TU Delft

Delft University of Technology

Short- and long-term prediction of length of day time series using a combination of MCSSA and ARMA

Shirafkan, Shayan; Sharifi, Mohammad Ali; Modiri, Sadegh; Belda, Santiago; Khazraei, Seyed Mohsen; Amiri-Simkooei, Alireza

DOI
10.1186/s40623-025-02166-0

Publication date
2025

Document Version
Final published version

Published in
Earth, Planets and Space

Citation (APA)
Shirafkan, S., Sharifi, M. A., Modiri, S., Belda, S., Khazraei, S. M., & Amiri-Simkooei, A. (2025). Short- and long-term prediction of length of day time series using a combination of MCSSA and ARMA. *Earth, Planets and Space*, *77*(1), Article 35. https://doi.org/10.1186/s40623-025-02166-0

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Earth, Planets and Space

FULL PAPER Open Access

Short- and long-term prediction of length of day time series using a combination of MCSSA and ARMA

Shayan Shirafkan[1], Mohammad Ali Sharifi[1*], Sadegh Modiri[2], Santiago Belda[3], Seyed Mohsen Khazraei[4] and Alireza Amiri-Simkooei[5]

Abstract

Accurately predicting Earth's rotation rate, as represented by Length of Day (LOD) variations, is essential for applications such as satellite navigation, climate studies, geophysical research, and disaster prevention. However, predicting LOD is challenging due to its sensitivity to various geophysical and meteorological factors. Current methods, including statistical approaches, often struggle with short-term forecasting accuracy. In this study, we use Monte Carlo Singular Spectrum Analysis (MCSSA) to distinguish between deterministic and non-deterministic components within the LOD time series. The deterministic components are extended using the SSA prediction algorithm. To enhance robustness, we refine Allen and Smith's methodology (testing significance of eigenmodes against an autoregressive (AR) (1) noise null hypothesis) by integrating an autoregressive moving average (ARMA) model to account for noise, providing valuable insights into the non-deterministic behaviors present in the series. We comprehensively evaluate our methodology through a comparative analysis. For long-term prediction (365 days), we compare our method against the combined LS and autoregressive (AR) method. For short-term prediction (next 10 days), we compare it against the results of the second Earth Orientation Parameters Prediction Comparison Campaign (second EOP PCC). Using the IERS 20 C04 time series, our hybrid model demonstrates a superior long-term prediction accuracy with a mean absolute error (MAE) of 0.201 ms/day on the 365th day. Additionally, the short-term prediction performance is comparable to the second EOP PCC results. These results illustrate that the proposed method efficiently predicts LOD, showing significant improvement in long-term accuracy and robustness in short-term forecasting.

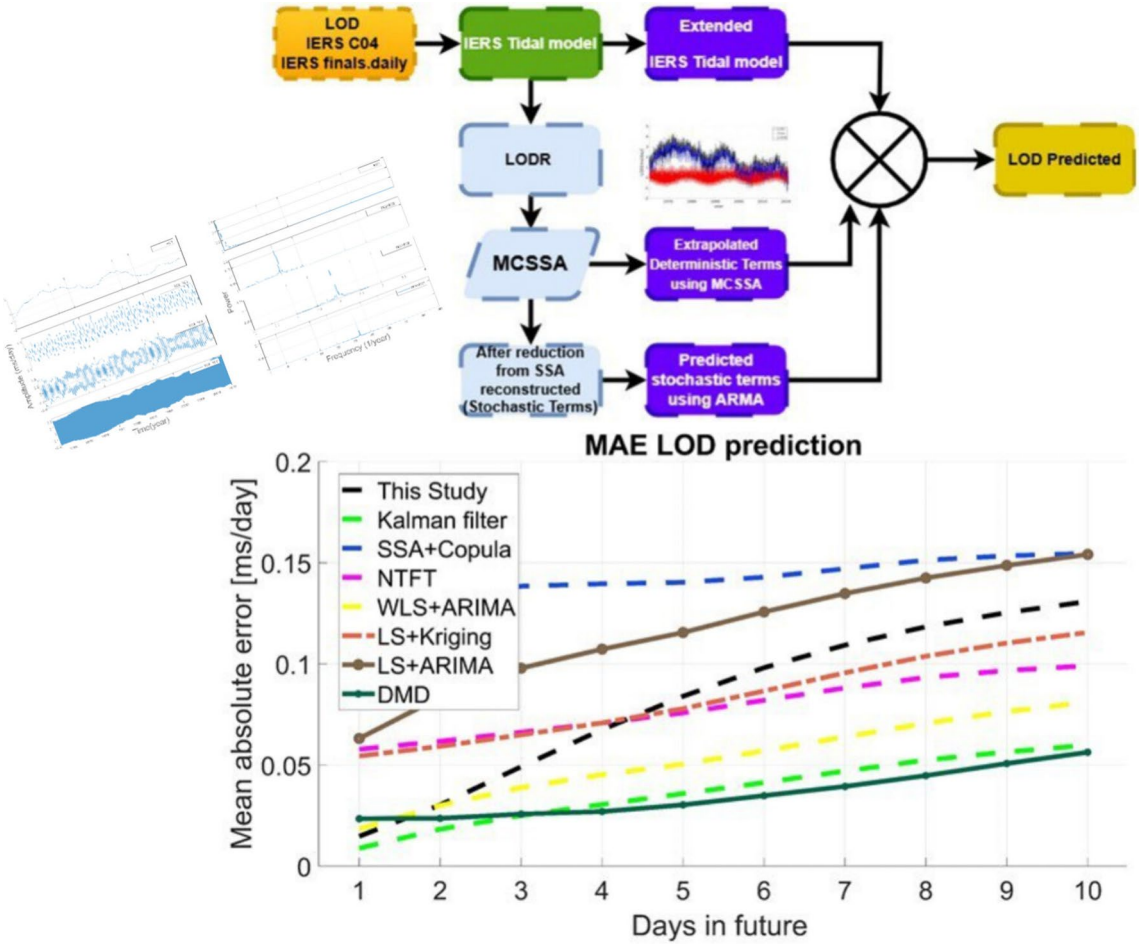
Keywords Prediction, Length of day, Singular spectrum analysis, Monte Carlo, ARMA, Least square

*Correspondence:
Mohammad Ali Sharifi
sharifi@ut.ac.ir
Full list of author information is available at the end of the article

Springer Open

© The Author(s) 2025. **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/.

Graphical abstract



1 Introduction

Earth Orientation Parameters (EOP) characterize the movement of Earth within space and relate points between terrestrial and celestial reference systems. The variation of the celestial intermediate pole concerning the terrestrial frame is referred to as polar motion, while its displacement concerning the celestial frame is termed celestial pole motion (Seitz and Schuh 2010). Additionally, UT1-UTC (denoted as dUT1) represents the interpretation of Earth's rotational angle (Aoki et al. 1982). Alongside the five EOP, the Length of Day (LOD) serves to delineate variations in Earth's rotational rate (Freedman et al. 1994).

EOP serve as foundational components in various fields, including astronomy, geodesy, geophysics, meteorology, and space navigation (Barnes et al. 1983; Mathews et al. 1991; Chen and Wilson 2005; Gross 2007; Belda et al. 2018; Modiri et al. 2021; Malkin et al. 2022; Raut et al. 2022). They play crucial roles in establishing accurate astronomical and geodetic reference frames, conducting geophysical and meteorological research, and facilitating precise space navigation systems (Altamimi et al. 2023; Hellmers et al. 2022). The determination of EOP values is achieved through diverse space geodetic techniques such as the Global Navigation Satellite System (GNSS), Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS), Satellite Laser Ranging (SLR), and Very Long Baseline Interferometry (VLBI) (Coulot et al. 2010; Angermann et al. 2010). Notably, VLBI stands out as the sole technique capable of monitoring the complete set of EOP (Schuh and Behrend 2012). These observations are consolidated by the International Earth Rotation and Reference Systems Service (IERS), which subsequently disseminates EOP data to

users (Dick and Richter 2004). However, due to delays in data availability, particularly when observations require rigorous processing and validation, the prediction of EOP becomes necessary.

The IERS conducted two Earth Orientation Parameter Prediction Comparison Campaigns (EOP PCC) to evaluate various prediction techniques across different time intervals. The first campaign, which took place between 2005 and 2009, involved active participation from several groups. This initiative revealed that no single prediction method excels for all EOP or prediction intervals (Kalarus et al. 2010). Instead, hybrid models were recommended as a favorable approach. Due to the high demand for accurate EOP prediction, the second EOP PCC commenced on September 1, 2021, and extended until December 31, 2022, though it remains open to receiving EOP predictions from participants (Śliwińska-Bronowicz et al. 2024). During this campaign, different groups participated to explore novel prediction methods and enhance existing approaches.

