used by the model. The corresponding evaluation metrics are visualized using a color scale. For each configuration trained under different loss functions, the models were evaluated under identical conditions, and their performance is displayed in the same figure for direct comparison.

From the heatmaps in Fig. 4, which present the MAE and Corrcoef evaluation results of different models under various *lookback*, *steps*, and loss function configurations, several key observations can be made:

- **First**, for a fixed lookback length, prediction accuracy generally decreases as the forecasting horizon increases, reflecting the accumulation of uncertainty over longer prediction intervals.
- **Second**, for a fixed prediction horizon, models trained with longer historical input windows do not always achieve higher accuracy. This suggests that while incorporating extended temporal context can better capture the underlying dynamics of polar motion, it also increases model complexity, which may offset potential performance gains.
- **Third**, overall model performance tends to degrade when both the lookback and forecasting steps become large, indicating that simultaneously extending the input and output sequences makes the prediction task more challenging. As observed from the trend of MAE with increasing steps, the resulting prediction errors become unacceptably large beyond a step length of 1100; therefore, larger steps are excluded from our experiments.
- **Fourth**, models trained with *mse-int* loss achieve the best overall performance, whereas models trained with *mse-corr* loss perform well under most configurations but fail when both lookback and steps are large. Models trained with *mae* loss exhibit the poorest performance as the prediction horizon increases.
- **Fifth**, the MAE and Corrcoef metrics are highly correlated across models, indicating that models unable to capture the correct trend of polar motion also tend to struggle in achieving low absolute errors.

Overall, these observations highlight the trade-off between temporal context length, forecasting horizon, and model complexity in achieving accurate and reliable polar motion prediction.

From the heatmaps in Fig. 5, which present the MAE evaluation results of PMX and PMY separately across different models, we find that the MAE values for PMX and PMY are slightly different but generally comparable. The absence of a significant discrepancy between the two indicates that the models achieve similar accuracy in predicting both PMX and PMY components of the polar motion. This consistency can be attributed to the model's input-output design, where each sample simultaneously includes the two-dimensional data of PMX and PMY for the same day.

From the heatmaps in Fig. 6, which present the Tolerance Accuracy (TA) evaluation results of PMX and PMY separately across different models, several key observations can be made:

- **First**, the TA metric exhibits trends consistent with those of MAE, indicating that the practical usefulness of the models can also be effectively reflected by their MAE values.

