

FULL PAPER

Open Access



# Improving the accuracy of polar motion prediction using a hybrid least squares and long short-term memory model

Chenxiang Wang<sup>1\*</sup> and Pengfei Zhang<sup>1,2</sup>

## Abstract

Earth rotation parameters (ERPs) are essential for transforming between the celestial and terrestrial reference frames, and for high-precision space navigation and positioning. Among the ERPs, polar motion (PM) is a critical parameter for analyzing and understanding the dynamic interaction between the solid Earth, atmosphere, ocean, and other geophysical fluids. Traditional methods for predicting the change in ERPs rely heavily on linear models, such as the least squares (LS) and the autoregressive (AR) model (LS + AR). However, variations in ERP partly reflect non-linear effects in the Earth system, such that the predictive accuracy of linear models is not always optimal. In this paper, long short-term memory (LSTM), a non-linear neural network, is employed to improve the prediction of ERPs. Polar motion prediction experiments in this study are conducted using the LSTM model and a hybrid method LS + LSTM model based on the IERS EOP14C04 time series. Compared with Bulletin A, the PMX and PMY prediction accuracy can reach a maximum of 33.7% and 31.9%, respectively, with the LS + LSTM model. The experimental results show that the proposed hybrid model displays a better performance in mid- and long-term (120–365 days) prediction of polar motion.

**Keywords** Earth rotation parameters, Polar motion, Long short-term memory, Least squares and the autoregressive model, IERS EOP 14C04, Bulletin A, LS + LSTM model

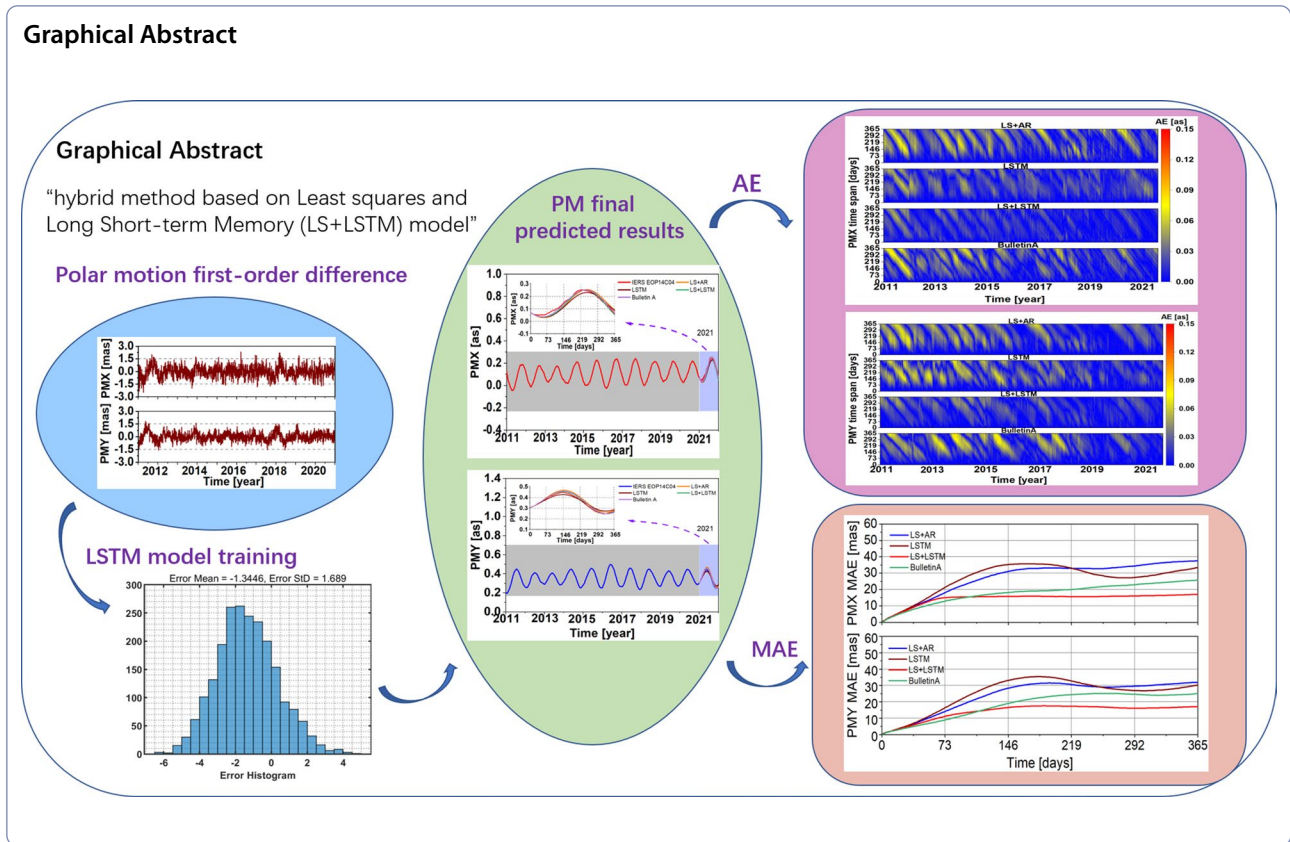
\*Correspondence:

Chenxiang Wang  
wangchenxiang@whu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



**Introduction**

Due to the rapid development of space technology, the accuracy of Earth Orientation Parameter (EOP) estimates dramatically improved in the 1990s (Schuh et al. 2002) and has remained at a high level ever since. The EOPs consist of precession–nutations, polar motion (PM), the difference between Universal Time (UT1) and Coordinated Universal Time (UTC), that is, UT1-UTC, and Length of Day (LOD) (Petit and Luzum 2010). Earth Rotation Parameters (ERPs) comprise the PM (including PMX and PMY), UT1-UTC, and LOD. The instantaneous movement of the Earth’s rotation axis with respect to the terrestrial reference frame is described by the PM. As for PM, a major complication is that it is caused by partially unpredictable mass redistributions on the surface and in the interior of the Earth (Gross 2007; Dobslaw et al. 2010; Sun et al. 2019; Börger et al. 2023). Modern space navigation and deep space exploration are increasingly required for accurate real-time prediction of ERPs. Given the complicated data processing of modern geodesic techniques, such as Global Positioning System (GPS) technology, the acquisition of ERP results must be delayed by 15–20 h. Obtaining ERPs requires several days for the Very Long Baseline Interferometry (VLBI) and Satellite Laser Ranging (SLR) technologies. These factors make it

challenging to acquire real-time ERPs, emphasizing the need for accurate predictions (Zhang et al. 2012). Several national and international services publish predicted values of EOPs, such as the International Earth Rotation and Reference Systems Service (IERS) Rapid Service/Prediction Center (RS/PC), operated by the US Naval Observatory (USNO) (Guo et al. 2013), and published in the IERS Bulletin A files, for a year into the future in the daily interval, or the EOP service of the Institute of Applied Astronomy of Russian Academy of Sciences (IAA RAS) (Suvorkin et al. 2015). The products provided by these agencies comprise estimates for PM, UT1-UTC, LOD, and other parameters, usually for a year into the future at daily sampling.

The polar motion includes a regular deterministic and an irregular stochastic component. The deterministic part consists of the long-term trend, Chandler wobbles (CW) (Chandler 1981; Zharkov and Molodensky 1996), Annual wobbles (AW), and Semi-Annual wobbles (SAW) (Wang et al. 2016; Gross 2000). Chandler Wobble is a resonant rotational mode of the Earth that decays freely due to the viscoelastic nature of the Earth. Studies have shown that CW will freely decay within 68 years to the minimum rotational energy state without excitation. It is generally believed that the oscillation period and

amplitude of CW vary over time, with the period fluctuating between 1.13 and 1.20 years (Schuh et al. 2001). The annual oscillation of the pole curve includes both prograde and retrograde components. The intensity of the prograde part is 10 times that of the retrograde part, and there is a significant change in the period of the forward annual oscillation of the PM, which oscillates between 356 and 376 days (Joachim 2004). Considering the characteristics of secular drift, CW, and AW, scholars have conducted extensive studies and proposed various methods for predicting the ERPs. In General, these methods fall into linear and non-linear models. Kalman Filtering (Babcock and Wilkins 1989), Least Squares (LS) extrapolation, fuzzy interface systems (Akyilmaz and Kutterer 2004), autoregression models (AR) (Sun and Xu 2012), autocovariance models (Kosek 2002), and different combinations of these methods (Kosek and Popiński 2005; Kosek et al. 2004; Kosek et al. 2008) are all linear models. Methods such as threshold autoregression models, artificial neural networks (Liao et al. 2012; Egger 1992), and fuzzy reasoning are non-linear models.

More hybrid and machine learning methods have been introduced in recent years for predicting ERP variations. The rapid expansion in computing power and data volume has made applying deep learning in geodesy increasingly promising. In particular, the long short-term memory (LSTM) network (Hochreiter and Schmidhuber 1997), one of the most popular forms of recurrent neural networks (RNNs), is advantageous for geodetic time series prediction. The LSTM network can capture the non-linear structure between different time epochs in the time series due to the unique structure of its cells (Gers et al. 2000; Graves and Schmidhuber 2005). Some researchers have used the LSTM model in predicting the LOD (Gou et al. 2021), which might also be suitable for PM prediction problems. This study investigates the potential of utilizing LSTM combined with traditional methods for predicting PM. The method proposed is novel in that the non-linear part of PM is not predicted by the linear method AR model but through the deep learning method LSTM model.

This paper is structured as follows: In the second section, we describe the LSTM and LS + AR algorithms. Section three introduces the dataset and processing strategy, including the data used in each experiment, the amplitude variation and characteristics of AW, SAW, and CW in PM through the Fast Fourier Transform (FFT) spectrum analysis, and the detailed PM prediction process by LSTM, LS + AR, and LS + LSTM model. Next, we present different models to estimate PM variability, including LS + AR, LSTM, and LS + LSTM, all of which draw on IERS EOP 14C04 data from 2011 to the end of 2020. At the same time, Bulletin A from the IERS RS/PC is used to

compare the prediction accuracy with the results derived in this paper. A summary of the findings is given in the last section.

## Materials and methods

### LSTM prediction model

#### Introduction of the general concept of LSTM

LSTM is now widely used and has proven to perform well on various problems such as handwriting recognition, speech recognition, and time series prediction (Schmidhuber 2015; Alex et al. 2018). However, a neural architecture would not be widely utilized in practice without a solid theoretical foundation. Greff et al. (2017) recently reviewed several LSTM variants and their performances relative to the so-called vanilla model (Greff et al. 2017). The variant LSTM is an improved model based on the original LSTM (Hochreiter and Schmidhuber 1997; Gers and Schmidhuber 2000). The main change of the variant LSTM model compared to the original LSTM model is the addition of cell state information to the inputs of the three control gates. Unlike feedforward neural networks, the RNNs have a cycle function, which can take the activation in the previous steps as the network’s input and play a decisive role in the current input. However, training recurrent or very deep neural networks is challenging because they frequently suffer from exploding and vanishing gradient problems (Hochreiter 1991; Hochreiter et al. 2001). To solve the problems mentioned above, the LSTM architecture was developed to address this deficiency and the learning long-term dependencies. Figure 1 depicts the LSTM network structure, which is detailed in Appendix A.