Among EOP, LOD is the most challenging parameter to predict, as variations in LOD stem from gravitational influences exerted by external celestial bodies, such as the Moon and Sun, as well as geophysical processes within Earth's layers and meteorological factors. Effective Angular Momentum (EAM), which encapsulates contributions from atmospheric, oceanic, and hydrological angular momentum, has been shown to significantly improve EOP prediction accuracy (Śliwińska-Bronowicz et al. 2024). Consequently, analyzing the LOD time series can enhance our comprehension of various geophysical and meteorological phenomena within the Earth system. One such factor is the El Niño-Southern Oscillation (ENSO), a climate pattern characterized by anomalous warming of surface waters in the eastern tropical Pacific Ocean (Chao 1989; Modiri et al. 2020; Raut et al. 2022). Many other studies have demonstrated the impact of this climatic index on LOD signals (Holton and Dmowska 1989; Gross et al. 1996; Agnew 2024; Shahvandi et al. 2024).

Several methodologies have been developed to improve the accuracy of LOD predictions. Based on findings from the first EOP PCC, prediction algorithms can be categorized into single and hybrid methods, leveraging past LOD data to forecast values for short-, medium-, and long-term intervals. In the single technique category, various approaches have been explored, including auto-covariance (AC) analysis (Kosek 2002), wavelet decomposition (Akyilmaz and Kutterer 2004), and neural networks (Schuh et al. 2002). These methods were examined and utilized during the campaign. Additionally, non-parametric techniques have been integrated into hybrid models. Examples of hybrid models include combinations of least squares (LS) and auto-regressive (AR) methods (Kosek et al. 2004; Kosek et al. 2007; Xu and Zhou 2015; Wu et al. 2021), LS extrapolation combined with ARIMA (Guo et al. 2013), and combinations of least squares, AR model, and Kalman filter.

In addition to the categorization observed in the first EOP PCC, insights from the second EOP PCC have further classified prediction methods into four main groups. Group 1 utilizes a combination of least squares (LS) and auto-regressive (AR) techniques. Group 2, which incorporates EAM as input alongside least squares (LS) and auto-regressive (AR) methods, has demonstrated the benefits of leveraging EAM for EOP predictions (Śliwińska-Bronowicz et al. 2024). Group 3 employs machine learning (ML) approaches for prediction tasks (Guessoum et al. 2022; Kiani Shahvandi et al. 2023; Gou et al. 2023). Finally, Group 4 encompasses miscellaneous methods, including those not fitting into the aforementioned categories. In Group 4, recent advancements have seen the application of hybrid methods incorporating singular spectrum analysis (SSA) and autoregressive moving average (ARMA), or Copula techniques for predicting EOP time series (Modiri et al. 2018; Shen et al. 2018; Modiri et al. 2021; Jin et al. 2021). For instance, Modiri et al. (2020) introduced a novel technique aimed at enhancing the short-term prediction accuracy of LOD time series. This approach leveraged Copula-based analysis to model dependencies between effective angular momentum (EAM) and LOD, resulting in improved predictive capabilities. Our study contributes to Group 4, focusing on alternative methodologies that explore deterministic and stochastic components of the LOD signal independently. This study employs SSA, a non-parametric method capable of capturing arbitrary statistical processes, to characterize the deterministic behavior of dominant frequencies in LOD time series. By applying SSA, we extract deterministic components after removing the tidal effects using IERS models. The residual, exhibiting random behavior, is subsequently modeled using autoregressive moving average (ARMA) techniques. SSA offers versatility in time series analysis, accommodating variations in signal amplitude and phase dynamics. SSA has found wide-ranging applications across disciplines, including oceanography (Kondrashov and Berloff 2015), climatology (Rial et al. 2013), and geodesy (Yi and Sneeuw 2021). Unlike parametric methods, SSA is not confined to predefined basis functions like sines and cosines, though discerning between deterministic and stochastic behaviors remains challenging within non-parametric frameworks. To address this, Allen and Smith (1996) introduced the Monte Carlo SSA (MCSSA) method, specifically designed to statistically differentiate signals from first-order autoregressive noise (AR(1)). Subsequently, Khazraei and Amiri-Simkooei

(2019) extended this concept to accommodate various noise models. MCSSA employs significance testing on detected signals, safeguarding against misinterpretation of random fluctuations as deterministic signals. Khazraei and Amiri-Simkooei (2019) demonstrated the efficacy of MCSSA in extracting annual and semiannual signals from GNSS position time series contaminated with flicker and white noise. In our investigation, we propose to leverage the MCSSA methodology to enhance the accuracy of LOD prediction. To assess the capability of the proposed method, we consider the same conditions as the second EOP PCC to demonstrate the effectiveness of the presented method for short-term (10 days) and long-term (365 days) prediction. We compare the prediction results with those of existing techniques, specifically the second EOP PCC and LS+AR, for short-term and long-term prediction, respectively. While EAM is not integrated into our current framework, the results demonstrate that the proposed approach can efficiently predict LOD. Future studies may combine our approach with EAM to further enhance predictive accuracy.

2 Methods

The Length of Day (LOD) prediction algorithm presented in this study integrates deterministic and stochastic approaches. We utilize MCSSA as a deterministic tool to investigate oscillations within LOD signals. Subsequently, the discrepancy between the original LOD time series and the reconstructed time series obtained through MCSSA is modeled using Autoregressive Moving Average (ARMA) techniques. Singular Spectrum Analysis (SSA), introduced by Broomhead and King (1986), is a non-parametric method that leverages the time domain of data to extract information from short time series containing noise, without requiring prior knowledge of the dynamics affecting the time series. This technique decomposes the time-delayed embedding of a time series into a set of data-adaptive orthogonal components, which can be categorized into trends, oscillation patterns, and noise. In the subsequent sections, we provide a detailed overview of the theoretical framework supporting our proposed methodology.

2.1 SSA algorithm

For a time series of m observations $\mathbf{y} = \left[y(1), y(2), ..., y(m)\right]^\top$ sampled at $t_1, t_2, ..., t_m$, the SSA algorithm can be implemented through the following steps:

- The lag-covariance matrix can be formed using the formulations proposed by (Broomhead and King 1986) and (Vautard and Ghil 1989). Their corresponding lag-covariance matrices are denoted by C_{BK} and C_{VG}, respectively. To form C_{BK}, one needs to first construct an $m' \times M$ trajectory matrix, sliding an M -length window across the time series and placing the constructed sub-vectors in consecutive rows. m' and the window length M are the SSA parameters, where $1 < M < m$ and $m' = m - M + 1$.

$$\mathbf{Y} = \begin{bmatrix} y(1) & y(2) & \cdots & y(M) \\ y(2) & y(3) & \cdots & y(M+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(m'-1) & y(m') & \cdots & y(m-1) \\ y(m') & y(m'+1) & \cdots & y(m) \end{bmatrix} \quad (1)$$

The matrix C_{BK} can then be formed using the following equation:

$$C_{BK} = \frac{1}{m'}\mathbf{Y}^\top\mathbf{Y} \quad (2)$$

The matrix C_{VG} can also be formed using the well-known unbiased covariance estimator. Accordingly, the entry in the ith row and jth column of C_{VG} $(i, j = 1, ..., M)$ is given by:

$$(C_{VG})_{i,j} = c(\tau) = \frac{1}{m-\tau}\sum_{k=1}^{m-\tau} y(k)y(k+\tau) \quad (3)$$

where $\tau = |i - j|$ is the time lag. This formulation leads to a Toeplitz lag-covariance matrix. For more information on C_{BK} and C_{VG}, refer to (Allen and Smith 1996) and (Ghil and Taricco 1997).

- The second step is to obtain the eigenvalues and eigenvectors of the log-covariance matrix C using eigenvalue decomposition:

$$\Lambda = \mathbf{E}^\top C\mathbf{E} \quad (4)$$

where Λ is a diagonal matrix of M eigenvalues λ_k $(k = 1, ..., M)$ and $\mathbf{E}$ is an $M \times M$ matrix with the corresponding eigenvectors e_k $(k = 1, ..., M)$ as its columns. The eigenpairs (λ_k, e_k) are sorted according to the variance λ_k in the direction of the orthogonal eigenvectors e_k.

- The third step involves computing the reconstructed components (RCs) corresponding to each kth eigenmode $(k = 1, \ldots, M)$. The entries of the kth principal component (PC) are calculated using the following equation:

$$a_k(i) = \sum_{j=1}^{M} y(i+j-1)e_k(j), \quad 1 \le i \le m' \quad (5)$$

The entries of the kth RC then read:

$$r_k(t) = \frac{1}{M_t} \sum_{j=L_t}^{U_t} a_k(t-j+1)e_k(j), \quad 1 \le i \le m \tag{6}$$

where M_t, L_t, and U_t are defined as:

$$(M_t, L_t, U_t) = \begin{cases} (t, 1, t), & 1 \le t \le M-1 \\ (M, 1, M), & M \le t \le m' \\ (m-t+1, t-m+M, M), & m \le t \le m' \end{cases}. \tag{7}$$

As demonstrated by the preceding equations, the choice of window length (M) is crucial to the SSA algorithm. In this article, the proposed method by (Modiri et al. 2020) is used to determine the window's length.