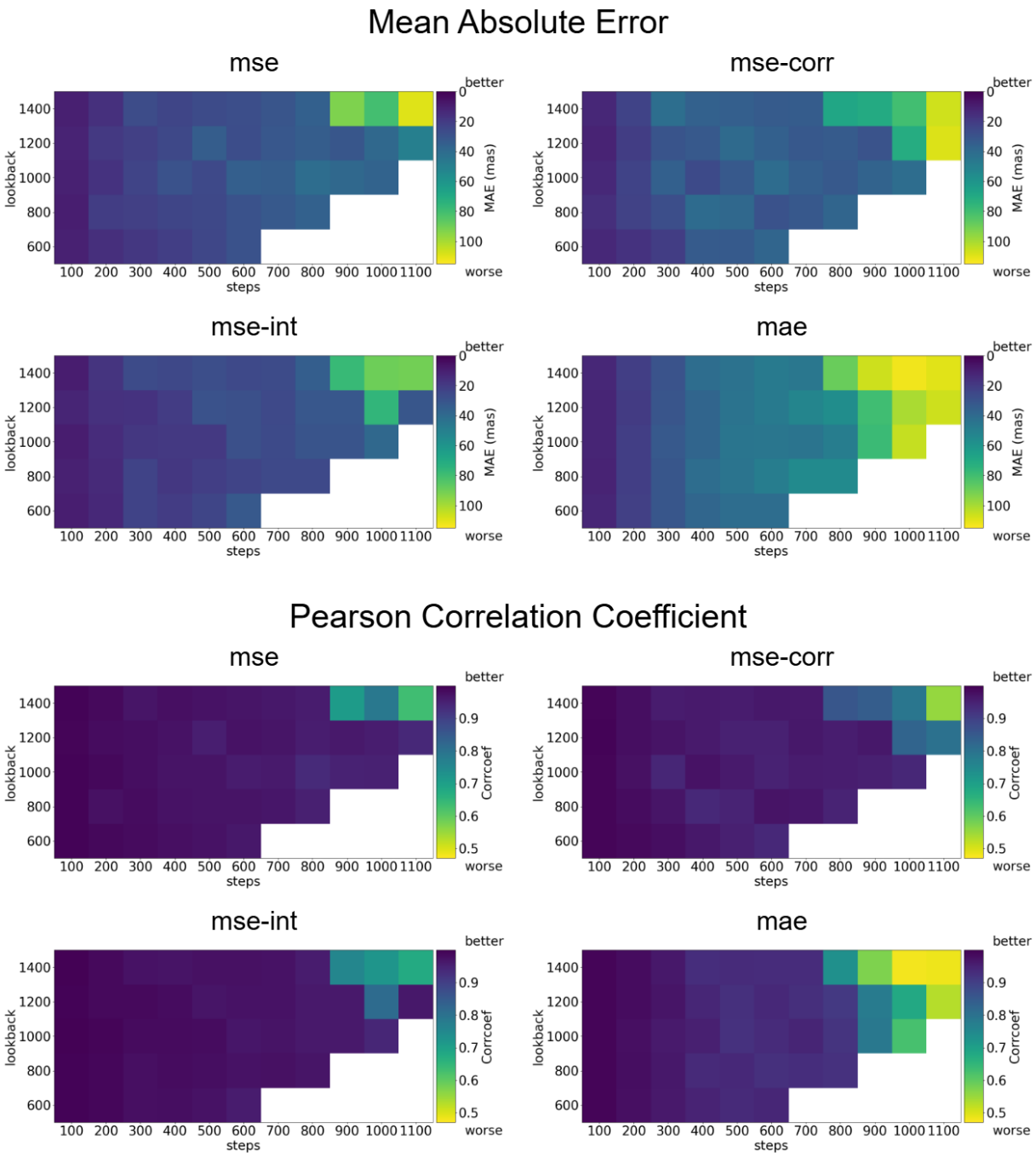

Figure 4. Heatmap of MAE and Corrcoef for Different Models by Lookback and Steps.

- **Second**, unlike other metrics, models trained with the *mse-int* loss slightly outperform those trained with *mse* and *mse-corr* losses on the TA metric when *steps* $\leq$ 800. This suggests that incorporating the integrated error term in the loss function enhances the overall stability of the predicted sequences, even though it does not further reduce the MAE.
- **Third**, quantitative results show that TA reaches 97% and 97% for PMX and PMY at (*lookback* = 1000, *steps* = 400, *loss* = mse-int); 94% and 94% at (*lookback* = 1400, *steps* = 600, *loss* = mse-int); 91% and 89% at (*lookback* = 1200, *steps* = 900, *loss* = mse-corr); and drops to 87% and 87% at (*lookback* = 1200, *steps* = 1100, *loss* = mse-int). These results indicate that the models remain highly reliable for *steps* $\leq$ 600, maintain acceptable accuracy forn *steps* $>$ 600.