#### LSTM training results analysis

In the LSTM network training, the hidden layers are set to 2 and the number of LSTM cells per hidden layer is 50. Time steps are set to 365 and training iterations are set to

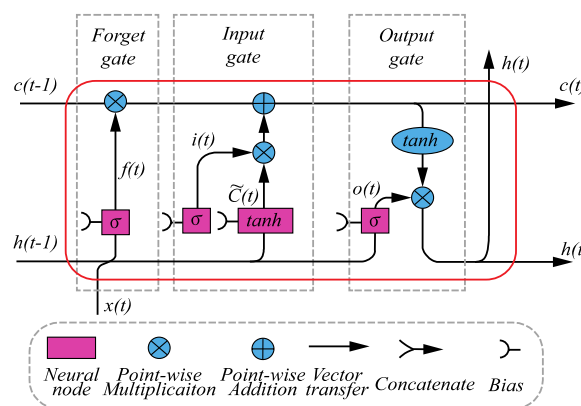
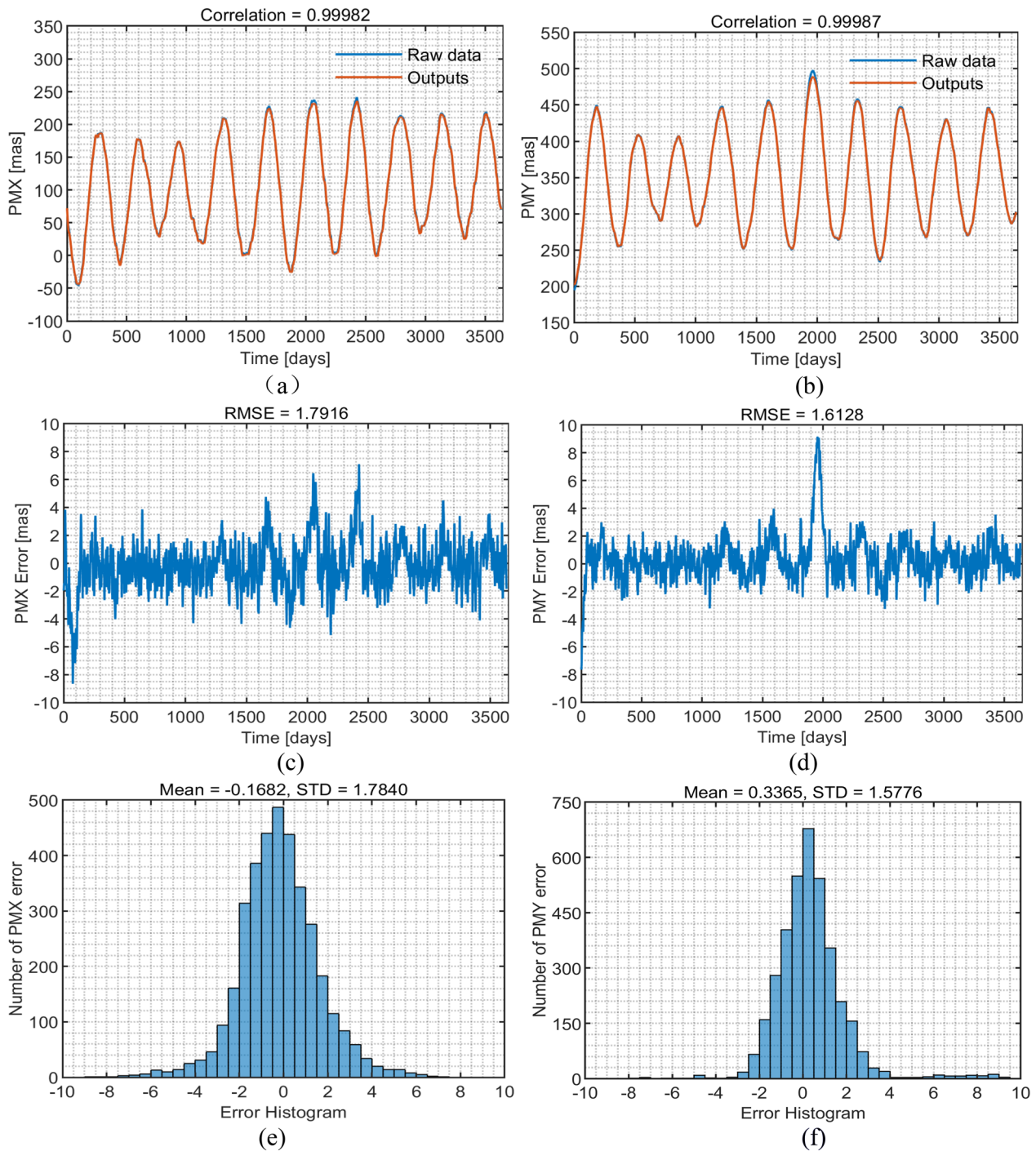


Fig. 1 The architecture of LSTM with a forget gate

1000. The learning efficient dropout is 0.1. The Savitzky–Golay (SG) smoothing filter is used in the experiments of this paper. The initial learning rate is set to 0.1 and the learn rate drop factor is 0.2 (Greff et al. 2017; Ren et al. 2020). The gradient threshold is set to 1 (Din et al.

2019). Other parameter settings are listed in Appendix A, Table 4. Figure 2 shows the LSTM network training based on the PM time series. Figure 2a and b indicates that the correlation between the original and output sequences of PMX and PMY is 0.99982 and 0.99987, respectively.



**Fig. 2** Prediction values of the PMX and PMY. **a, b** Are the LSTM network training, PMX and PMY raw data in the blue line and outputs in the red line, **c, d** are the errors, **e, f** are the error histogram of PM from 2011 to 2020

Figure 2c and d shows that the Root Mean Square Error (RMSE) of PMX and PMY is 1.7916 mas and 1.6128 mas, respectively. Figure 2e and f shows that the Mean and the Standard Deviation (STD) of PMX and PMY are  $-0.1682$  mas and  $0.3365$  mas, and  $1.7840$  mas and  $1.5776$  mas, respectively.

**LS + AR prediction model**

**LS model**

We use the following model to fit the trend and periodic terms of EOP, whose parameters can be estimated using the least squares method. The residuals are then analyzed by the AR and other models. The least squares model can be described as

$$\begin{aligned}
 F(t) = & A + Bt + C_1 \cos\left(\frac{2\pi t}{P_{SA}}\right) + C_2 \sin\left(\frac{2\pi t}{P_{SA}}\right) \\
 & + D_1 \cos\left(\frac{2\pi t}{P_A}\right) + D_2 \sin\left(\frac{2\pi t}{P_A}\right) + E_1 \cos\left(\frac{2\pi t}{P_C}\right) \\
 & + E_2 \sin\left(\frac{2\pi t}{P_C}\right) + \omega,
 \end{aligned}
 \tag{1}$$

where  $A$  is the constant,  $B$  is the trend term parameter in the model,  $C_1$  and  $C_2$  are the SAW parameters,  $D_1$  and  $D_2$  are the AW parameters, and  $E_1$  and  $E_2$  are the CW parameters. The fitting model calculates  $P_{SA}$ ,  $P_A$ , and  $P_C$  in years, representing the SAW, AW, and CW, respectively.

We additionally conducted an FFT analysis of the EOP 14C04 series. From Fig. 3a and b, it can be seen that CW and AW dominate the PM spectrum, manifested by cusps of power between 413 to 439 days (CW) and 356 to 376 days (AW). These values are relatively consistent with

estimates given elsewhere (Mccarthy and Luzum 1991; Schuh et al. 2001; Joachim 2004). In our model (Eq. 1), the AW period is 365.25 days, CW is 434 days and SAW is 182.62 days.  $\omega$  is the random error,  $t$  is the UTC of the series, and the unit is converted into years when LS fitting. Similarly, the meaning of each corresponding parameter table in the PMX series is identical to that of the PMY series.

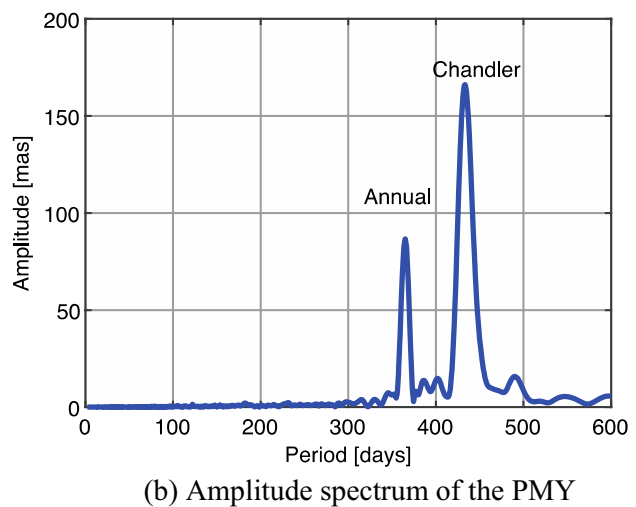
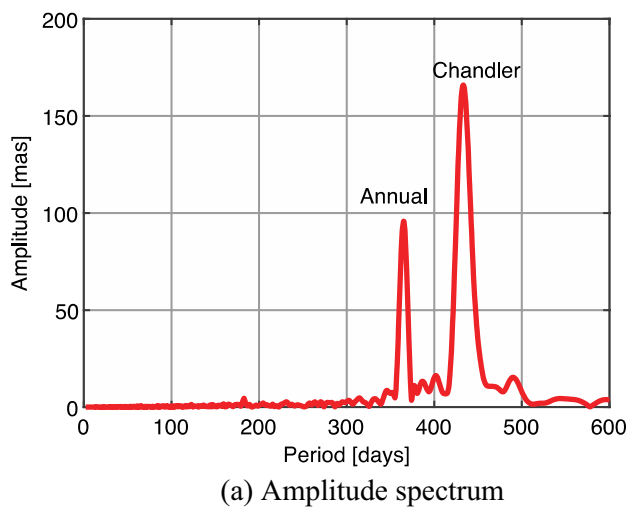
**AR model**

AR( $p$ ) model is the description of the relationship between a random series  $z_t$  ( $t=1, 2, \dots, N$ ) before time  $t$  and the current time. Its expression can be written as follows:

$$z_t = \sum_{i=1}^p \phi_i z_{t-i} + \omega_t,
 \tag{2}$$

where  $\phi_1, \phi_2, \dots, \phi_p$  represent the autoregressive coefficients obtained by solving the Yule–Walker equations using the Levinson–Durbin recursion (Brockwell and Davis 1997),  $\omega_t$  is the white noise with zero means, and  $p$  stands for the model order. The above equation denoted by AR( $p$ ) is the AR model of the order  $p$ , and how to determine the order  $p$  is crucial. Usually, there are three methods for the determination of  $p$ , Akaike’s final prediction error (FPE) criterion, the information criterion, and the delivery function criterion. In this paper, the FPE criterion is adopted to determine the order  $p$  and corresponds to the smallest FPE (Akaike 1971):

$$FPE_p = P_p(N + p + 1)/(N - p - 1),
 \tag{3}$$



**Fig. 3** Amplitude spectrum of the two components of polar motion (PMX in red, PMY in blue) as deduced from FFT of the respective time series. Units are [mas]

$$P_p = 1/(N - p) \sum_{t=p+1}^N \left( z_t - \sum_{j=1}^p \varphi_j z_{t-j} \right)^2 \tag{4}$$

The mean absolute error (MAE) is utilized to evaluate the prediction accuracy. It can be expressed as follows:

$$MAE_j = \frac{1}{n} \sum_{i=1}^n (|P_i - X_i|), \tag{5}$$

where  $P_i$  represents the predicted value of  $i$ -th prediction,  $X_i$  stands for the corresponding observation value,  $n$  is the total prediction number, and  $MAE_j$  is the MAE at span  $j$ .