2.2 Monte-Carlo SSA (MCSSA) via ARMA model

MCSSA is recognized as a method to assess the significance of SSA results. In SSA, the set of eigenvectors and eigenvalues are usually arranged in descending order of eigenvalues. However, this informal classification is not necessarily equivalent to the meaningful order of the components. Only by performing a statistical test against a null hypothesis can the significant deterministic behavior be detected. In the Monte Carlo test proposed by Allen and Smith (1996) for SSA, the variance absorbed by the eigenvalues of the data is compared with the variance of a set of surrogate data realized on the assumption of AR(1) noise. The MCSSA process consists of three steps: estimating the noise model parameters, generating surrogate data, and constructing confidence intervals (Allen and Smith 1996).

- The first step is to estimate the unknown parameters of the ARMA noise model. The ARMA model can be written as follows:

$$y(t) = \mu + e(t) + \sum_{i=1}^{p} \beta_i y(i-t) + \sum_{i=1}^{q} \theta_i e(i-t) \tag{8}$$

  where β_i are the Autoregressive (AR) coefficients, θ_i are the Moving Average (MA) coefficients, μ is the constant term, and $e(t)$ is the white noise series.
- The second step is to generate surrogate data simulating noise using the estimated parameters from the previous step and the ARMA model. These simulated realizations are called surrogate data and are denoted by (y_{Rd}), where ($d = 1, \ldots, D$), and (D) is the number of surrogate datasets.
- The third step involves forming confidence intervals. Allen and Smith (1996) proposed various approaches for constructing confidence intervals for data using a set of surrogated covariance matrices. One of these possible ways is to project each of the matrices C_{Rd}, $d = 1, ..., D$, onto the original data's eigenvectors E using the decomposition of the log-covariance matrix as in equation 9:

$$\Lambda_{(Rd)}^{(E)} = \mathbf{E}^{\top} C_{Rd} \mathbf{E} \tag{9}$$

where $\lambda_{(k)}^{(Rd)}$, $k = 1, \ldots, M$, represent the diagonal entries of $\Lambda_{(Rd)}^{(E)}$ for each surrogate realization between $d = 1$ and $d = D$. The statistical distribution obtained from a set of Monte Carlo simulations results in confidence intervals. If for each successive realization at $d = 1, \ldots, D$, we denote the elements of the principal diagonal of each of the matrices $\Lambda_{(Rd)}^{(E)}$ by $\lambda_{(k)}^{(Rd)}, k = 1, \ldots, M$, assuming zero for the kth component, a different eigenvalue is obtained from different surrogate data. If the number of surrogate realizations is large enough (large D), confidence intervals can be effectively established. These confidence intervals are based on the experimental distribution of surrogate data. The eigenvalues of the data are considered significant with a confidence level of $((1-\alpha))$ if each eigenvalue is outside the $((1-\alpha))$ middle eigenvalues corresponding to the successive realizations. In other words, the acceptance range of the null hypothesis for each eigenvalue will be the interval between $((\alpha/2))$ and $((1-\alpha/2))$ of the corresponding eigenvalues. This means that the eigenvalues of the data do not differ significantly from the pure color noise simulated in the surrogate realizations under the null hypothesis. For example, the range of 2.5% above and below the eigenvalues of the surrogate values, for each component ((k)th component), specifies the rejection area of the null hypothesis for that component at the 95% confidence level. In fact, if the eigenvalue (k) of the data ((λ_k)) is outside the range containing 95% of the noise, it can be assumed that this eigenvalue of the data is significantly different from the null hypothesis. This is because the noise model considers the nature of this component. Otherwise, that particular component cannot be considered significantly different from the assumed zero noise.

2.3 Error analysis

In this study, the mean absolute error (MAE) is utilized to evaluate the accuracy of EOP predictions. The formula is given by

$$\mathrm{MAE}_l = \frac{1}{N} \sum_{i=1}^{N} \left| P_i^l - O_i^l \right| \tag{10}$$

where P_i^l represents the predicted value of the ith prediction day, O_i^l denotes the corresponding observed value, N stands for the number of samples in the test set, and l indicates the prediction span.

3 Data and implementation details

3.1 Data description

This research utilized LOD data from the following sources:

(1) C04 14 time series available at https://hpiers.obspm.fr/iers/eop/eopc04_14/eopc04_IAU2000.62-now.
(2) finals.daily file provided by IERS Rapid Service/Prediction Center (RS/PC) at U.S. Naval Observatory. This series is available at https://maia.usno.navy.mil/ser7/finals.daily.extended.

To be consistent with the 2nd EOP PCC, we have combine the C04 14 and final.daily to create LOD time series to be used as input for the prediction model as follows:

$$LOD_{\text{date}} = \{C04_{\text{date}-n}, C04_{\text{date}-n+1}, \ldots, C04_{\text{date}-30}, \text{final.daily}_{\text{date}-29}, \ldots, \text{final.daily}_{\text{date}}\}$$

where 'date' represents the starting epoch of prediction (shown in Fig. 1). The LOD encompasses various factors, such as the influence of solid Earth tides, which have periods ranging from 5 days to 18.6 years. These tides are effectively modeled, and we utilize the solid tide model recommended by the IERS Conventions 2010 to remove them (Petit and Luzum 2010). The resulting LOD time series without these effects was referred to as the LODR time series.

Figure 2 illustrates the time series of LOD (black curve), solid Earth plus ocean tide terms (red curve), and the LODR (blue curve).

3.2 MCSSA analysis of LOD time series

MCSSA proposes an analysis routine to discriminate deterministic and non-deterministic variations in the LODR time series. Non-deterministic variations, as proposed in the current investigation, can be modeled using the ARMA model. First, the linear trend and well-known periodic patterns in the LODR data, including signals with annual and semi-annual, and seasonal terms, are subtracted from the LODR time series and it considered as noise behavior (Wu et al. 2021). The Akaike Information Criterion (AIC) model was applied to identify the optimal order of ARMA noise, with orders $p = 1$ and $q = 6$ determined to be optimal (Fig. 3). The unknown parameters of the ARMA model are estimated using the maximum likelihood method.

The next step of MCSSA involves conducting singular spectrum analysis on the LODR time series and identifying dominant eigenvalues that statistically differ from the estimated ARMA model. This study has been conducted for both long-term and short-term prediction.

3.3 Long-term prediction

The window length of the SSA is set to 365 days, as the annual signal dominates the LODR time series (Modiri et al. 2020). A significance hypothesis test for eigenvalues is then performed, forming 95% confidence intervals with an ensemble of 1000 surrogate data as different realizations of the estimated ARMA model. Figure 4 illustrates the results of this step.