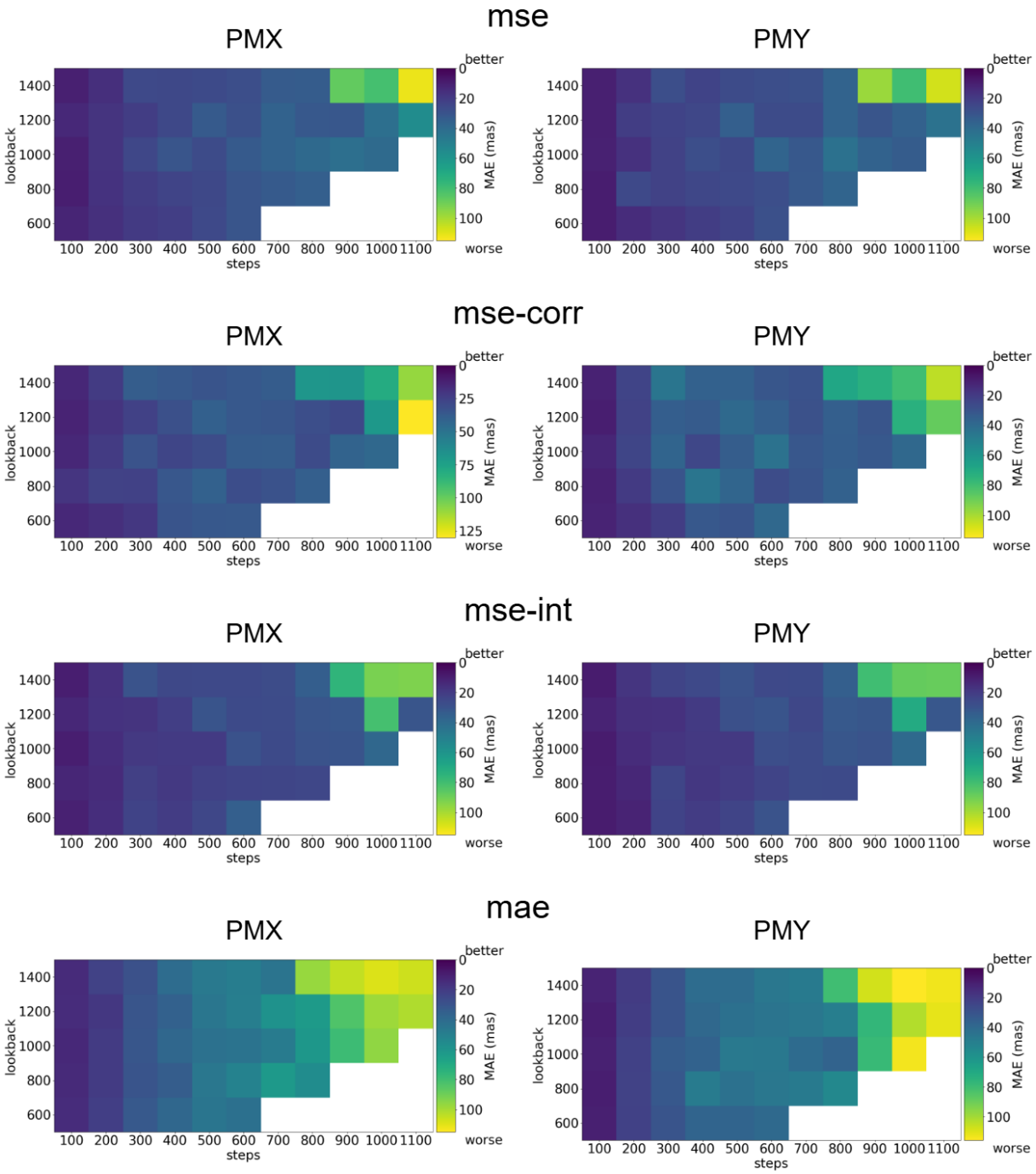

Figure 5. Heatmap of MAE by PMX and PMY for Different Models by lookback and steps.

Overall, the TA metric provides a complementary and practical perspective on model robustness, revealing how predictive reliability degrades as the forecasting horizon increases and highlighting the trade-off between accuracy and stability across different configurations.

To further illustrate the practical effectiveness of the proposed model, Fig. 7 presents a comparison between the recent polar motion observations, our model's future predictions, and the official Bulletin A forecasts. This figure is provided solely for visual demonstration and is not included in the quantitative evaluation. As shown, our model's predictions closely follow the recent trend of polar motion and maintain consistency with the Bulletin A forecasts, demonstrating the model's capability to generate stable and physically reasonable long-term predictions.