### Data description and processing strategy

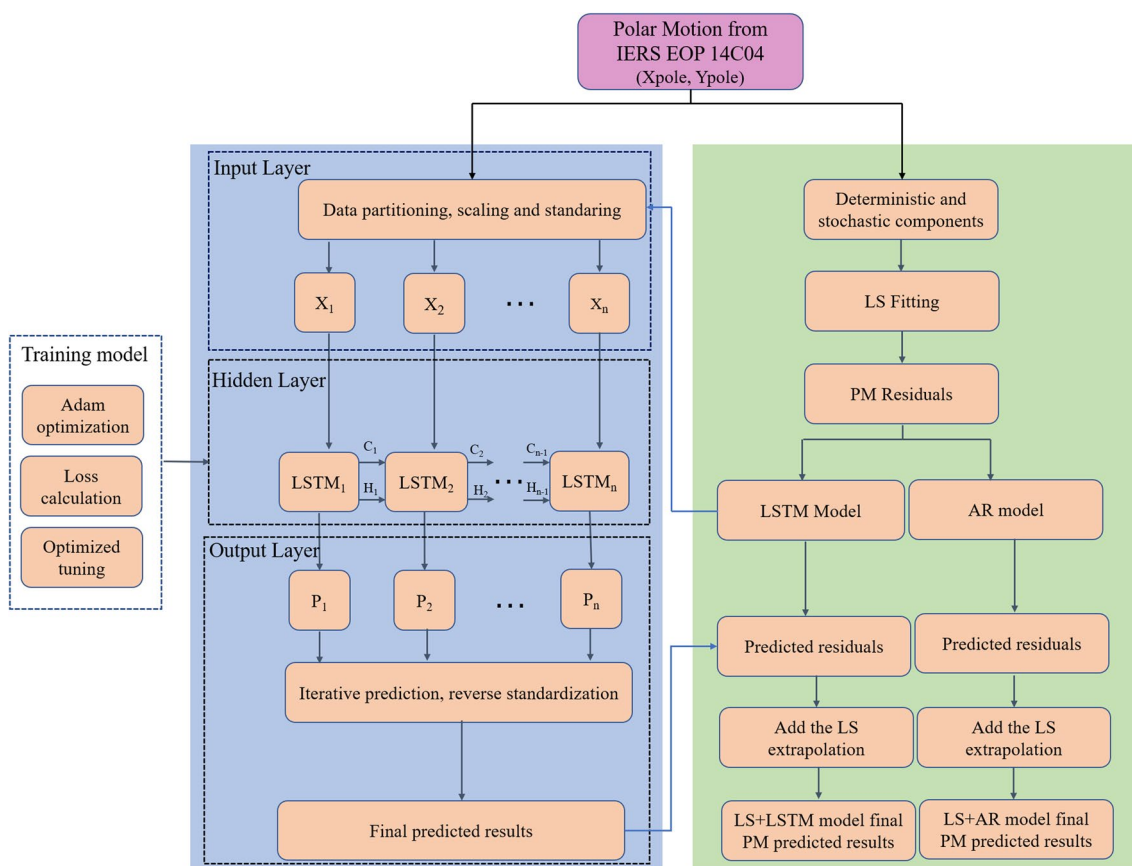
#### Data description

In particular, we use the PM time series from IERS EOP 14C04 with daily sampling interval is available at <https://hpiers.obspm.fr/eopppc/eop/eopc04/>. In this study, we use the PM series from January 8, 2011 to December 31, 2021. The results will be compared to Bulletin A (558 files) available at <https://www.iers.org/IERS/EN/DataP>

[roducts/EarthOrientationData/eop.html](https://www.iers.org/IERS/EN/DataProducts/EarthOrientationData/eop.html), for the same period of time as IERS EOP 14C04. The LSTM network training is based on the PM time series from January 1, 2011 to December 31, 2020.

#### PM prediction processing strategy

Figure 4 depicts a schematic representation of the methodology adopted for predicting PM with various models. The observed PM can be divided into deterministic and stochastic components. The known component is referred to as a priori model, consisting of the long-term trend, CW, AW, and SAW. In this study, the LS+AR model (first method) is applied to forecast PMX and PMY and compared to results based on the LS+LSTM (second method) and LSTM models (third method). Figure 4 describes the respective processing schemes. For the first two methods, training patterns are derived from the residuals after subtracting the a priori model. These patterns are used for training the LSTM. The predicted residuals are then added to the a priori model to obtain the final predicted values of the PMX and PMY. The third method for predicting PM uses the LSTM model directly, relying on the IERS EOP14C04 time series.



**Fig. 4** Flowchart of the LS+AR, LS+LSTM, and LSTM approaches for PM prediction

## Results and discussion

### PM prediction using the LS + AR model

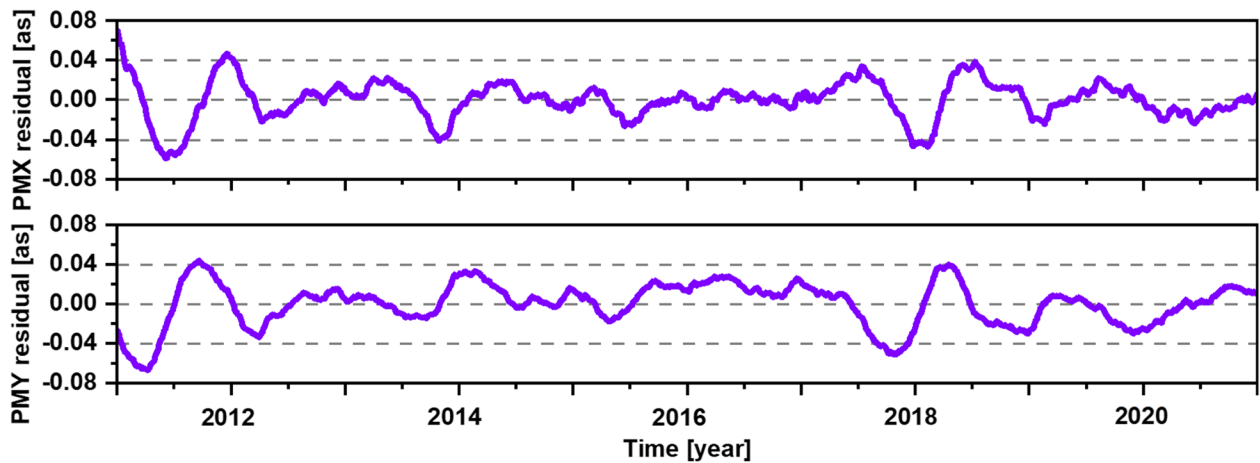
We initially preprocess the PM series using the LS model and deduce residuals (that is, the stochastic components) by subtracting the LS analysis results from the original pole coordinates. Figure 5a shows the PM residual results (purple line) derived from the IERS EOP 14C04 from 2011 to 2020. The residuals of the PMX and PMY are within  $\pm 0.08$  arcsecond (as). Due to the nature of the LS fitting model, the fluctuations at the start and end of the time series are somewhat larger compared to the middle part. Figure 5b depicts the first-order difference of the residual sequence (brown line) for PMX and PMY. Most of the residual values and the first difference values are within  $\pm 0.1$  arcseconds (as) and  $\pm 2.0$  milliarcseconds (mas), respectively.

The evaluation of the Autoregressive Integrated Moving Average (ARIMA) model type is essential. When

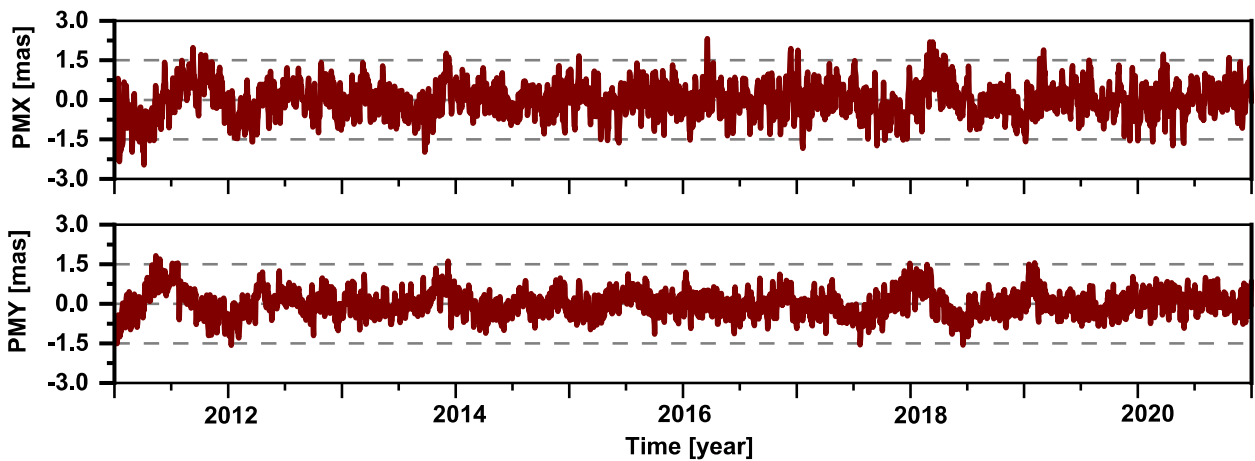
$p = 0$ , the ARIMA( $p, q$ ) (autoregressive integrated moving average) model can be expressed as MA( $q$ ), i.e.,  $q$  order moving average model. When  $q = 0$ , the model can be described as AR( $p$ ), i.e., a  $p$ -order autoregressive model (Box et al. 1976). For the PM time series, Table 1 illustrates the criteria for which model can be evaluated according to the autocorrelation and partial correlation functions of the time series. Figure 6 depicts the

**Table 1** Judgment criterion of the ARIMA model

Autocorrelation function	Partial autocorrelation function	Strategy
Tails off to zero	Truncation	AR( $p$ )
Truncation	Tails off to zero	MA( $q$ )
Tails off to zero	Tails off to zero	ARIMA( $p, q$ )

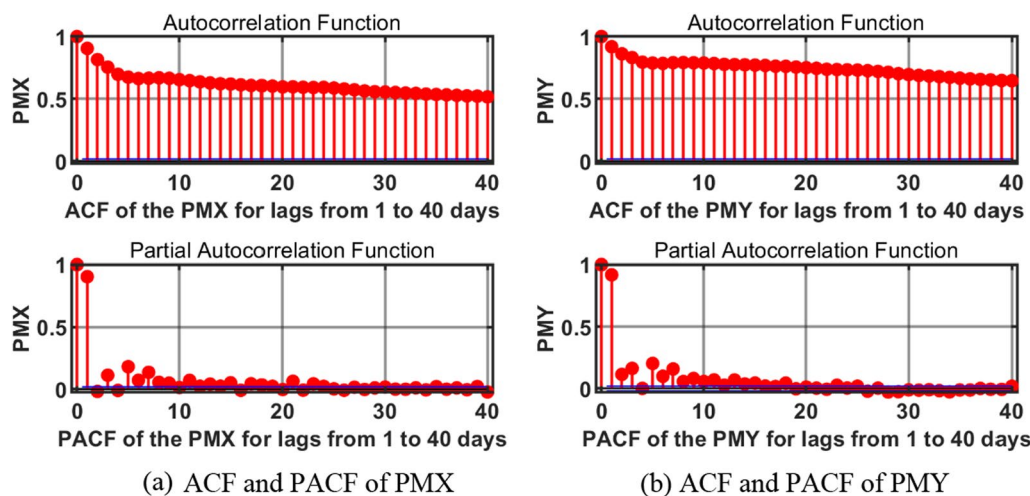


(a) Residual series of PMX and PMY



(b) First-order difference series of PM residuals series

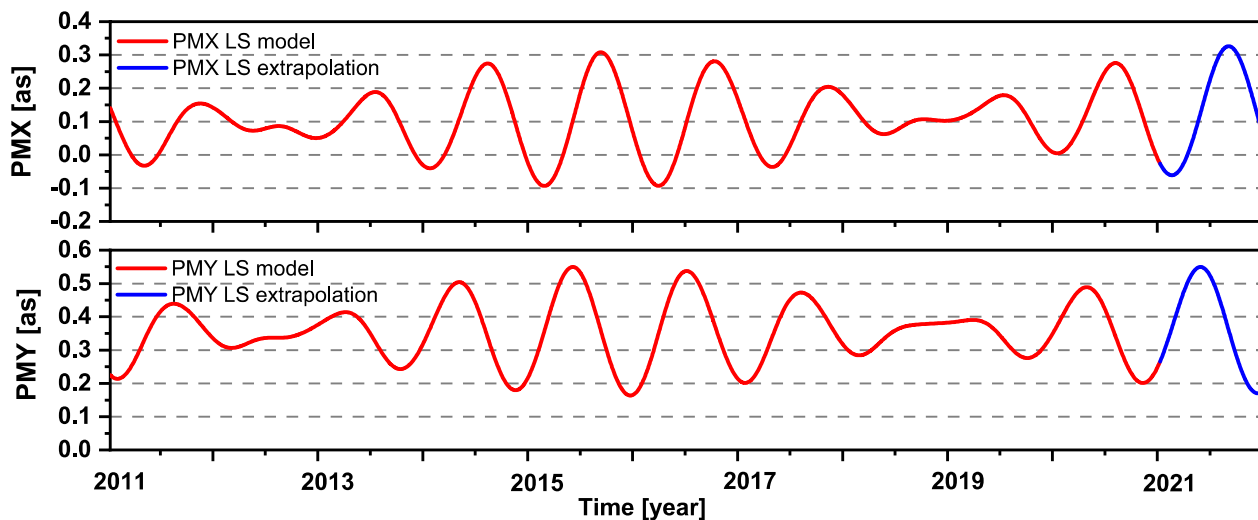
**Fig. 5** Residual series (purple line) of the LS fitting model and the first-order difference series (brown line) of the residuals from 2011 to 2020



**Fig. 6** The ACF and PACF of first-order differential of PM residuals time series for lags from 1 to 40 days. All results in units of [mas]

subsequent calculation of the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the first-order differential of residuals time series with a delay of 1 to 40 days. The results indicate that the autocorrelation function of the first-order difference sequence of the residual error is tailing, and the partial correlation function is truncated, allowing the  $AR(p)$  model to be used for prediction, i.e.,  $q = 0$  (Schaffer et al. 2021). In this research, the FPE is used to determine order  $p$ , and this method is described in Eqs. (3) and (4). The optimal order  $p$  for the AR model, as determined by the final prediction error criterion, is set to 50.