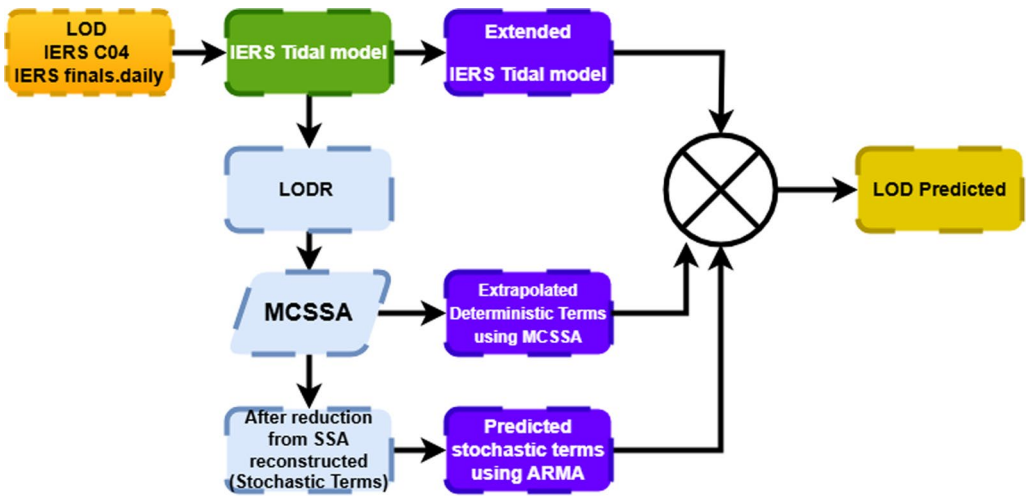


Fig. 1 Flowchart illustrating the LOD prediction process employed in this study

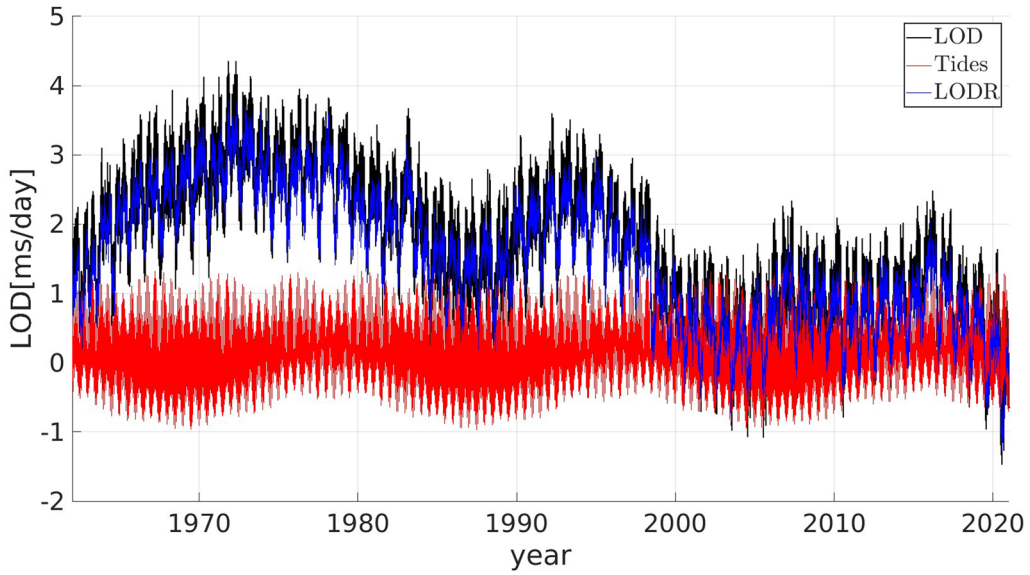


Fig. 2 LOD time series (black), solid Earth plus ocean tide terms (red), and LODR time series (blue)

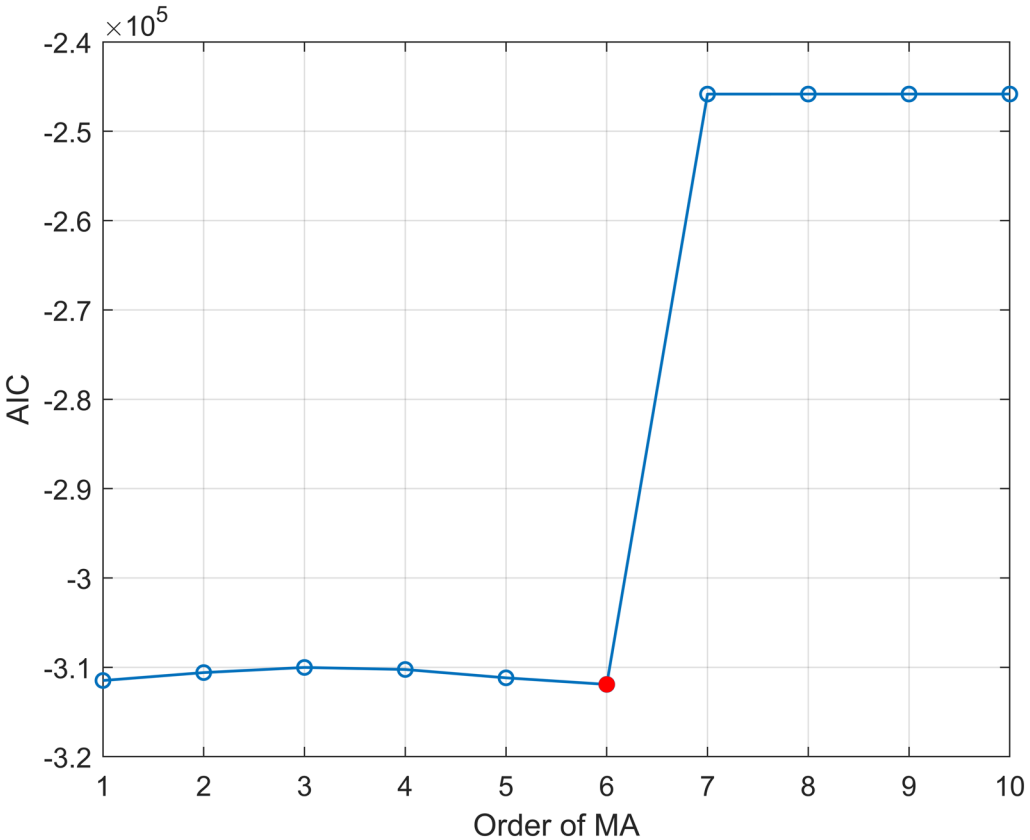


Fig. 3 AIC values for ARMA noise model with AR order set to 1 and MA order varied from 1 to 10

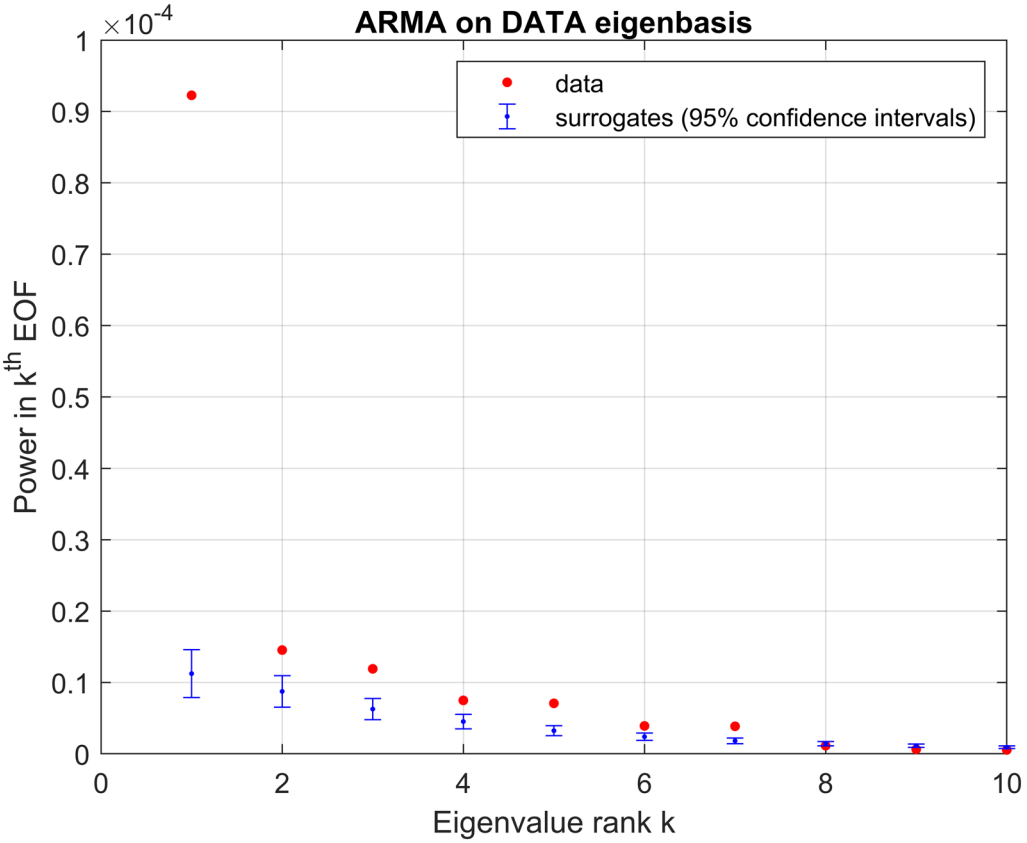


Fig. 4 MCSSA algorithm results for the first 10 reconstructed components by SSA. The data eigenvalues (red dots) and their 95% corresponding confidence intervals (blue error bars) are plotted in their rank order. Confidence intervals are constructed from an ensemble of 1000 surrogate data generated as realizations of the estimated ARMA model

According to Fig. 4, the first seven reconstructed components is significant against ARMA noise. This indicates that the non-linear trend, annual, semi-annual, and 13-day frequencies have the highest contribution to the SSA prediction algorithm. Additionally, by reducing the reconstructed components of the LODR time series and calculating the root mean square (RMS) of the time series after removing each of these components, the results show that the first seven components contribute the most to the formation of the LODR time series (see Fig. 5).