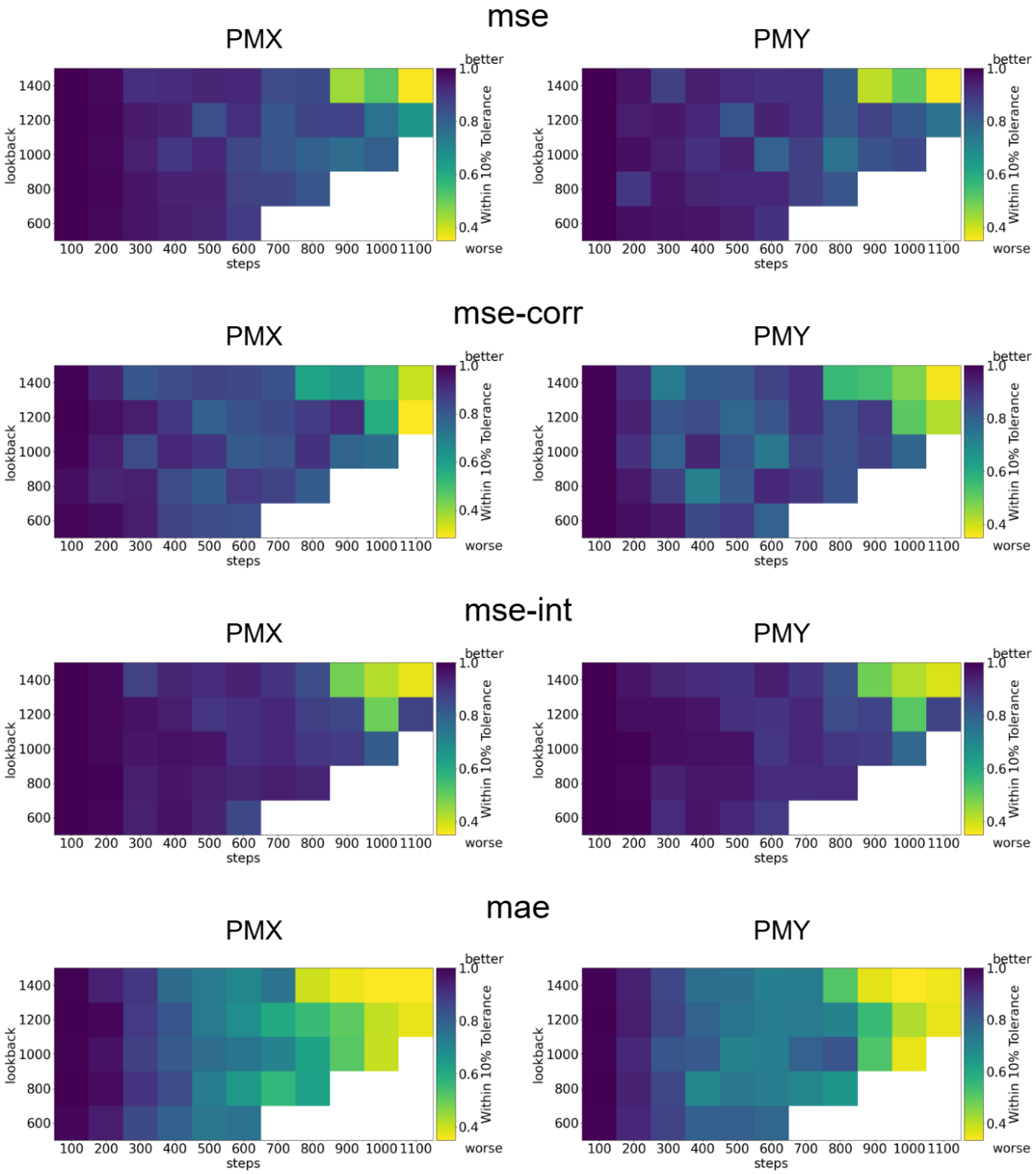

Figure 6. Heatmap of TA for Different Models by lookback and steps.