In this experiment, we extrapolate the deterministic part of 2021 (365 days) using the LS model, based on the least squares fitting series of the IERS EOP14C04 from January 1, 2011 to December 31, 2020 (10 years). Figure 7 shows the LS model time series (red line) and LS extrapolation time series (blue line) of PM. The experimental purpose is to use the LS+AR model for PM prediction. The final prediction results of the LS+AR model are the sum of the LS extrapolation using the determined part and the prediction results of the AR model based on the residual sequence.



**Fig. 7** The LS fitting results of the PMX and PMY (red line) between 2011 and 2020, and the LS extrapolation of PMX and PMY (blue line) in the future 365 days of 2021. All outcomes are given in units of [as]



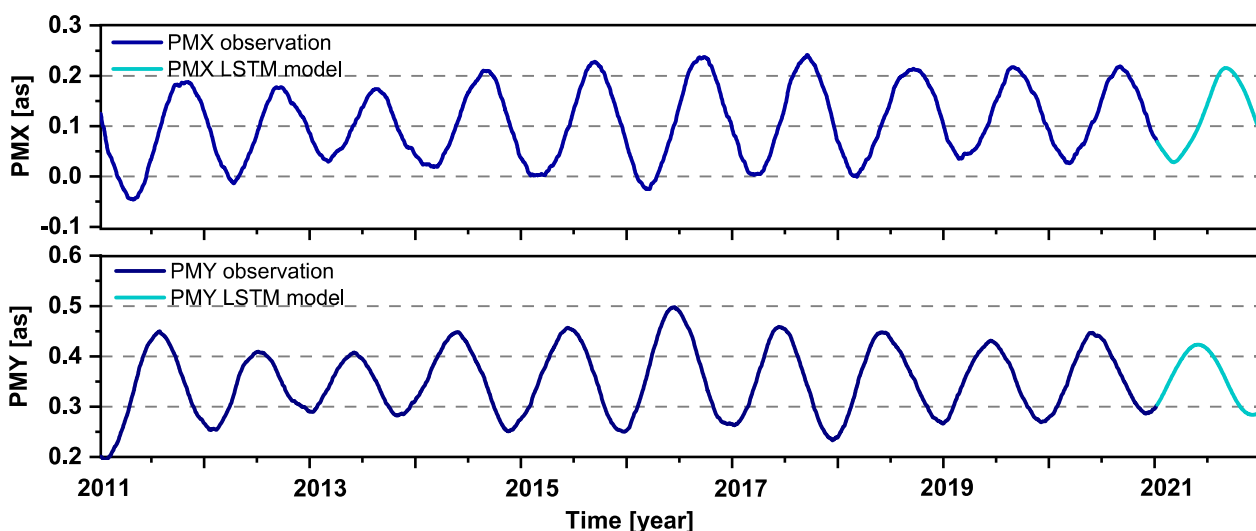
**PM prediction using the LSTM model**

Figure 8 displays the PM prediction based on the IERS EOP14C04 time series (blue line) utilizing the LSTM model. The green line represents the prediction outcomes for 2021 (365 days). It can be seen that the prediction results of PMX and PMY based on the LSTM model for 365 days in 2021 are consistent with the overall trend of PMX and PMY time series from the IERS EOP 14C04. Most PMX prediction values fall between 0.15 as and 0.21 as and the PMY prediction values are between 0.28 as and 0.42 as.

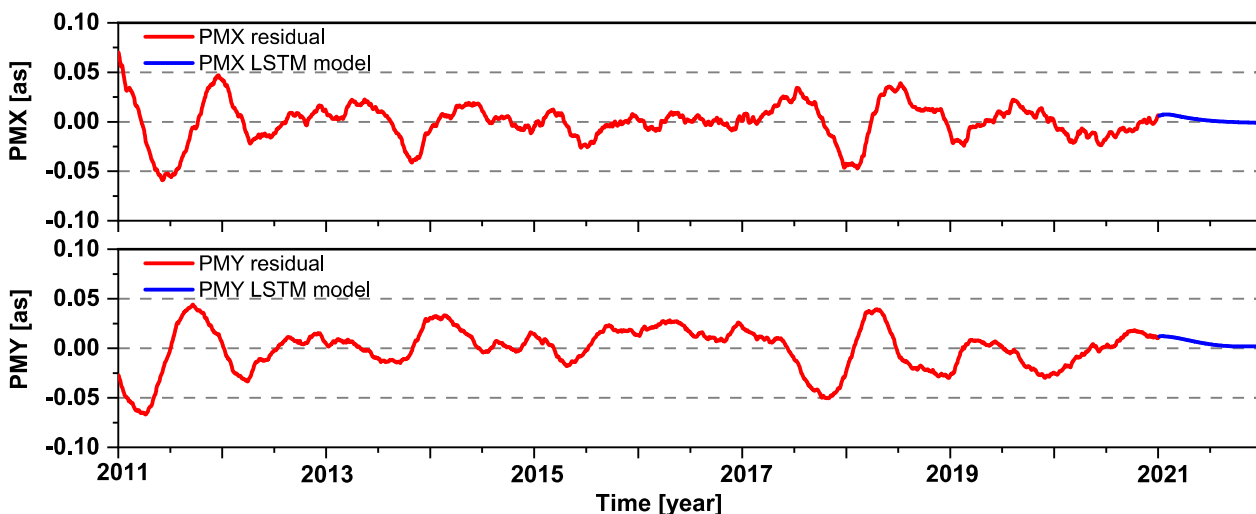
**PM prediction using the LS + LSTM model**

To investigate the contribution of the LS+LSTM model in PM prediction, LSTM is applied to forecast the residual part in 2021, using the residuals' basic time series. Figure 9 presents the 2021 residual values (blue line) prediction from the LSTM model. The final PM prediction using the LS+LSTM model is the total results of the LS extrapolation from the LS model and the residuals determined by the LSTM model.

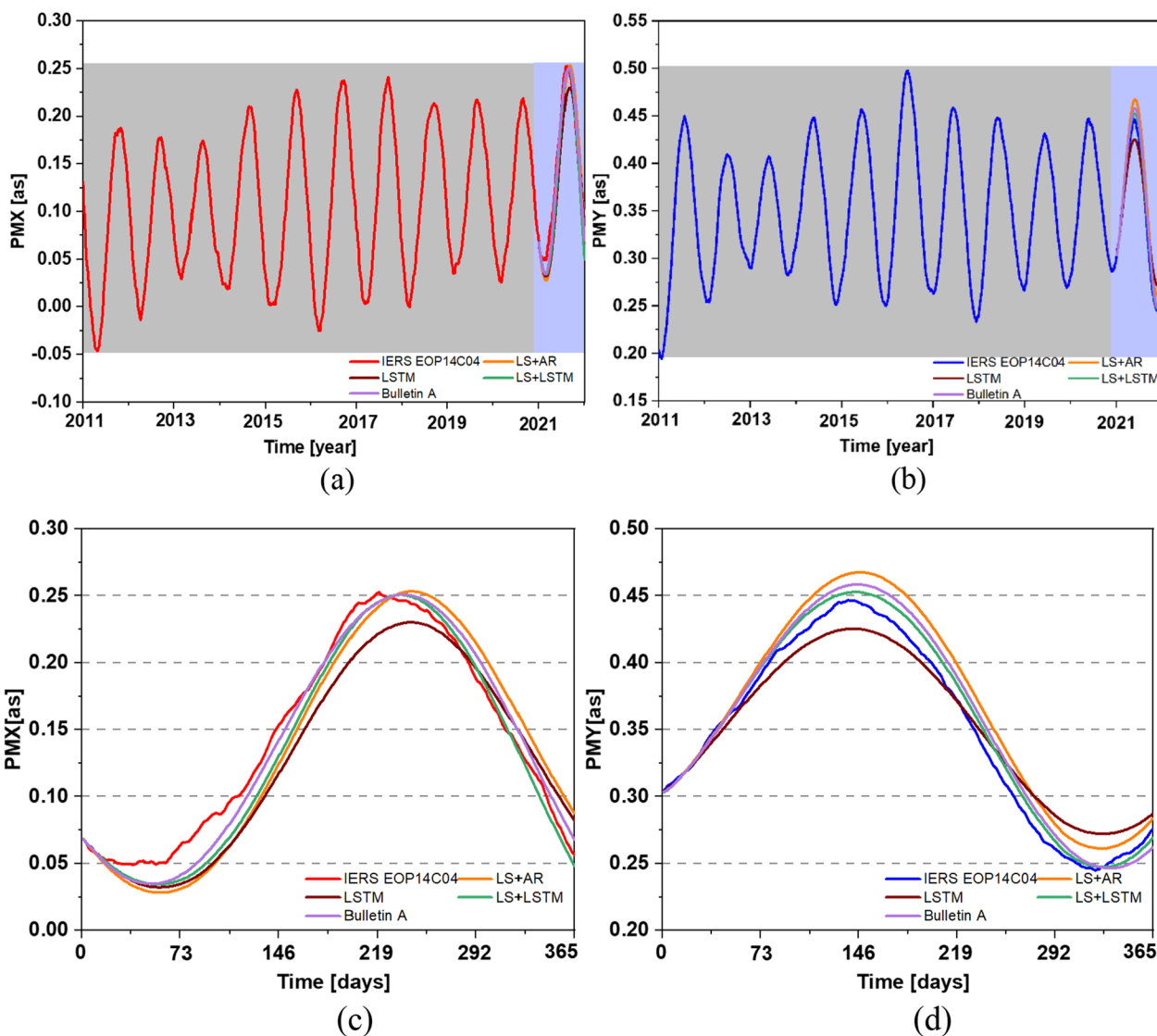
Figure 10a and b depicts the final prediction of the PM with different methods, including the LS+AR, LSTM,



**Fig. 8** The PM prediction results (green line) from the LSTM model in 2021, based on the PMX and PMY observation (blue line) from IERS EOP14C04 between 2011 and 2020. All results in units of [as]



**Fig. 9** Residuals of PM (red line) from 2011 to 2020, and the residual prediction results (blue line) of the PMX and PMY based on the LSTM model in 2021. All results are given in units [as]