3.3.1 Combination of SSA with ARMA for LOD time series prediction

The combination of SSA with ARMA consists of six steps:

(1) Assuming the length of the prediction interval is n, a zero vector with dimensions n is added to the end of the LODR time series.
(2) Applying the SSA algorithm to the time series obtained in the previous step. After reconstructing the components, this process continues until the values of RC_1 converge. The convergence condition is satisfied when the RMS between the values of R of two consecutive repetitions is less than 0.01 ms. This threshold is derived empirically, based on experimental tuning for prediction convergence.
(3) In the third step, the prediction algorithm is also given the RC_2 values. To achieve this, the second step algorithm iterates until the $RC_1 + RC_2$ values converge.
(4) The process continues until the first seven components (identified by the proposed Monte Carlo method) are included.
(5) ARMA models are fitted to the residual time series, which consists of the eighth-to-last SSA components.
(6) The values obtained from the SSA and ARMA prediction models are combined.

After performing the statistical test associated with the Monte Carlo method, it was determined that nonlinear trends and annual, semi-annual, and 13-day frequencies are statistically significant against ARMA noise. The SSA

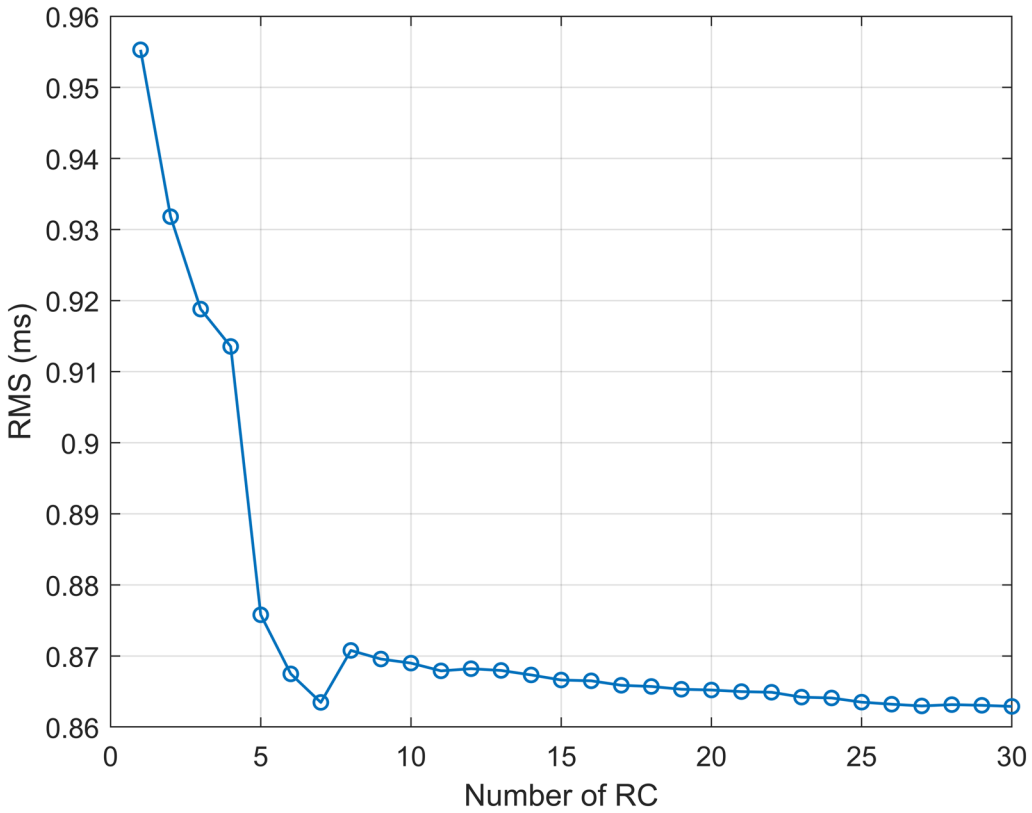


Fig. 5 RMS of the residuals after subtracting the signals derived from SSA for the first 30 components

prediction algorithm requires the selection of the optimal and appropriate components for accurate prediction. Therefore, in the prediction stage, nonlinear trends and significant frequencies are employed. The SSA algorithm was implemented on the LODR time series between January 1, 2015, and December 31, 2018, for the first forecast period. The first seven components were employed to forecast the next 365 days. Figure 6 illustrates the

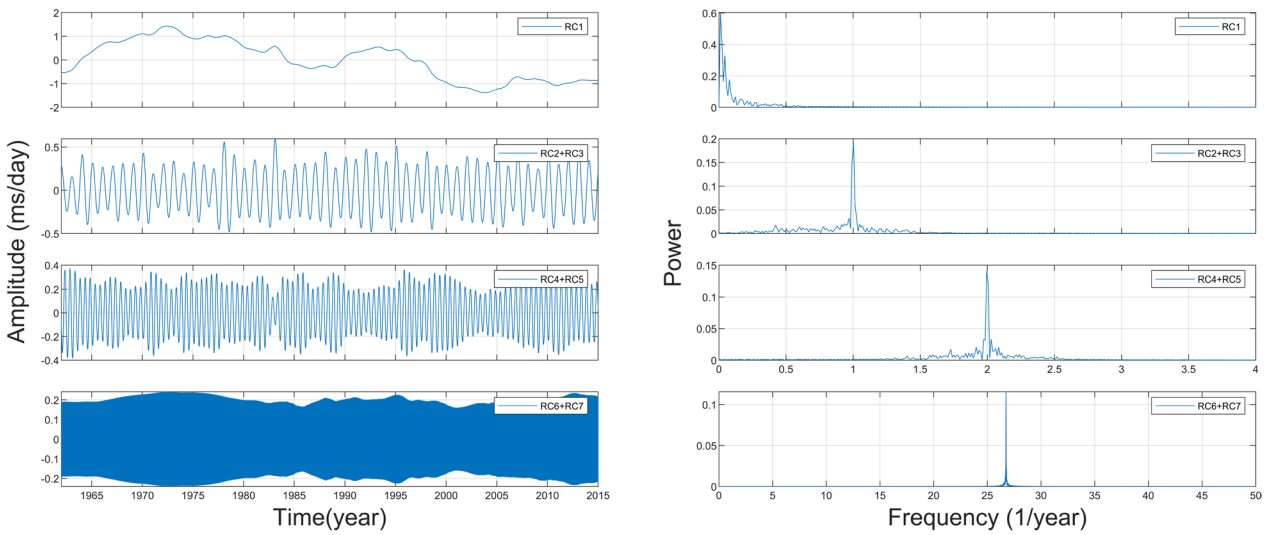


Fig. 6 RCs of LODR time series and its power spectrum analysis

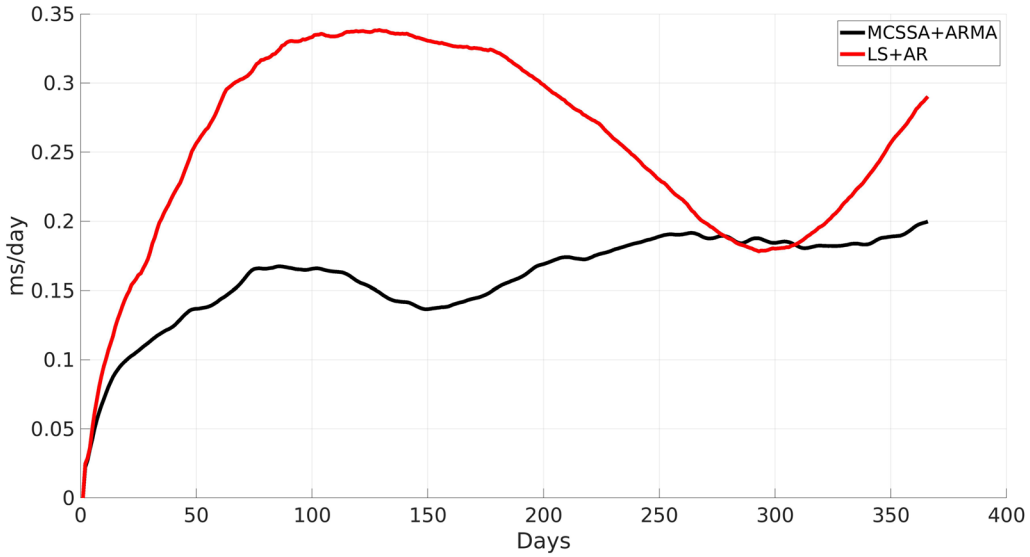


Fig. 7 MAE of MCSSA+ARMA and LS+AR

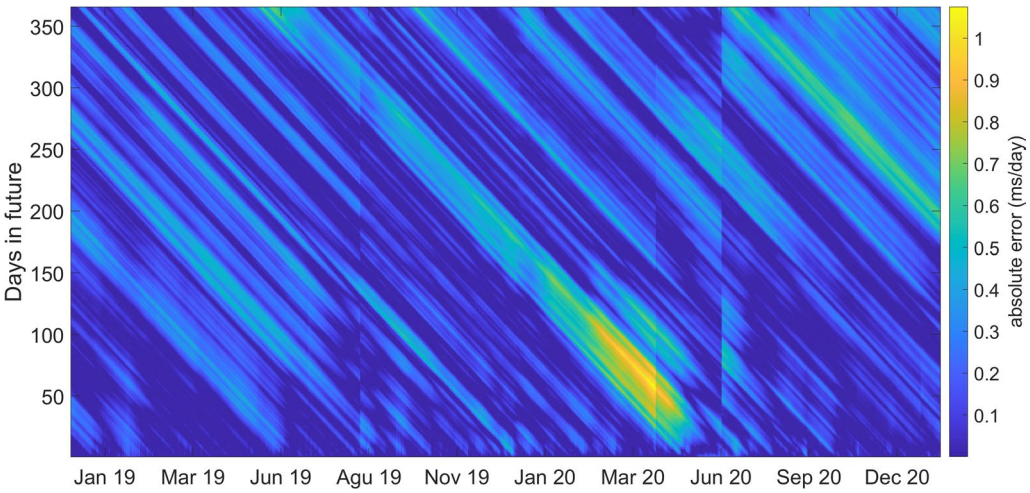


Fig. 8 Daily error of the LOD prediction using MCSSA+ARMA approach between Jan 19 and Dec 2020