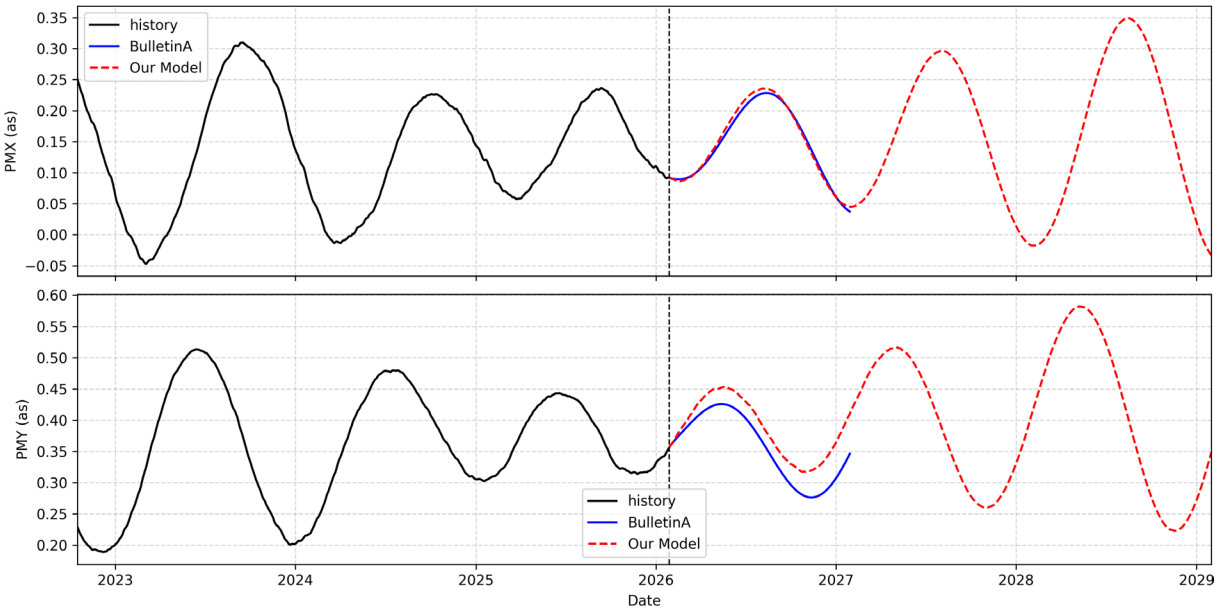

Figure 7. Comparison of recent polar motion history, model prediction, and Bulletin A forecast.

4. Discussion

The results presented in this study demonstrate that the proposed hybrid LSTM-attention model provides robust and accurate long-term forecasts of polar motion (PM), achieving stable performance even for prediction horizons exceeding three years. Compared with traditional linear models such as LS+AR or harmonic extrapolation, the deep learning framework exhibits stronger adaptability to nonlinear temporal dynamics and can capture both short-term fluctuations and long-period oscillations such as the Chandler and annual wobbles. This highlights the advantage of data-driven approaches in representing the complex, multi-scale characteristics of Earth rotation behavior.

From the quantitative evaluation results, it can be observed that prediction accuracy gradually declines as the forecast length increases, consistent with the general tendency observed in PM prediction. However, even at extended horizons (exceeding 1000 days), the model maintains a high correlation with observed values and acceptable error levels, suggesting that the learned representations effectively capture long-range dependencies within the PM series. The multi-head attention mechanism plays a key role in this performance, as it enables the model to dynamically reweight temporal features and extract multi-frequency relationships that conventional LSTM networks alone may fail to represent.

In addition to overall accuracy, the differenced-input design contributes to greater training stability and mitigates the common issue of amplitude damping in long-term forecasts. By modeling incremental variations rather than absolute positions, the network better preserves oscillatory patterns and reduces bias accumulation over extended prediction intervals. However, this differencing strategy may also amplify local noise, implying a trade-off between stability and sensitivity. Future work may further optimize this balancing through hybrid differencing or frequency-domain preprocessing techniques.

The results also indicate that optimizing with composite loss functions enhances temporal consistency in long-range forecasts. Loss formulations that combine both pointwise and cumulative error terms (e.g., MSE+NIAE) improve the smoothness and stability of predictions without compromising accuracy, showing that multi-objective learning can guide the model toward physically plausible trajectories.

Overall, the model effectively reproduces both amplitude and phase characteristics of the true PM series, maintaining realistic periodicity and minimal phase drift within the three-year forecast horizon. This robustness indicates that the network captures intrinsic oscillatory modes of Earth rotation rather than merely fitting short-term trends.

Despite its promising performance, several limitations remain. First, the current framework relies exclusively on historical PM data, without explicit incorporation of geophysical excitations such as atmospheric, oceanic, or hydrological angular momentum. Although this purely data-driven approach ensures generality, integrating

physical drivers could further enhance interpretability and improve response to irregular perturbations. Second, the model assumes temporal stationarity of PM dynamics, even though long-term variations (e.g., secular trends or amplitude modulation of the Chandler wobble) may violate this assumption. Future extensions could adopt adaptive or transfer-learning schemes to better accommodate time-varying behavior. Third, although differencing helps mitigate bias accumulation, error propagation remains a challenge for extremely long-term forecasts. Multi-stage or autoregressive correction mechanisms could be explored to further stabilize iterative predictions.