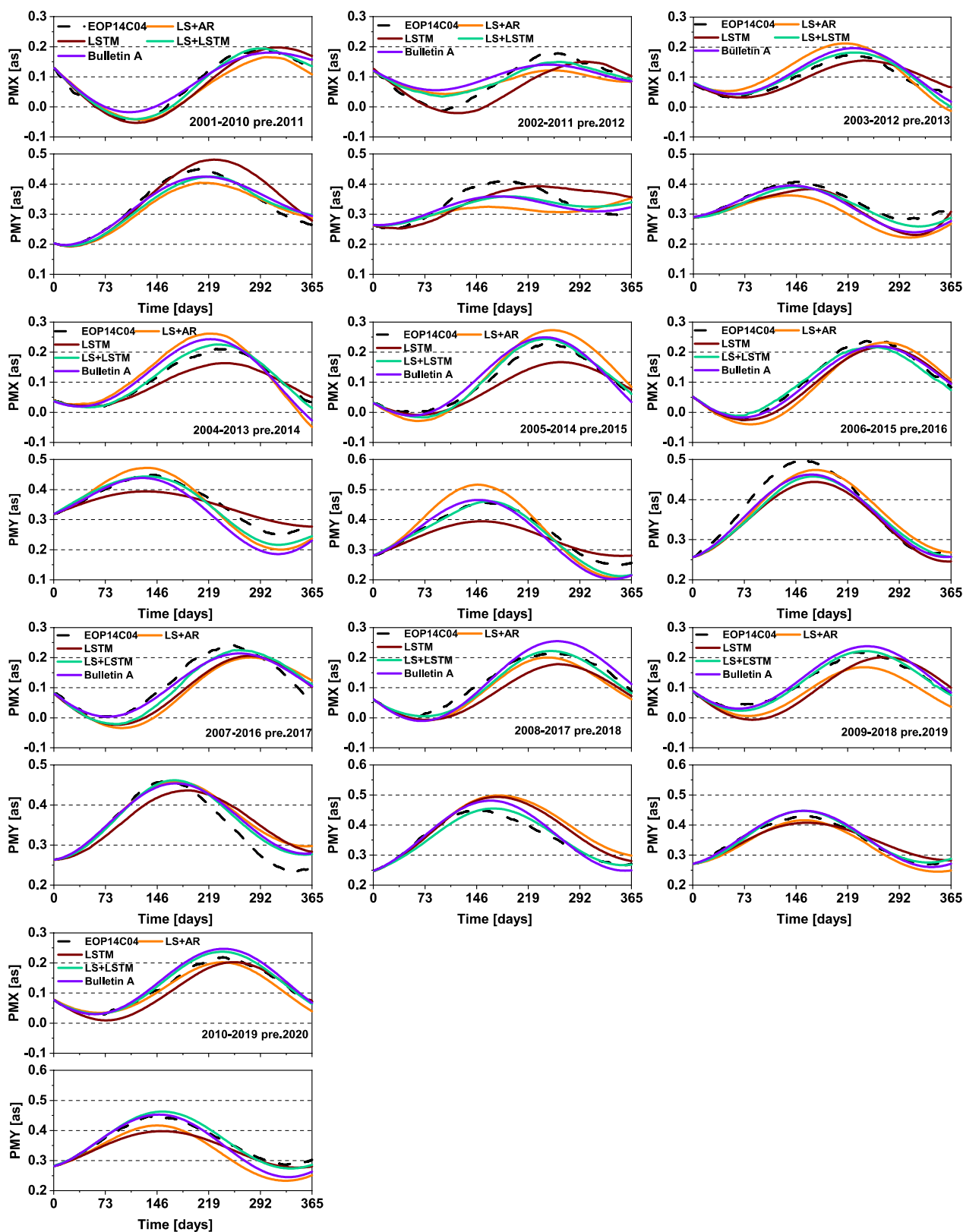
**Fig. 10** The PMX and PMY predictions from different models, i.e., LS+AR, LSTM, LS+LSTM, and Bulletin A model, are shown in **a** and **b** panel, respectively, and the IERS EOP 14C04 as a reference. Panels **c** and **d** show the 2021 forecasts as separate figures. All results in units of [as]

and LS+LSTM models. In addition, the forecast results of Bulletin A in 2021 are included (purple line). The IERS EOP14C04 time series is considered a benchmark for comparing the estimated outcomes of various techniques. In terms of PM prediction, the results predicted by the LS+LSTM model (green line) in Fig. 10c and d are the closest to the IERS EOP14C04 time series (red) over the mid- and long-term prediction. Although the improvement is marginal, the findings predicted by the LS+LSTM model are very close to or better than Bulletin A in the mid- and long-term prediction of PMX, and it also can be seen that the PMX prediction accuracy from the LSTM model is higher than that from the

LS+AR model in the mid- and long term. For PMX, the RMSE of the results is 0.035 as, 0.031 as, 0.018 as, and 0.030 as for LS+AR, LSTM, LS+LSTM, and Bulletin A, respectively. For PMY, the RMS of the results is 0.038 as, 0.035 as, 0.015 as, and 0.035 as for LS+AR, LSTM, LS+LSTM, and Bulletin A, respectively.

**Evaluating the PM prediction results**

Based on the previous analysis (Fig. 10), the LS+LSTM model prediction accuracy is higher than other models. To assess the accuracy of this method in a more extended way, we compare the PM prediction in different periods using the LS+LSTM model, LS+AR model, and



**Fig. 11** Prediction values of PMX and PMY with different models, including the LS+AR, LSTM, and LS+LSTM, respectively. All results in units of [as]. IERS EOP 14C04 from 2011 to 2020 is shown as reference for comparisons

LSTM model as shown in Fig. 11. The prediction span is 365 days with a ten-year basic sequence, and the statistical period is from 2011 to 2020. The model proposed in this experiment is based on the IERS EOP 14C04 for PM prediction. In Fig. 11 the orange, brown, green, and purple lines represent the LS+AR, LSTM, LS+LSTM, and Bulletin A prediction of PM, respectively. We also compare the PM prediction results from the LS+LSTM model to the IERS EOP 14C04. In the mid- and long-term prediction of the PM, the prediction results based on the LS+LSTM model are closer to the observed IERS EOP 14C04 time series than those based on Bulletin A.

#### **AE of PM prediction with different models**

Experiments have demonstrated that the LS+LSTM model is superior for predicting PM, especially in the mid- and long term. To further explore the advantages of the LS+LSTM model in the accuracy of PM prediction, four different cases were designed to predict PM for 11 years (from 2011 to 2021). In this experiment, the prediction span was 365 days with a weekly sliding window. The experiment is divided into four parts, considering the following methods:

Case 1: PMX and PMY prediction based on the LS+AR model;

Case 2: PMX and PMY prediction based on the LSTM model;

Case 3: PMX and PMY prediction based on the LS+LSTM model;

Case 4: PMX and PMY prediction from Bulletin A achieved from the IERS RS/PC.

The four cases listed above correspond to the LS+AR, the LSTM, the LS+LSTM, and the Bulletin A provided by the IERS, respectively. Authors generally rely on the 10-year IERS EOP14C04 time series in the PM prediction as the basic series (Xu et al. 2012; Xu and Zhou 2015; Kenyon et al. 2012). In the following experiments, using various methods, we also choose a ten-year base sequence to predict the PM for the next 365 days.

Figure 12 shows the PM prediction's absolute errors (AE) using four cases. All experimental results take the IERS EOP14C04 time series as a reference. It can be seen that the accuracy of the four cases (LS+LSTM, LSTM, LS+LSTM, and Bulletin A) in 2011–2015 is inferior to that in 2016–2021. One potential reason could be the 2011 earthquake on the Pacific coast of Tōhoku (the 3.11 Japan earthquake). Based on the data from the Jet Propulsion Laboratory (JPL) of the National Aeronautics and Space Administration (NASA), the 3.11 Japan earthquake shifted the Earth's rotation axis by 25 cm and accelerated the Earth's rotation rate by 1.8 microseconds (Gross 2007). Earthquakes not only cause significant changes in the Earth's rotation on the day they occur, but they also

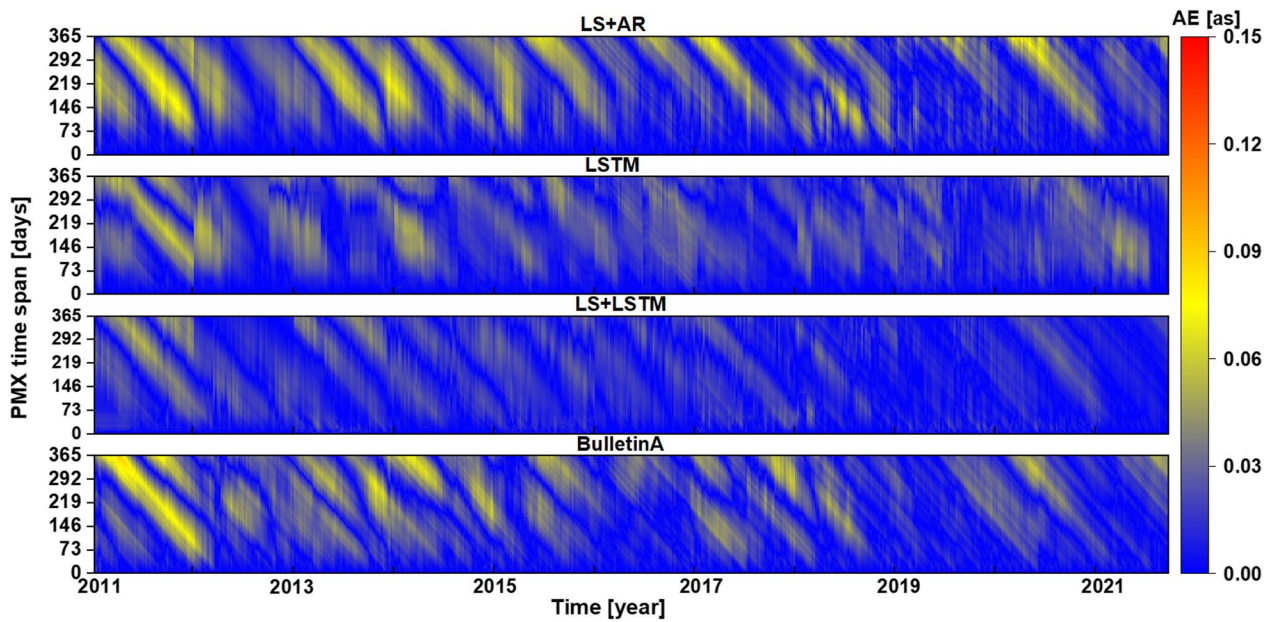
impact the location changes of surface stations over the next 3–5 years, thus affecting ERP monitoring (Souriau 1986; Bizouard, 2005; Bogusz et al. 2015). IERS introduced post-seismic deformation (PSD) in 2017 when establishing the most recent international terrestrial coordinate framework (ITRF2014) to reduce the influence of earthquakes on ground stations and obtain more accurate ERPs data. It is worth noting that ITRF2014 is the most recent ITRF solution at the time of the study. The IERS EOP14C04 also corrected the PSD model of the large earthquake in Japan in March 2011 to precisely solve the PM change phenomenon at this stage; hence, the PM results at this phase deviated from the previous comprehensive trend. However, this deviation was not considered when the models described in this study were used to predict PM, likely resulting in prediction errors. Thus, our preliminary conclusion is that the larger deviations of the prediction results from the observed values (EOP 14C04) between 2011 and 2015 are attributable to the effects of large earthquakes. To improve the accuracy of PM prediction following a major earthquake, further PSD model processing of the prediction algorithm is required. The results predicted by the LS+LSTM are closer to the IERS EOP14C04 series than those by the LS+AR, LSTM, and Bulletin A in mid- and long-term prediction.