performance of these components. Additionally, an ARMA model was used to predict the behavior of the residuals. After combining the SSA and ARMA prediction models, tidal effects subtracted from the LOD time series were added for the predicted days. In other words, all days of 2020 were predicted during this step. In subsequent intervals, with a 1-day shift on the first day of forecasting, the number of forecasts reached 751. The most recent interval was considered to be between January 21, 2021, and January 19, 2022.

In addition, the efficacy of the proposed method was compared with that of an LS+AR hybrid method. Figure 7 depicts the MAE of LS+AR (red curve) and our

Table 1 MAE results for MCSSA+ARMA and LS+AR

Lead day	This study (ms/day)	LS with AR (ms/day)	Percentage
1	0.022	0.025	12
5	0.050	0.061	18
10	0.076	0.101	24.7
30	0.115	0.184	37.5
90	0.167	0.330	49.3
180	0.153	0.322	52.4
365	0.201	0.294	31.6

proposed method (black curve). The results demonstrate that MCSSA+ARMA significantly improves LOD prediction for short-, medium-, and long-term intervals. According to Table 1, the MAE values for the 10th, 90th, 180th, 270th, and 365th days using the MCSSA+ARMA hybrid method are 0.076, 0.167, 0.153, 0.188, and 0.201 ms/day, respectively. In addition, the MAE of LS+AR methods for the same days is 0.101, 0.330, 0.322, 0.198, and 0.294 ms/day. Table 1 reveals that while our proposed method improved prediction accuracy (MAE) by as much as 60 percent, particularly for medium-term forecasts, some days (such as the 300th day) show no improvement for LOD forecasts. As indicated by (Malkin et al. 2022), Fig. 8 depicts various features, and certain patterns exhibit high errors from the beginning of March 2020 to June 2020, which may be caused by the El Nino effect or certain prospective geomagnetic jerk events.

3.4 Short-term prediction

Participants at the second EOP PCC primarily examined short-term prediction horizons due to their practical application for evaluating and benchmarking LOD prediction models. Using insights from the second EOP PCC, this subsection compares the performance of LOD prediction models based on short-term data. In this study, we used IERS Bulletin A (final.daily) and C04 data, following Dill et al. (2019), to enhance signal relevance for short-term predictions. Our methodology involved merging C04 and final.daily time series data and applying predictions from September 1, 2021, to December 31, 2022. This mirrors the conditions of the second EOP PCC.

In the first step, we used Monte Carlo test to determine the optimal number of eigenvalues for SSA reconstructions. Using 4 years of data, as described by Dill et al. (2019), the number of eigenvalues is 26, as shown in Fig. 9. As a result, SSA-extrapolated residuals for LOD and ARMA-predicted stochastic behaviors were incorporated into extended IERS tide models for the following 10 days.

In this study, we evaluate the efficacy of the proposed MCSSA+ARMA model in predicting the Length of Day (LOD) by comparing its Mean Absolute Error (MAE) against established methods from the fourth cluster of

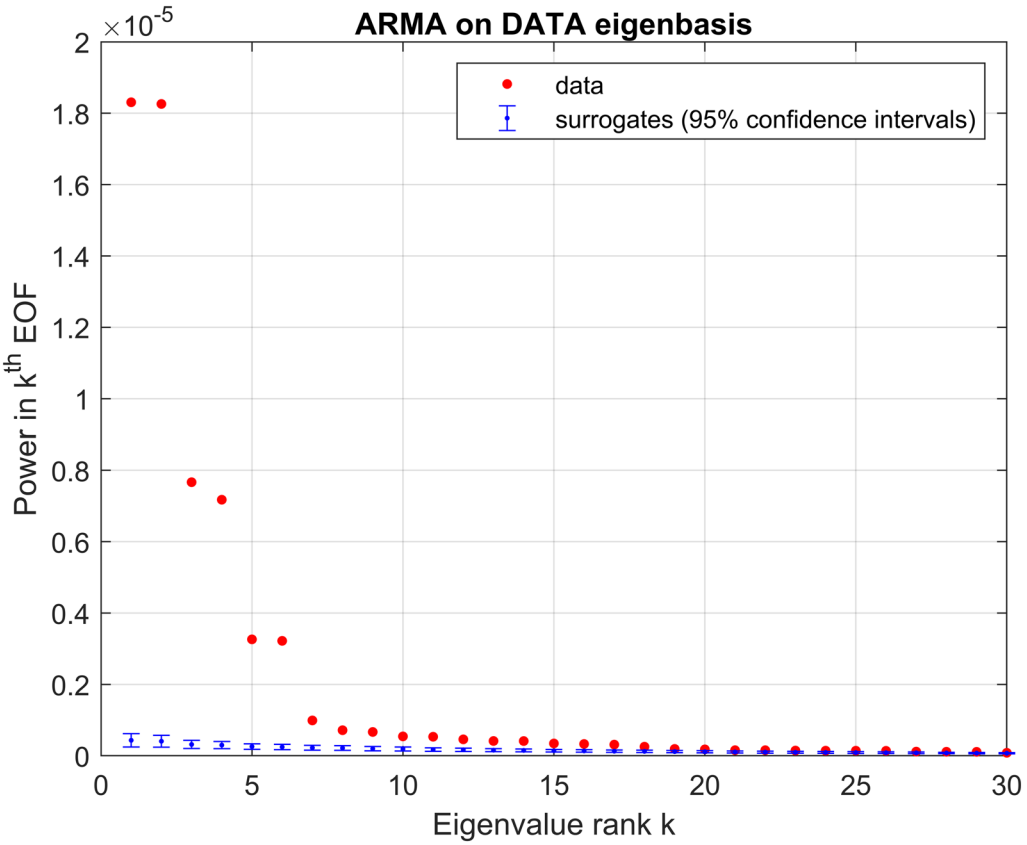


Fig. 9 MCSSA algorithm results for the first 30 reconstructed components by SSA. The data eigenvalues (red dots) and their 95% corresponding confidence intervals (blue error bars) are plotted in their rank order. Confidence intervals are constructed from an ensemble of 1000 surrogate data generated as realizations of the estimated ARMA model

the EOP PCC benchmark. The MCSSA+ARMA model demonstrates competitive performance, with an MAE of approximately 0.01 ms/day on day 1, and up to day 6 the MAE remains below 0.01 ms/day, surpassing the majority of competing models within the cluster. These models include SSA+Copula (ID = 117), Normal Time-Frequency Transform (NTFT) (ID = 121), Weighted Least Squares (WLS) + ARIMA (ID = 122), LS + Kriging (ID = 141), and LS + ARIMA (ID = 156). While the Kalman filter (ID = 104) and Dynamic Mode Decomposition (DMD) (ID = 157) achieve a slightly lower MAE, the MCSSA+ARMA model represents a viable alternative, offering comparable accuracy and potentially enhanced interpretability due to its ARMA component (Fig. 10). Among these models, ID=104 and ID=117 have been integrated as input data to bolster prediction accuracy and stability.

It is pertinent to acknowledge that the MCSSA+ARMA model does possess certain limitations compared to the Kalman filter. While the MCSSA+ARMA model demonstrates comparable accuracy, the Kalman filter enjoys the advantage of being a recursive algorithm capable of real-time prediction updates, rendering it better suited for dynamic systems and leveraging additional input datasets. The MCSSA+ARMA model, notwithstanding its limitations, stands out for its competitive performance and accuracy relative to the Kalman filter. Furthermore, the incorporation of the ARMA component in the MCSSA+ARMA model offers the potential for improved interpretability. Nonetheless, it remains imperative to recognize that the Kalman filter's recursive nature, facilitating real-time updates, grants it a distinct advantage, particularly in dynamic system contexts.