Another limitation of this study is that prediction uncertainty is not explicitly quantified. The proposed framework focuses on deterministic point forecasts and does not provide confidence intervals or probabilistic estimates, which may limit its applicability in risk-sensitive or operational scenarios. Recent studies have highlighted the importance of uncertainty-aware prediction for Earth rotation parameters, exploring probabilistic and ensemble-based approaches to characterize forecast reliability (Kiani et al., 2023, 2024a, 2024b). These methods provide complementary insights by quantifying the confidence of predictions in addition to their accuracy. Incorporating uncertainty estimation into deep learning-based PM prediction frameworks will be an important direction for future research.

In summary, this study demonstrates the feasibility of applying a deep learning-based system to perform three-year polar motion forecasts with competitive accuracy. The LSTM-attention architecture proves capable of modeling complex, nonlinear, and quasi-periodic geophysical signals, advancing the methodological frontier of Earth rotation prediction. Future work will focus on incorporating physical excitations, exploring transformer architectures, and developing adaptive real-time update mechanisms to further enhance the precision and operational applicability of data-driven Earth rotation forecasts.

5. Conclusion

In this study, we proposed a deep learning framework for long-term polar motion prediction, combining the sequential learning capability of Long Short-Term Memory (LSTM) networks with the global contextual modeling power of a multi-head attention mechanism. By jointly predicting the PMX and PMY components from differenced time series inputs, the model effectively captures both short-term fluctuations and long-period oscillations in the Earth's pole trajectory.

Experimental evaluations on the IERS C04 dataset show that the proposed LSTM-attention model performs consistently well across a wide range of forecasting horizons. It achieves high accuracy and strong correlation in both short-term and ultra-long-term predictions, with reliable results extending up to 1100 days (approximately three years). Compared with traditional linear models, the proposed framework demonstrates enhanced stability and generalization capability, particularly in its ability to maintain long-term trend fidelity.

The findings confirm that integrating attention mechanisms into recurrent architectures can significantly improve the interpretability and accuracy of geophysical time series modeling. Future work will further explore hybrid approaches that incorporate physical constraints, including excitation information derived from effective angular momentum (EAM), or multi-source geodetic observations, aiming to extend the model's robustness and expanding its applicability within real-time Earth rotation prediction systems.

Data availability statement. The historical polar motion data used in this study, as well as the prediction data adopted as baselines, are obtained from the IERS website and are publicly available at https://datacenter.iers.org/eop.php. The 2nd EOP PCC data utilized in this work is sourced from the GFZ repository at https://datapub.gfz-potsdam.de/download/10.5880.GFZ.1.3.2023.001-Ncehfb/. The python program source code of this work can be found at https://github.com/PM-Prediction-Model/lstm-attention. All models in our research were trained and evaluated with the code.

Acknowledgements. This research was funded by the Spark Program of Earthquake Technology of the China Earthquake Administration (CEA), grant number XH23023YA, and by the Scientific Research Project of the Fujian Earthquake Agency, grant number M202503. The APC was funded by the same projects. During the preparation of this manuscript, the author(s) used GPT-5 and Claude-sonnet-4.5 for the purposes of text optimization and code segment generation. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems, software, https://www.tensorflow.org/.

Akaike, H. (1974). A new look at the statistical model identification, IEEE Trans. Autom. Control, 19, 6, 716-723, doi:10.1109/TAC.1974.1100705.

Akulenko, L. D., S. A. Kumakshev, Y. G. Markov and L. V. Rykhlova (2002). Forecasting the polar motions of the deformable Earth, Astron. Rep., 46, 10, 858-865, doi:10.1134/1.1515097.

Bruni, S., E. Schoenemann, V. Mayer, M. Otten et al. (2021). ESA's Earth Orientation Parameter product, in Proc. EGU General Assembly 2021, Copernicus, doi:10.5194/egusphere-egu21-12989.

Byram, S. and C. Hackman (2012). High-precision GNSS orbit, clock and EOP estimation at the United States Naval Observatory, in Proc. IEEE/ION PLANS 2012, 659-663, doi:10.1109/PLANS.2012.6236940.

Coulot, D., A. Pollet, X. Collilieux and P. Berio (2010). Global optimization of core station networks for space geodesy: Application to the referencing of the SLR EOP with respect to ITRF, J. Geod., 84, 1, 31-50, doi:10.1007/s00190-009-0342-1.