#### **MAE of PM prediction with different models**

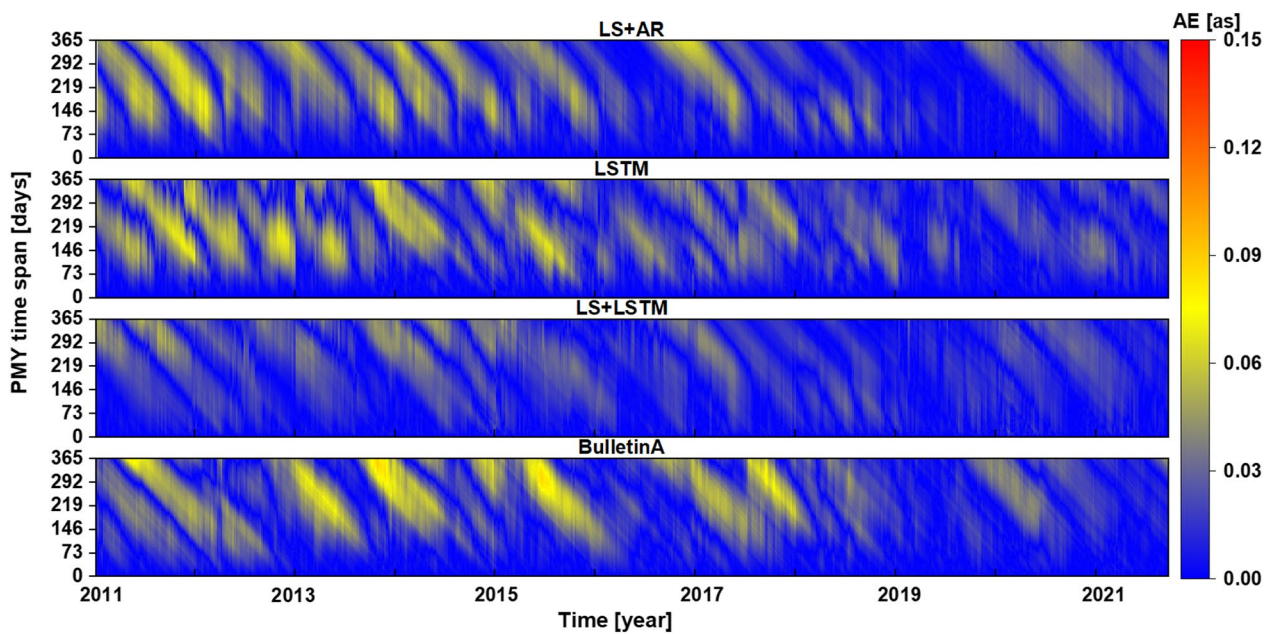
Figure 13 shows the Mean absolute errors (MAE) of PMX and PMY prediction in four cases. Compared to the other models, the MAE of the proposed LS+LSTM model yields smaller errors in mid- and long-term prediction. Since the LS+LSTM model better considers the overall characteristics of the base series, it obtains a more accurate long-term trend and long period term than the LS model during extrapolation, thus improving the mid- and long-term PM prediction accuracy.

As Table 2 reveals, the estimation accuracy of the PM is determined using different models, i.e., the LS+AR, LSTM, and LS+LSTM models. The MAE of predicted PM at various periods (1, 5, 10, 15, 20, 30, 45, 90, 120, 180, 270, 320, 365 days) is listed in Table 2. Combined with the PM prediction accuracy statistics, the improvement of LS+LSTM over Bulletin A is clear after 120 days. The improvement gradually increases with the lengthening of the prediction span, reaching a maximum of 33.7% and 31.9% in PMX and PMY, respectively. Generally, the LS+LSTM model has more advantages than the LSTM model, the traditional linear prediction model (LS+AR model), and Bulletin A in mid- and long-term PM prediction.

However, Bulletin A exhibits a smaller MAE in the four cases for short-term prediction. Table 2 demonstrates



(a) Absolute errors of PMX final predicted results

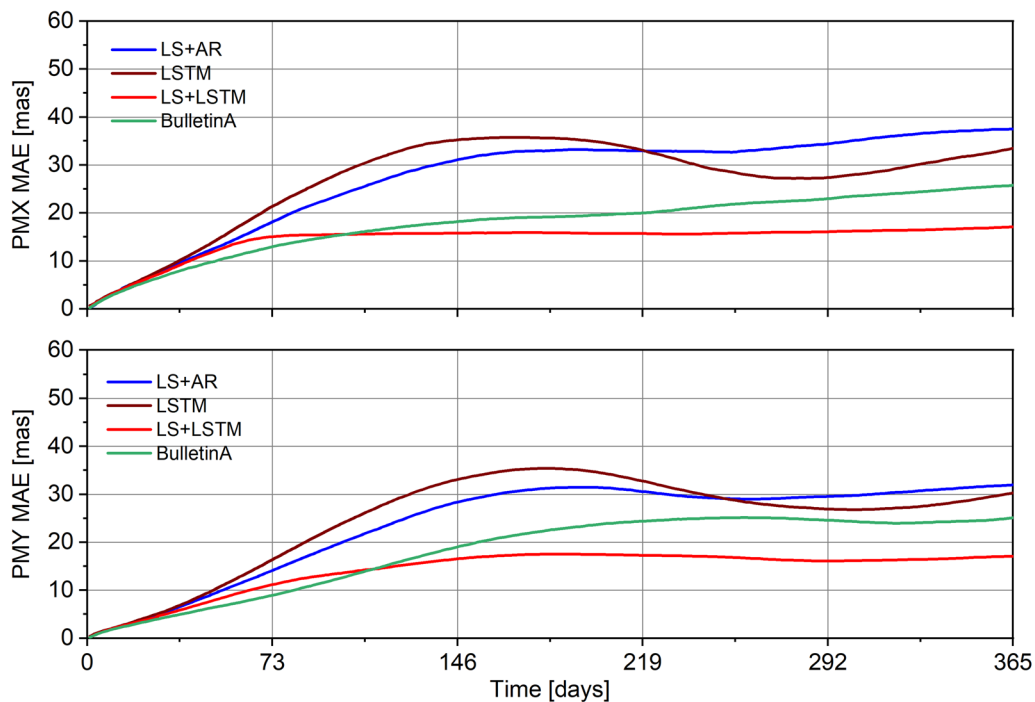


(b) Absolute errors of PMY final predicted results

**Fig. 12** Absolute errors (AE) of the PMX and PMY prediction from 2011 to 2021, and all outcomes with the unit [as]. All results are based on 10-year time series of IERS EOP 14C04, using the LS+AR, LSTM, and LS+LSTM models, respectively. The MAE of Bulletin A relies on the PM prediction, coming from the IERS RS/PC

that Bulletin A outperformed the other models regarding short-term prediction, especially the ultra-short term (the first ten days in the future). This advantage is primarily due to Bulletin A considering the effects of

atmospheric angular momentum (AAM) and oceanic angular momentum (OAM). In addition, the statistical results demonstrate that the LSTM model is superior to



**Fig. 13** Mean absolute errors (MAE) of the PMX and PMY prediction using the four cases, namely, the LS + AR model (blue line), LSTM model (brown line), LS + LSTM model (red line), and Bulletin A is indicated by the green line. All results in units of [mas]

**Table 2** Mean absolute errors (MAE) of the polar motion (PMX, PMY) prediction [mas] in 2021, using the LS + AR, LSTM, LS + LSTM, and Bulletin A time series from IERS, respectively

No.	PMX					PMY				
	LS + AR	LSTM	LS + LSTM	Bulletin A	PCT	LS + AR	LSTM	LS + LSTM	Bulletin A	PCT
1	0.257	0.676	0.284	0.269	-5.6%	0.197	0.466	0.204	0.196	-4.2%
5	1.603	1.898	1.690	1.563	-8.1%	1.099	1.108	1.129	1.073	-5.2%
10	3.046	3.289	3.152	2.892	-9.0%	1.940	1.812	2.009	1.843	-9.0%
15	4.593	4.531	4.374	3.988	-9.7%	2.751	2.598	2.729	2.480	-10.0%
20	5.751	5.793	5.546	5.033	-10.2%	3.607	3.509	3.386	3.077	-10.1%
30	7.995	8.362	7.414	6.809	-12.1%	5.407	5.501	4.876	4.261	-14.4%
45	11.483	12.561	10.995	9.227	-19.2%	8.292	8.879	7.137	5.904	-20.9%
60	14.915	17.281	13.778	11.260	-22.4%	11.382	12.764	9.426	7.454	-26.5%
90	21.950	25.895	15.419	14.619	-5.5%	17.785	21.144	12.868	11.176	-15.1%
120	27.506	32.397	15.685	16.819	6.7%	23.904	28.571	14.905	15.500	3.8%
180	32.933	35.620	15.904	19.148	16.9%	31.231	35.412	16.424	22.356	26.5%
270	33.364	27.344	15.926	22.318	28.6%	29.096	27.771	17.218	25.066	31.3%
320	36.116	29.332	16.393	24.096	32.0%	30.430	27.190	16.350	23.980	31.8%
365	37.491	33.456	17.086	25.770	33.7%	31.949	30.263	17.100	25.107	31.9%

The improvement represents the prediction accuracy of LS + LSTM relative to Bulletin A

the traditional model (LS+AR) in the long-term (270–365 days) prediction of PM.

## Conclusions

Polar motion is a crucial parameter describing the instantaneous movement of the Earth's rotation axis relative to the body-fixed reference frame. Among existing prediction models, linear models are often used to predict PM, such as the LS+AR model. Here, we have analyzed the PM series data from January 8, 2011 to September 11, 2021 by different models, including LS+AR, LSTM, LS+LSTM, and Bulletin A. The residual series used in this research is obtained by removing the long-term trend term and the calculated AW, SAW, and CW values. In this paper, based on the characteristics of PM and its inherent periodic and trend terms, the LSTM prediction model is proposed. To verify the advantages of LSTM and its combination with LS in PM prediction, the basic sequence length of 10 years is selected, which is optimal for the LS+AR model to predict PM. The experimental findings demonstrated that the LS+LSTM model is superior for mid- and long-term forecasting of PM. Compared to Bulletin A, published by IERS, the LS+LSTM model demonstrates improved PMX and PMY prediction accuracy by up to 33.7% and 31.9%, respectively, and the LSTM model outperforms the LS+AR model in the mid- and long term.

The study's findings rely heavily on the 10-year snippet of PM time series between 2011 and 2021. Future research will investigate the relationship between the length of the basic time series, the seismic factors, and the accuracy of LSTM and LS+LSTM models in predicting PM. The prediction model such as LS+LSTM, based on proper base sequence length and seismic factor correction, will be established to improve the short-term PM prediction. In addition, the benefits of combining the LSTM with LS and other traditional methods for PM short-term prediction need to be further explored.

## Appendix

### Description of the LSTM training process

LSTM adopts three unique gate designs to avoid gradient explosion and long-term dependence (Yu et al. 2019). Since each cycle uses information from the previous cycle and each output state is affected by the previous state, the LSTM network can better remember long-term laws more effectively and is widely used in time series prediction, such as financial time series prediction (Zhang et al. 2019), GNSS time series prediction (Wang et al. 2021), and weather forecasting (Karevan and Suykens 2020). A vanilla LSTM unit contains a cell, an input gate, an output gate, and a forget gate. This forget gate was not initially a part of the LSTM network but was proposed by Gers

et al. (2000). Figure 1 depicts the LSTM network structure adopted in this paper. At time  $t$ , the first layer network comprises two information flows. The information flow from  $C_{t-1}$  to  $C_t$  represents the transmission of cell state. The entire line linearly interacts with the following information flow through three gate control structures.

The gate structure allows information to selectively pass through, i.e., determining whether the information is removed or added to the cell state from  $C_{t-1}$  to  $C_t$ . That is to say, this part is the screening of information input in the gating structure. The  $\sigma$  activation function layer and the  $\tanh$  activation function layer can convert the input between (0, 1) and (-1, 1), generate the weight of the input data, and filter the input data. There are three gate structures in each layer of the LSTM network to control the cell state:

#### (1) Forget gate

The first step in the LSTM is to decide what information will be removed from the cell state. This decision is made by the  $\sigma$ -layer called the "forget gate layer." It looks at  $h_{t-1}$  and  $x_t$  and outputs a number between 0 and 1 for each number in the cell state  $C_{t-1}$ . 1 is "completely keep  $h_{t-1}$  and  $x_t$  information," while the 0 represents "completely get rid of  $h_{t-1}$  and  $x_t$  information." The formula of the "forget gate"  $f_t$  is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (6)$$

where  $\sigma$  is the activation function,  $W_f$  is the weight,  $h_{t-1}$  is the recurrent information at time  $t-1$ ,  $x_t$  denote the input information, and  $b_f$  is the bias of the forget gate.