To isolate the potential influence of EAM on prediction accuracy, particularly for models like SSA+Copula (ID = 117) and the Kalman filter (ID = 104), we removed an offset and centered the mean absolute error (MAE) at zero. The MCSSA+ARMA model remained competitive, achieving an MAE of 0.1 ms/day up to day 8 and 0.12 ms/day for day 10, outperforming most other methods. However, the Kalman filter (ID = 104) still exhibited a slight advantage with a marginally lower centered MAE (Fig. 11). Interestingly, after offset removal, SSA+Copula (ID = 117) showed a greater improvement in centered MAE compared to the MCSSA+ARMA model, suggesting it may benefit more from the inclusion of EAM.

4 Discussion

The presented methodology, which integrates Monte Carlo Singular Spectrum Analysis (MCSSA) with an autoregressive moving average (ARMA) model, demonstrates significant improvements in the prediction accuracy of the Length of Day (LOD) for both short-term and long-term intervals. By extracting deterministic components using SSA and modeling the residuals with ARMA, the proposed method effectively addresses the complex

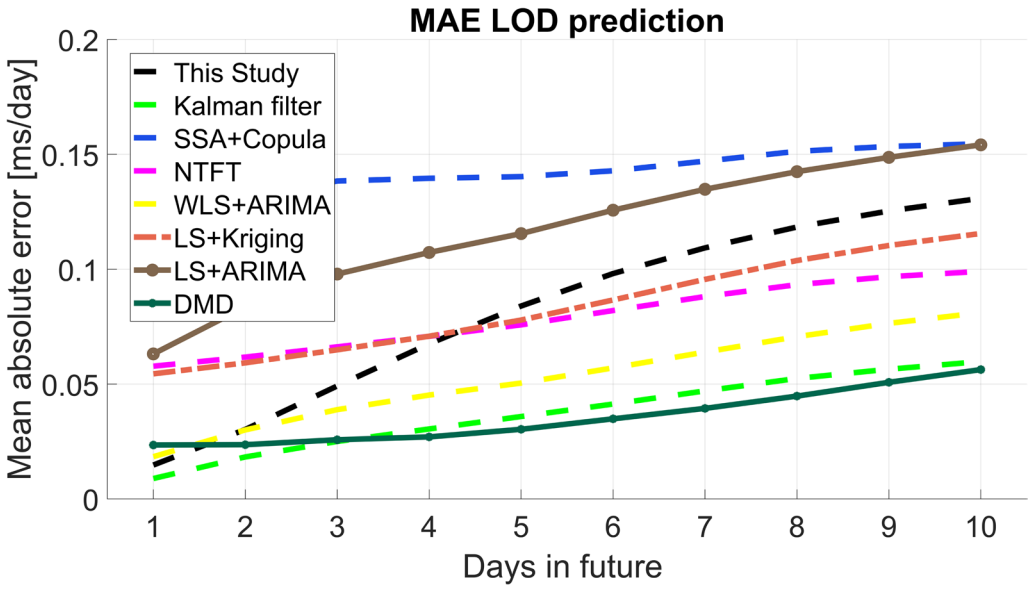


Fig. 10 MAE comparison of the MCSSA+ARMA model proposed in this study with black line against other prediction models, including Kalman filter (ID=104 with light green line, with 69 submissions), SSA+Copula (ID=117 with blue line, with 57 submissions), Normal Time-Frequency Transform (NTFT) (ID=121 with pink line, with 67 submissions), Weighted Least Squares (WLS) + ARIMA (ID=122 with yellow line, with 61 submissions), LS + Kriging (ID=141 with crimson line, with 38 submissions), LS + ARIMA (ID=156 with brown line, with 31 submissions), and Dynamic Mode Decomposition (DMD) (ID=157 with green line, with 7 submissions)

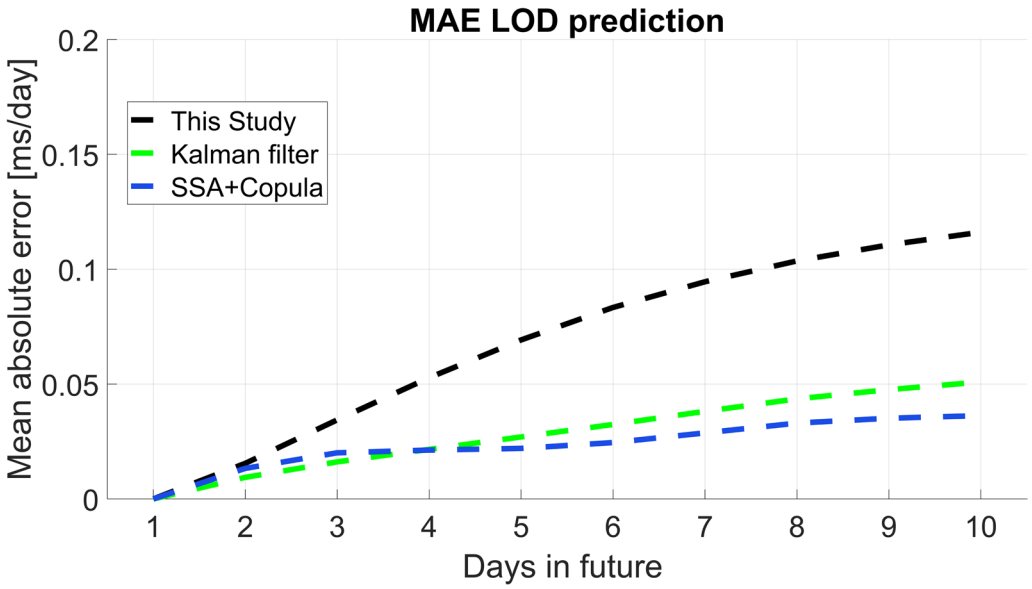


Fig. 11 MAE comparison of the MCSSA+ARMA model proposed in this study against other prediction models, including the Kalman filter (ID=104 with light green line, with 69 submissions) and SSA+Copula (ID=117 with blue line, with 57 submissions), in which the offset is removed and shifted to 0

nature of LOD variations driven by geophysical and astronomical factors.

For long-term predictions, the results indicate that the MCSSA+ARMA approach significantly outperforms traditional methods, such as the least squares and autoregressive (LS+AR) hybrid model. The inclusion of non-linear trends and statistically significant frequencies, such as annual, semi-annual, and 13-day periods, enhances the predictive capability of the SSA component. The iterative approach to convergence, coupled with the ARMA model for residuals, results in reduced prediction errors over extended periods. This is evident in the mean absolute error (MAE) values, which show substantial improvements at various forecast intervals compared to the LS+AR model. Specifically, the MCSSA+ARMA method achieves up to a 60% reduction in MAE for medium-term forecasts, demonstrating its robustness and effectiveness.

The short-term prediction results further validate the efficacy of the MCSSA+ARMA method. Compared to other models from the second EOP PCC, such as the Kalman filter, SSA+Copula, and various combinations of least squares and ARIMA methods, the MCSSA+ARMA approach exhibits competitive performance with an MAE around 0.12 ms/day. This performance is on par with the Kalman filter, which is renowned for its real-time updating capability. However, it is important to note that the typical 2-day delay in LOD rapid products was not considered in this study, which could influence the accuracy assessment when comparing to operational predictions. This limitation should be addressed in future work to ensure a more realistic comparison with operational frameworks. The MCSSA+ARMA method's ability to achieve similar accuracy without relying on recursive updates highlights its potential for practical applications in short-term LOD forecasting.

The improved accuracy of LOD predictions has significant implications for various fields reliant on precise Earth orientation parameters, such as astronomy, geodesy, and space navigation. The enhanced predictability of LOD variations can lead to better alignment of terrestrial and celestial reference frames, improving the accuracy of satellite positioning and navigation systems. Additionally, understanding the deterministic components of LOD variations can provide insights into underlying geophysical processes, aiding in the study of Earth's internal dynamics and external gravitational influences.

While the MCSSA+ARMA method shows promising results, it is essential to acknowledge its limitations. The method relies on historical LOD data and the assumption that past patterns will continue into the future, which may not always hold true in the presence of unexpected geophysical events or significant changes in external gravitational forces. Moreover, the computational complexity of SSA and the need for extensive data processing may limit its real-time applicability compared to more dynamic approaches like the Kalman filter.

Future work could explore hybrid models that integrate the real-time updating capabilities of the Kalman filter

with the robustness of the MCSSA+ARMA approach. Additionally, incorporating more comprehensive datasets, such as effective angular momentum (EAM) data, could further enhance prediction accuracy and stability. Continued participation in initiatives like the EOP PCC will provide valuable benchmarks and opportunities to refine and validate prediction models.