Dill, R., H. Dobslaw and M. Thomas (2019). Improved 90-day Earth orientation predictions from angular momentum forecasts of atmosphere, ocean, and terrestrial hydrosphere, J. Geod., 93, 3, 287-295, doi:10.1007/s00190-018-1158-7.

Dill, R., L. Stumpe, J. Saynisch-Wagner, M. Thomas et al. (2025). Benefits of refined 10-day effective angular momentum forecasts for Earth rotation parameter prediction, J. Geod., 99, 2, 15, doi:10.1007/s00190-025-01941-x.

Dobslaw, H. and R. Dill (2018). Predicting Earth orientation changes from global forecasts of atmosphere-hydrosphere dynamics, Adv. Space Res., 61, 4, 1047-1054, doi:10.1016/j.asr.2017.11.044.

Gers, F. A., J. Schmidhuber and F. Cummins (2000). Learning to forget: Continual prediction with LSTM, Neural Comput., 12, 10, 2451-2471, doi:10.1162/089976600300015015.

Gross, R. S., T. M. Eubanks, J. A. Steppe, A. P. Freedman et al. (1998). A Kalman-filter-based approach to combining independent Earth-orientation series, J. Geod., 72, 4, 215-235, doi:10.1007/s001900050162.

Gross, R. S. (2007). Earth rotation variations – long period, in Treatise on Geophysics, Elsevier, 239-294, doi:10.1016/B978-044452748-6.00057-2.

Gross, R. S. (2015). Earth rotation variations – long period, in Treatise on Geophysics, Elsevier, 215-261, doi:10.1016/B978-0-444-53802-4.00059-2.

Jin, X., J. Guo, Y. Shen, X. Liu et al. (2021). Application of singular spectrum analysis and multilayer perceptron in the mid-long-term polar motion prediction, Adv. Space Res., 68, 9, 3562-3573, doi:10.1016/j.asr.2021.06.039.

Kalarus, M., H. Schuh, W. Kosek, O. Akyilmaz et al. (2010). Achievements of the Earth orientation parameters prediction comparison campaign, J. Geod., 84, 10, 587-596, doi:10.1007/s00190-010-0387-1.

Kehm, A., H. Hellmers, M. Bloßfeld, R. Dill et al. (2023). Combination strategy for consistent final, rapid and predicted Earth rotation parameters, J. Geod., 97, 1, 3, doi:10.1007/s00190-022-01695-w.

Kiani Shahvandi, M., S. Adhikari, M. Dumberry, S. Modiri et al. (2024). Contributions of core, mantle and climatological processes to Earth's polar motion, Nat. Geosci., 17, 7, 705-710, doi:10.1038/s41561-024-01478-2.

Kiani Shahvandi, M., R. Dill, H. Dobslaw, A. Kehm et al. (2023). Geophysically informed machine learning for improving rapid estimation and short-term prediction of Earth orientation parameters, J. Geophys. Res., Solid Earth, 128, 10, e2023JB026720, doi:10.1029/2023JB026720.

Kiani Shahvandi, M., S. Mishra and B. Soja (2024). BaHaMAs: A method for uncertainty quantification in geodetic time series and its application in short-term prediction of length of day, Earth Planets Space, 76, 1, 127, doi:10.1186/s40623-024-02066-9.

Kosek, W., D. D. McCarthy and B. J. Luzum (1998). Possible improvement of Earth orientation forecast using autocovariance prediction procedures, J. Geod., 72, 4, 189-199, doi:10.1007/s001900050160.

Kur, T., J. Śliwińska-Bronowicz, M. Wińska, H. Dobslaw et al. (2024). Prospects of predicting the polar motion based on the results of the second Earth orientation parameters prediction comparison campaign, Earth Space Sci., 11, 11, e2023EA003278, doi:10.1029/2023EA003278.

Liao, D. C., Q. J. Wang, Y. H. Zhou, X. H. Liao et al. (2012). Long-term prediction of the Earth orientation parameters by the artificial neural network technique, J. Geodyn., 62, 87-92, doi:10.1016/j.jog.2011.12.004.

Luzum, B. J., J. R. Ray, M. S. Carter and F. J. Josties (2001). Recent improvements to IERS Bulletin A combination and prediction, GPS Solut., 4, 3, 34-40, doi:10.1007/PL00012853.