#### (2) Input gate

One part of the "input gate" linearly combines  $x_t$  with the hidden state  $h_{t-1}$  at the previous time to obtain  $i_t$  through  $\sigma$ -layer activation from Eq. (7) (Wang et al. 2021). This part determines which information needs to be updated; this is part of the forgetting gate selected to be forgotten. In the other part,  $h_{t-1}$  and  $x_t$  are passed through a  $\tanh$  layer to generate a vector  $\tilde{C}_t$ , which is alternatively employed to update the new content. Then the two parts are combined to update the state  $C_{t-1}$  to  $C_t$ . The expression of the "input gate" is as follows (Hochreiter and Schmidhuber 1997; Wang et al. 2021):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (7)$$

$$\tilde{C}_t = \tanh[W_C(h_{t-1}, x_t) + b_C], \quad (8)$$

cell state equation at time  $t$ ,

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (9)$$

where  $C_t$  denotes the cell state of LSTM,  $\tanh$  is the activation function,  $W_i$ ,  $W_C$  are the weights,  $*$  represents

convolution,  $b_i, b_c$  are the bias of the input gate,  $\tilde{C}_t$  is the cell status update value, and other parameters are consistent with those mentioned above.

(3) Output gate

The "output gate" updates the value of the hidden layer output at the current time, i.e.,  $h_t$ , through Eq. (11), is the hidden state at time  $t$ . LSTM can remember long-term historical information because every cycle uses the information  $C_t$  and  $h_{t-1}$  of the previous cycle, and each output state is affected by the previous state. Especially, the forget gate can decide what information will be removed from the cell state. The formula of the "output gate" is as follows:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{10}$$

$$h_t = o_t * \tanh(C_t), \tag{11}$$

where  $o_t$  is the output gate,  $W_o$  is the weight associated with  $x_t$ ,  $b_o$  is for the bias weight vector, and  $h_t$  is the hidden state. Equations (6)–(10),  $W_\zeta, b_\zeta, \zeta \in \{f, i, c, o\}, \xi \in \{f, i, c, o\}$ , respectively, represent the output weight and offset matrix, which are also parameters to be learned in training.

In network training, one should pay attention to possible overfitting. It can minimize the loss function by constantly adjusting the parameters. For example, when the total number of samples is  $N$ , the output value  $Y_i^*$  trained by the network and the expected output value  $Y_i$  can be expressed by the mean squared error (MSE) loss function, also known as the L2 Loss function. Its basic form is as follows:

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N (Y_i^* - Y_i)^2. \tag{12}$$

LSTM Model training is the process of determining the parameters in the model structure. First, the original PM time series defines as  $F_o = \{f_1, f_2, \dots, f_n\}$  in the input layer. The training set and test set can be divided into  $F_{tr} = \{f_1, f_2, \dots, f_m\}$  and  $F_{te} = \{f_{m+1}, f_{m+2}, \dots, f_m\}$ , satisfying the constraint conditions  $m < n$ , and  $m, n \in \mathbb{N}$ . Then standardize the element  $f_t$  in the training set using the classic z-score standardization formula as Eq. (13) (Liu et al. 2019). The  $x'(t)$  represents the PM values at  $t$ ,  $x_{\text{mean}}'$  and  $x_{\text{std}}'$  are the mean and standard deviation of  $x'(t)$ , respectively. The standardized training set can be expressed as

$$z(t) = \frac{x'(t) - x_{\text{mean}}'}{x_{\text{std}}'}, \tag{13}$$

$$F_{tr}' = \{f_1', f, \dots, f_m'\}, \tag{14}$$

$$f_t' = \left( f_t - \sum_{t=1}^n \frac{f_t}{n} \right) / \sqrt{\frac{\sum_{t=1}^n (f_t - \sum_{t=1}^n \frac{f_t}{n})^2}{n}} \tag{15}$$

$$1 \leq t \leq m, t \in \mathbb{N}$$

In order to adapt to the characteristics of hidden layer input, a data segmentation method is applied to  $F_{tr}'$ . Set the split window length value to  $L$ , and the model input after the split is

$$X = \{X_1, X_2, \dots, X_L\}, \tag{16}$$

$$X_p = \{f_{p'}, f_{p+1}', \dots, f_{m-L+p-1}'\} \quad 1 \leq p \leq L; p, L \in \mathbb{N}. \tag{17}$$

The corresponding theoretical output is

$$Y = \{Y_1, Y_2, \dots, Y_L\}, \tag{18}$$

$$Y_p = \{f_{p+1}', f_{p+2}', \dots, f_{m-L+p}'\}. \tag{19}$$

Next, input the  $X$  to the hidden layers, which contain  $L$  isomorphic LSTM cells connected at the front and back times. The output of  $X$  after passing through the hidden layer is represented as

$$P = \{P_1, P_2, \dots, P_L\}, \tag{20}$$

$$P_p = \text{LSTM}_{\text{forward}}(X_p, C_{p-1}, H_{p-1}), \tag{21}$$

where  $C_{p-1}$  and  $H_{p-1}$  represent the state and output of the previous LSTM cell, respectively;  $\text{LSTM}_{\text{forward}}$  represents the forward calculation method of LSTM cells. If the cell state vector size is set to  $S_{\text{state}}$ , the sizes of both  $C_{p-1}$  and  $H_{p-1}$  vectors are both  $S_{\text{state}}$ . It can be seen that the hidden layer output  $P$ , model input  $X$ , and theoretical output  $Y$  are two-dimensional arrays with dimensions of  $(m-L, L)$ . The mean square error is selected as the error calculation formula, and the loss function of the training process can be expressed as

$$\text{Loss} = \sum_{i=1}^{L(m-L)} (p_i - y_i)^2 / [L(m-L)]. \tag{22}$$

**Application of the trained LSTM model**

Set the minimum loss function as the optimization goal, and given the random seed for network initialization, learning efficiency is  $\eta$ , and training steps, apply the Adam optimization algorithm to continuously update the network weight to obtain the final hidden side network.



**Table 3** Based on LSTM PM time series models training and prediction algorithm

<b>Algorithm1 Training series prediction LSTM neural network</b>	
Input: $F_o, m, L, S_{state}, seed, steps, \eta$	
Output:	
1:	get $F_{tr}, F_{te}$ from $F_o$ by $m$
2:	$F'_{tr} = zscore(F_{tr})$
3:	get $X$ and $Y$ from $F'_{tr}$ by $L$
4:	Create $LSTM_{cell}$ by $S_{state}$
5:	connect $LSTM_{net}$ by $LSTM_{cell}$ and $L$
6:	initialize $LSTM_{net}$ by seed
7:	for each step in 1: steps
8:	$P = LSTM_{forward}(X)$
9:	Loss = $\sum_{i=1}^{L(m-L)} (p_i - y_i)^2 / [L(m-L)]$
10:	update $LSTM_{net}$ by Adam with loss and $\eta$
11:	get $LSTM_{net}$
12:	for each $j$ in 0 : (n-m-1)
13:	$P_{f+j} = LSTM_{net}^*(Y_{f+j})$
14:	append $P_o$ with $P_{f+j}[-1]$
15:	$P_{te} = de\_zscore(P_o)$
16:	error measure $\varepsilon(P_{te}, F_{te}), \varepsilon_e(P_{tr}, F_{tr})$ ,
<b>Algorithm2 Parameter optimization of LSTM prediction model</b>	
Input: $F_o, seed, steps, m, L, S_{state}, \eta, step_L, step_{state}, step_\eta$	
Output:	
1:	Predefine values of seed, steps
2:	Predefine value ranges of $L, S_{state}, \eta$
3:	for each $\eta$ in $\eta_1, \eta_2 \dots \eta_r$ by $step_\eta$
4:	for each $L$ in 2 : $L_{max}$ by $step_L$
5:	for each $S_{state}$ in 2 : $S_{max}$ by $step_{state}$
6:	execute $LSTM_{predict}(F_o, m, L, S_{state}, seed, steps, \eta)$
7:	append results with $L, S_{state}, \eta, \varepsilon(P_{te}, F_{te}), \varepsilon_e(P_{tr}, F_{tr})$
8:	end (for); end (for); end (for)
9:	return results ranked by $\varepsilon(P_{te}, F_{te})$ ,

This section applies the trained LSTM network ( $LSTM_{net}^*$ ) for prediction, and the prediction process adopts an iterative method. First, the last row of data in theoretical output  $Y$  is

$$Y_f = \{ f'_{m-L+1}, f'_{m-L+2}, \dots, f'_m \}. \tag{23}$$

Enter  $Y_f$  into  $LSTM_{net}^*$ , the output result can be expressed as

$$P_f = LSTM_{net}^*(Y_f) = (p_{m-L+2}, p_{m-L+3}, \dots, p_{m+1}). \tag{24}$$

**Table 4** Experiments parameters

LSTM model parameters	Values
Total number of hidden layers	2
Number of LSTM cells per hidden layer	50
Time steps	365
Training iterations	1000
Dropout	0.1
Learn Rate Drop Factor	0.2
Loss function	MSE
Optimization algorithm	Adam
Smoothing method	SG smoothing filter
Gradient threshold	1

The predicted value at time  $m + 1$  is  $p_{m+1}$ . Then, combine the last  $L - 1$  data point of  $Y_f$  and  $P_{m+1}$  into a new row of data

$$Y_f = \{ f'_{m-L+1}, f'_{m-L+2}, \dots, f'_m \}. \tag{25}$$

Enter  $Y_{f+1}$  into  $LSTM_{net}^*$ ; then the predicted value at time  $m + 2$  is  $P_{m+2}$ , and so on. The resulting prediction order is

$$P_o = \{ P_{m+1}, P_{m+2}, \dots, P_n \}. \tag{26}$$

Next, by performing z-score de-normalization on  $P_o$  (represented as  $de\_zscore$ ), and the formula of  $P_{te}$  is the de-normalization as shown in Eq. (27) (Liu et al. 2019), the final prediction sequence corresponding to the test set  $F_{te}$  is obtained as.

$$P_{te} = P_o * x'_{std} + x_{mean}, \tag{27}$$

$$P_{te} = de_{zscore}(P_o) = (p_{m+1}^*, p_{m+2}^*, \dots, p_{m+n}^*), \tag{28}$$

$$P_k^* = \sqrt{\frac{\sum_{t=1}^n (f_t - \frac{\sum_{t=1}^n f_t}{n})^2}{n}} + \sum_{t=1}^n \frac{f_t}{n}, m + 1 \leq k \leq n, k \in \mathbf{N}. \tag{29}$$

Similarly, using each row of  $X$  as model input can obtain a fitting sequence  $P_{tr}$  corresponding to the training set  $F_{tr}$ . The training of LSTM-based PM time series models, prediction algorithms, and parameter optimization of LSTM prediction models are shown in Table 3, and other parameter settings are shown in Table 4.

## Abbreviations

AW	Annual wobbles
AR	Autoregressive models
ARIMA	Autoregressive integrated moving average
ACF	Autocorrelation function
AE	Absolute errors
AAM	Atmospheric angular momentum
CW	Chandler wobbles
EOPs	Earth orientation parameters
ERPs	Earth rotation parameters
FFT	Fast Fourier Transform
FPE	Final prediction error
GPS	Global Position System
IERS	International Earth Rotation and Reference Systems Service
IAA	Institute of Applied Astronomy
JPL	Jet Propulsion Laboratory
LOD	Length of Day
LS	Least squares
LSTM	Long short-term memory
MAE	Mean absolute errors
NASA	National Aeronautics and Space Administration
OAM	Oceanic Angular Momentum
RNNs	Recurrent neural networks
PM	Polar motion
PMX	X coordination of PM
PMY	Y coordination of PM
PACF	Autocorrelation function
PSD	Post-seismic deformation
RS/PC	Rapid Service/Prediction Center
RAS	Russian Academy of Sciences
SG	Savitzky–Golay
SLR	Satellite Laser Ranging
SAW	Semi-Annual wobbles
UT1	Universal Time
UTC	Coordinated Universal Time
USNO	The US Navy Observatory
VLBI	Very Long Baseline Interferometry

## Acknowledgements

This study is under the support of the National Natural Science Foundation of China (NSFC) (Grant Nos. 41874035, 12103035, 42030105). The authors are grateful to IERS for the EOP 14C04 solution. It is worth stating that all the prediction models in this experiment were implemented based on our self-developed software compiled in the MATLAB platform. Interested readers may contact the authors by email.

## Author contributions

CW designed the framework of this research in this paper. CW and PZ designed and derived the theoretical formulations, carried out the observations of the microtremors, and wrote the article. CW performed the numerical calculations. All the authors read and approved the final manuscript.

## Funding

Funding was provided by the National Natural Science Foundation of China (NSFC) (Grant Nos. 41874035, 12103035, 42030105).

## Availability of data and materials

The datasets of IERS EOP 14C04 with a daily interval are available at <https://hpiers.obspm.fr/eopppc/eop/eopc04/>; the Bulletin A can find at <https://www.iers.org/IERS/EN/DataProducts/EarthOrientationData/eop.html>.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China. <sup>2</sup>Time and Frequency Geodesy Center, Wuhan University, 129 Luoyu Road, Wuhan 430079, China.

Received: 24 May 2023 Accepted: 23 September 2023

Published online: 07 October 2023

## References

- Akaike H (1971) Autoregressive model fitting for control. *Annals Inst Stat Math* 23:163–180. <https://doi.org/10.1007/BF02479221>
- Akyilmaz O, Kutterer H (2004) Prediction of Earth rotation parameters by fuzzy inference systems. *J Geod.* <https://doi.org/10.1007/s00190-004-0374-5>
- Alex G, Marcus L, Santiago F, Roman B, Horst B, Jürgen S (2018) A novel connectionist system for unconstrained handwriting recognition. *IEEE Trans Pattern Anal Mach Intell* 31:855–868. <https://doi.org/10.1109/TPAMI.2008.137>
- Babcock AK, Wilkins GA (1989) The Earth rotation and reference frames for Geodesy and Geodynamics. *Geophys J Int.* <https://doi.org/10.1111/j.1365-246X.1989.tb04457.x>
- Bizouard C (2005) Influence of the earthquakes on the polar motion with emphasis on the Sumatra event. In: *Journées Systèmes de Référence Spatio-Temporels, Proceedings*. pp 229–232
- Bogusz J, Brzezinski A, Kosek W, Nastula J (2015) Earth rotation and geodynamics. *Geodesy Cartogr* 64(2):201–242. <https://doi.org/10.1515/geocart-2015-0013>
- Börger L, Schindelegger M, Dobslaw H, Salstein D (2023) Are ocean reanalyses useful for earth rotation research? *Earth Space Sci.* <https://doi.org/10.1029/2022ea002700>
- Box GEP, Jenkins GM, Reinsel GC, Ljung GM (1976) *Time series analysis: forecasting and control*. Holden Day, San Francisco, pp 88–125
- Brockwell PJ, Davis RA (1997) *Introduction to time series and forecasting*, 2nd edn. Springer, New York, pp 81–106
- Chandler S (1981) On the variation of latitude. *Astron J* 156–61. <https://doi.org/10.1038/056040a0>
- Din AZU, Ayaz Y, Hasan M, Khan J, Salman M (2019) Bivariate short-term electric power forecasting using LSTM network. In: *2019 International Conference on Robotics and Automation in Industry (ICRAI)*. IEEE, pp 1–8
- Dobslaw H, Dill R, Grötzsch A, Brzeziński A, Thomas M (2010) Seasonal polar motion excitation from numerical models of atmosphere, ocean, and continental hydrosphere. *J Geophys Res.* <https://doi.org/10.1029/2009jb007127>
- Egger D (1992) Neuronales Netz Prädiziert Erdrotation. *AVN-Allgemeine Vermessungs-Nachrichten* 99:517–524
- Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with LSTM. *Neural Comput* 12:2451–2471. <https://doi.org/10.1049/cp:19991218>
- Gers FA, Schmidhuber J (2000) Recurrent nets that time and count. In: *IJCNN 2000, Neural Networks*. pp 189–194
- Gou J, Kiani Shahvandi M, Hohensinn R, Soja B (2021) Ultra-short-term prediction of LOD using LSTM neural networks. In: *EGU General Assembly Conference, Vienna, Austria.* <https://doi.org/10.5194/egusphere-egu21-2308>
- Graves A, Schmidhuber J (2005) Framewise phoneme classification with bidirectional LSTM and Other Neural Network architectures. *Neural Netw* 18:602–610. <https://doi.org/10.1016/j.neunet.2005.06.042>
- Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J (2017) LSTM: a search space Odyssey. *IEEE Trans Neural Netw Learn Syst* 28(10):2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Gross RS (2000) the excitation of the Chandler wobble. *Geophys Res Lett* 27(15):2329–2332. <https://doi.org/10.1029/2000gl011450>
- Gross RS (2007) Earth rotation variations-long period. In: Herring T (ed) *Treatise on geophysics*, vol 3. Elsevier, Amsterdam, pp 239–294. <https://doi.org/10.1016/B978-044452748-6/00057-2>
- Guo JY, Li YB, Dai CL, Shum CK (2013) A technique to improve the accuracy of Earth orientation prediction algorithms based on least squares

- extrapolation. *J Geodyn* 70:36–48. <https://doi.org/10.1016/j.jog.2013.06.002>
- Hochreiter S (1991) Untersuchungen zu dynamischen neuronalen Netzen. Diploma thesis, Germany, Technische Universität München
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hochreiter S, Bengio Y, Frasconi P, Schmidhuber J (2001) Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. In: JF Kolen (ed) IEEE Press, Los Alamitos. <https://doi.org/10.1109/9780470544037.ch14>
- Joachim H (2004) Low-frequency variations chandler and annual wobbles of polar motion as observed over one century. *Surv Geophys* 25:1–54. <https://doi.org/10.1023/B:GEO.0000015345.88410.36>
- Karevan Z, Suykens JAK (2020) Transductive LSTM for time-series prediction: an application to weather forecasting. *Neural Netw* 125:1–9. <https://doi.org/10.1016/j.neunet.2019.12.030>
- Kenyon SC, Pacino MC, Mart U (2012) Geodesy for planet earth. In: Kenyon SC, Pacino MC, Mart U (eds) 2009 IAG Symposium, Buenos Aires, Argentina, August 31–September 4. Springer, pp 513–520. [www.iag2009.com.ar](http://www.iag2009.com.ar)
- Kosek W (2002) Autocovariance prediction of complex-valued polar motion time series. *Adv Space Res* 30:375–380. [https://doi.org/10.1016/S0273-1177\(02\)00310-1](https://doi.org/10.1016/S0273-1177(02)00310-1)
- Kosek W, Popiński W (2005) Forecasting of pole coordinates data by combination of the wavelet decomposition and autocovariance prediction. In: Journées 2005 Systemes de Reference Spatio-Temporels, pp 139–140
- Kosek W, McCarthy DD, Johnson TJ, Kalarus M (2004) Comparison of polar motion prediction results supplied by the IERS Sub-bureau for Rapid Service and predictions and results of other prediction methods. In: Finkelstein A CN (ed) the Journées 2003 "Systemes deréférence spatio-temporels, Petersburg, pp 164–169
- Kosek W, Kalarus M, Niedzielski T (2008) Forecasting of the earth orientation parameters: comparison of different algorithms. In: Capitaine N (ed) Proceedings of the journées 2007, Paris, pp 155–158
- Liao DC, Wang QJ, Zhou YH, Liao XH, Huang CL (2012) Long-term prediction of the Earth Orientation Parameters by the artificial neural network technique. *J Geodyn* 62:87–92. <https://doi.org/10.1016/j.jog.2011.12.004>
- Liu Y, Guan L, Hou C, Han H, Liu Z, Sun Y, Zheng M (2019) Wind power short-term prediction based on LSTM and discrete wavelet transform. *Appl Sci*. <https://doi.org/10.3390/app9061108>
- Mccarthy DD, Luzum BJ (1991) Prediction of earth orientation. *Bulletin Géodésique* 65(1):18–21
- Petit G, Luzum B (2010) IERS Conventions 2010. IERS Technical Note 36. Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main
- Ren Z, Huangfu Y, Xie R, Ma R (2020) Modeling of Proton Exchange Membrane Fuel Cell Based on LSTM Neural Network. Paper presented at the 2020 Chinese Automation Congress (CAC)
- Schaffer AL, Dobbins TA, Pearson SA (2021) Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Med Res Methodol* 21(1):58. <https://doi.org/10.1186/s12874-021-01235-8>
- Schmidhuber J (2015) Deep Learning in neural networks an overview. *Neural Netw* 61:85–117. <https://doi.org/10.1016/J.NEUNET.2014.09.003>
- Schuh H, Nagel S, Seitz T (2001) Linear drift and periodic variations observed in long time series of polar motion. *J Geod* 74:701–710. <https://doi.org/10.1007/s001900000133>
- Schuh H, Ulrich M, Egger D, Müller J, Schwegmann W (2002) Prediction of Earth orientation parameters by artificial neural networks. *J Geod* 76(5):247–258. <https://doi.org/10.1007/s00190-001-0242-5>
- Souriau A (1986) The Influence of Earthquakes on the Polar Motion. In: Cazenave A (ed) Earth Rotation: Solved and Unsolved Problems. Springer Netherlands, Dordrecht %@ 978-94-009-4750-4, pp 229–240. [https://doi.org/10.1007/978-94-009-4750-4\\_16](https://doi.org/10.1007/978-94-009-4750-4_16)
- Sun Z, Xu T (2012) Prediction of earth rotation parameters based on improved weighted least squares and autoregressive model. *Geodesy Geodyn* 3(3):57–64. <https://doi.org/10.3724/sp.J.1246.2012.00057.1>
- Sun Z, Xu T, Jiang C, Yang Y, Jiang N (2019) An improved prediction algorithm for Earth's polar motion with considering the retrograde annual and semi-annual wobbles based on least squares and autoregressive model. *Acta Geod Geophys* 54(4):499–511. <https://doi.org/10.1007/s40328-019-00274-4>
- Suvorkin VV, Kurdubov SL (2015) I.S. G GNSS Processing in Institute of Applied Astronomy RAS. In: Malkin Z., Capitaine N (eds) In: Proceeding of the Journées 2014 "Systemes de référence spatio-temporels": Recent developments and prospects in ground-based and space astrometry, Petersburg, Russia. pp 261–262
- Wang G, Liu L, Su X, Liang X, Yan H, Tu Y, Li Z, Li W (2016) Variable Chandler and annual wobbles in Earth's Polar motion during 1900–2015. *Surv Geophys* 37(6):1075–1093. <https://doi.org/10.1007/s10712-016-9384-0>
- Wang J, Jiang W, Li Z, Lu Y (2021) A new multi-scale sliding window LSTM framework (MSSW-LSTM): a case study for GNSS time-series prediction. *Remote Sens*. <https://doi.org/10.3390/rs13163328>
- Xu X, Zhou Y (2015) EOP prediction using least square fitting and autoregressive filter over optimized data intervals. *Adv Space Res* 56(10):2248–2253. <https://doi.org/10.1016/j.asr.2015.08.007>
- Xu XQ, Zhou YH, Liao XH (2012) Short-term earth orientation parameters predictions by combination of the least-squares, AR model and Kalman filter. *J Geodyn* 62:83–86. <https://doi.org/10.1016/j.jog.2011.12.001>
- Yu Y, Si X, Hu C, Zhang J (2019) A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput* 31(7):1235–1270. [https://doi.org/10.1162/neco\\_a\\_01199](https://doi.org/10.1162/neco_a_01199)
- Zhang XH, Wang QJ, Zhu JJ, Zhang H (2012) Application of general regression neural network to the prediction of LOD change. *Chinese J Astron Ast* 36(1):86–96. <https://doi.org/10.1016/j.chinastron.2011.12.010>
- Zhang X, Liang X, Zhiyuli A, Zhang S, Xu R, Wu B (2019) AT-LSTM: an attention-based LSTM model for financial time series prediction. *IOP Conf Ser Mater Sci Eng*. <https://doi.org/10.1088/1757-899x/569/5/052037>
- Zharkov VN, Molodensky SM (1996) On the Chandler wobble of Mars. *Planet Space Sci* 44(11):1457–1462. [https://doi.org/10.1016/S0032-0633\(96\)00052-9](https://doi.org/10.1016/S0032-0633(96)00052-9)

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)