Niedzielski, T. and W. Kosek (2008). Prediction of UT1-UTC, LOD and AAM χ_3 by combination of least-squares and multivariate stochastic methods, J. Geod., 82, 2, 83-92, doi:10.1007/s00190-007-0158-9.

Niu, X., N. Wei and Y. Zhou (2025). Improved polar motion prediction using a genetic algorithm optimized backpropagation neural network and effective angular momentum functions, Adv. Space Res., doi:10.1016/j.asr.2025.04.065.

Petit, G. and B. Luzum (2010). IERS conventions 2010, IERS Tech. Note 36, IERS.

Schuh, H. and J. Böhm (2013). Very long baseline interferometry for geodesy and astrometry, in Sciences of Geodesy II, Springer, 339-376, doi:10.1007/978-3-642-28000-9_7.

Śliwińska, J., H. Dobslaw, T. Kur, J. Nastula et al. (2023). EOP predictions collected during the operational phase of the second Earth orientation parameters prediction comparison campaign, GFZ Data Serv., doi:10.5880/GFZ.1.3.2023.001.

Śliwińska-Bronowicz, J., T. Kur, M. Wińska, H. Dobslaw et al. (2024). Assessment of length-of-day and universal time predictions based on the results of the second Earth orientation parameters prediction comparison campaign, J. Geod., 98, 3, 22, doi:10.1007/s00190-024-01824-7.

Śliwińska-Bronowicz, J., J. Nastula, T. Kur, H. Dobslaw et al. (2025). IERS Technical Note No. 42: Second Earth orientation parameters prediction comparison campaign, IERS Tech. Note, 42, doi:10.60599/IERS-TN42.

Wang, L., W. Miao and F. Wu (2022). A new polar motion prediction method combined with the difference between polar motion series, Geod. Geodyn., 13, 6, 564-572, doi:10.1016/j.geog.2022.07.001.

Wang, L., W. Miao and F. Wu (2023). A new medium-long term polar motion prediction method based on sliding average within difference series, Meas. Sci. Technol., 34, 10, 105023, doi:10.1088/1361-6501/ace5c1.

Wang, L., H. Que and F. Wu (2025). The CNN-LSTM-attention model for short-term prediction of the polar motion, Meas. Sci. Technol., 36, 1, 016323, doi:10.1088/1361-6501/ad8be5.

Wang, X. Q. (2024). Leveraging the empirical wavelet transform in combination with convolutional LSTM neural networks to enhance the accuracy of polar motion prediction, Res. Astron. Astrophys., 24, 9, doi:10.1088/1674-4527/ad74dd.

Wińska, M., T. Kur, J. Śliwińska-Bronowicz, J. Nastula et al. (2024). Findings on celestial pole offsets predictions in the second Earth orientation parameters prediction comparison campaign, Earth Planets Space, 76, 1, 100, doi:10.1186/s40623-024-02042-3.

Wu, F., K. Deng, G. Chang and Q. Wang (2018). The application of a combination of weighted least-squares and autoregressive methods in predictions of polar motion parameters, Acta Geod. Geophys., 53, 2, 247-257, doi:10.1007/s40328-018-0214-3.

Xu, X. and Y. Zhou (2015). EOP prediction using least square fitting and autoregressive filter over optimized data intervals, Adv. Space Res., 56, 10, 2248-2253, doi:10.1016/j.asr.2015.08.007.

Yao, Y., S. Yue and P. Chen (2013). A new LS+AR model with additional error correction for polar motion forecast, Sci. China Earth Sci., 56, 5, 818-828, doi:10.1007/s11430-012-4572-3.

Yu, K., K. Yang, T. Shen, L. Li et al. (2023). Estimation of Earth rotation parameters and prediction of polar motion using hybrid CNN-LSTM model, Remote Sens., 15, 2, 427, doi:10.3390/rs15020427.

Yu, K., H. Shi, M. Sun, L. Li et al. (2024). Combined BiLSTM and ARIMA models in middle- and long-term polar motion prediction, Stud. Geophys. Geod., 68, 1-2, 25-40, doi:10.1007/s11200-023-0134-y.

***CORRESPONDING AUTHOR: Tianyi CHEN**,
AVIC Changcheng Institute of Metrology and Measurement, Beijing, China
e-mail: chenty031@avic.com