5 Conclusion

This study focuses on evaluating the effectiveness of prediction methods for the Length of Day (LOD) and demonstrates that the proposed MCSSA+ARMA algorithm can significantly improve predictive accuracy for both short-term and long-term intervals. Although the importance of input data in improving prediction accuracy is acknowledged, this study emphasizes the performance of the prediction framework independently of input variations. EAM data, which include contributions from atmospheric, oceanic, and hydrological processes, have been shown in previous studies to enhance LOD predictions. Future research will aim to integrate EAM data and other EOP time series into the framework to develop a more comprehensive and reliable prediction approach. While this study has some limitations, its findings provide a foundation for further research to refine prediction models and explore hybrid approaches that combine the strengths of different methods. The ongoing development and validation of these models, along with the integration of improved input datasets such as EAM data, will be essential for advancing the understanding and prediction of Earth's orientation parameters.

Abbreviations

EOP	Earth orientation parameters
LOD	Length of day
LODR	The time series obtained after removing these tides is denoted as LODR
UT1-UTC	Universal time
dx,dy	Precession–nutation corrections
EAM	Effective angular momentum
IERS	International Earth Rotation and Reference Systems Service
LS	Least squares
AR	Autoregressive models
AC	Auto-covariance models
ARMA	Auto-regressive moving average models
RC	Reconstructed component
KF	Kalman filtering
NNs	Neural networks
SSA	Singular Spectrum Analysis
EOPPCC	EOP Prediction Comparison Campaigns
ARIMA	Autoregressive integrated moving average models

Acknowledgements

We are grateful to the International Earth Rotation and Reference System Service IERS for providing the length-of-day data.

Author contributions

S.S. conducted most of the data analysis, the writing of the manuscript. S.M and S.B. wrote part of the manuscript and conceived and designed the study. M.A.S, A.A.S and S.M.K provided the supervision and helped to improve the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

SB was partially supported by Generalitat Valenciana (SEJIGENT/2021/001), the European Union - NextGenerationEU (ZAMBRANO 21-04) and Spanish Project PID2020-119383GB-I00 funded by Ministerio de Ciencia e Innovación (MCIN/AEI/10.13039/501100011033/).

Data availability

IERS rapid and final EOPs series can be found on https://www.iers.org/IERS/EN/DataProducts/EarthOrientationData/eop.html The data collected during the Second Earth Orientation Parameters Prediction Comparison Campaign used in this study can be accessed from the GFZ Data Services under the following link: https://doi.org/10.5880/GFZ.1.3.2023.001.

Declarations

Competing interests

The authors declare no competing interests.

Author details

[1]School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. [2]Department Geodesy, Federal Agency for Cartography and Geodesy (BKG), Frankfurt am main, Germany. [3]UAVAC, Department of Applied Mathematics, Universidad de Alicante, Alicante, Spain. [4]Department of Surveying Engineering, Faculty of Civil Engineering, Jundi-Shapur University of Technology, Dezful, Iran. [5]Department of Control and Operations, Faculty of Aerospace Engineering, Delft University of Technology, Delft, Netherlands.

Received: 8 July 2024 Accepted: 28 February 2025
Published online: 17 March 2025

References

Agnew DC (2024) A global timekeeping problem postponed by global warming. Nature 628(8007):333–336
Akyilmaz O, Kutterer H (2004) Prediction of earth rotation parameters by fuzzy inference systems. J Geodesy 78:82–93
Allen MR, Smith LA (1996) Monte Carlo SSA: detecting irregular oscillations in the presence of colored noise. J Clim 9(12):3373–3404
Altamimi Z, Rebischung P, Collilieux X, Métivier L, Chanard K (2023) Itrf 2020: an augmented reference frame refining the modeling of nonlinear station motions. J Geodesy 97(5):47
Angermann D, Seitz M, Drewes H (2010) Analysis of the Doris contributions to itrf2008. Adv Space Res 46(12):1633–1647
Aoki S, Kinoshita H, Guinot B, Kaplan G, McCarthy DD, Seidelmann PK (1982) The new definition of universal time. Astron Astrophys 105:359–361
Barnes R, Hide R, White A, Wilson C (1983) Atmospheric angular momentum fluctuations, length-of-day changes and polar motion. Proc R Soc Lond A Math Phys Sci 387(1792):31–73
Belda S, Ferrándiz J M, Heinkelmann R, Schuh H (2018) A new method to improve the prediction of the celestial pole offsets. Sci Rep 8
Broomhead DS, King GP (1986) Extracting qualitative dynamics from experimental data. Physica D 20(2–3):217–236
Chao BF (1989) Length-of-day variations caused by el nino-southern oscillation and quasi-biennial oscillation. Science 243(4893):923–925
Chen J, Wilson C (2005) Hydrological excitations of polar motion, 1993–2002. Geophys J Int 160(3):833–839
Coulot D, Pollet A, Collilieux X, Berio P (2010) Global optimization of core station networks for space geodesy: application to the referencing of the slr eop with respect to itrf. J Geodesy 84(1):31–50
Dick W R, Richter B (2004) The international earth rotation and reference systems service (IERS). In Organizations and strategies in astronomy. Springer, pp 159–168

Dill R, Dobslaw H, Thomas M (2019) Improved 90-day earth orientation predictions from angular momentum forecasts of atmosphere, ocean, and terrestrial hydrosphere. J Geodesy 93(3):287–295
Freedman A, Steppe J, Dickey J, Eubanks T, Sung L-Y (1994) The short-term prediction of universal time and length of day using atmospheric angular momentum. J Geophys Res Solid Earth 99(B4):6981–6996
Ghil M, Taricco C (1997) Advanced spectral-analysis methods. In Past and present variability of the solar-terrestrial system: measurement, data analysis and theoretical models. Ios Press, pp 137–159
Gou J, Kiani Shahvandi M, Hohensinn R, Soja B (2023) Ultra-short-term prediction of lod using lstm neural networks. J Geodesy 97(5):52
Gross RS (2007) Earth rotation variations-long period. Treat Geophys 3:239–294
Gross RS, Marcus SL, Eubanks TM, Dickey JO, Keppenne CL (1996) Detection of an enso signal in seasonal length-of-day variations. Geophys Res Lett 23(23):3373–3376
Guessoum S, Belda S, Ferrandiz JM, Modiri S, Raut S, Dhar S, Heinkelmann R, Schuh H (2022) The short-term prediction of length of day using 1d convolutional neural networks (1d cnn). Sensors 22(23):9517
Guo J, Li Y, Dai C, Shum C (2013) A technique to improve the accuracy of earth orientation prediction algorithms based on least squares extrapolation. J Geodyn 70:36–48
Hellmers H, Modiri S, Bachmann S, Thaller D, Bloßfeld M, Seitz M, Gipson J (2022) Combined ivs contribution to the itrf2020. In Geodesy for a Sustainable Earth: Proceedings of the 2021 Scientific Assembly of the International Association of Geodesy, Beijing, China, June 28–July 2, 2021. Springer, pp 3–13
Holton J R, Dmowska R (1989) El Niño, La Niña, and the southern oscillation. Academic press
Jin X, Liu X, Guo J, Shen Y (2021) Analysis and prediction of polar motion using MSSA method. Earth Planets and Space 73(1):1–13. https://doi.org/10.1186/s40623-021-01477-2
Kalarus M, Schuh H, Kosek W, Akyilmaz O, Bizouard C, Gambis D, Gross R, Jovanović B, Kumakshev S, Kutterer H et al (2010) Achievements of the earth orientation parameters prediction comparison campaign. J Geodesy 84:587–596
Khazraei S, Amiri-Simkooei A (2019) On the application of Monte Carlo singular spectrum analysis to GPS position time series. J Geodesy 93(9):1401–1418
Kiani Shahvandi M, Dill R, Dobslaw H, Kehm A, Bloßfeld M, Schartner M, Mishra S, Soja B (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
Kondrashov D, Berloff P (2015) Stochastic modeling of decadal variability in ocean gyres. Geophys Res Lett 42(5):1543–1553
Kosek W (2002) Autocovariance prediction of complex-valued polar motion time series. Adv Space Res 30(2):375–380
Kosek W, McCarthy D, Johnson T, 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. Systemes de Reference Spatio-Temporels, pp 164–169
Kosek W I, Kalarus M, Niedzielski T, Capitaine N (2007) Forecasting of the Earth orientation parameters: comparison of different algorithms. Citeseer
Malkin Z, Belda S, Modiri S (2022) Detection of a new large free core nutation phase jump. Sensors 22(16):5960
Mathews P, Buffett BA, Herring TA, Shapiro II (1991) Forced nutations of the earth: influence of inner core dynamics: 1. theory. J Geophys Res Solid Earth 96(B5):8219–8242
Modiri S, Belda S, Heinkelmann R, Hoseini M, Ferrándiz JM, Schuh H (2018) Polar motion prediction using the combination of SSA and copula-based analysis. Earth Planets Space 70(1):1–18. https://doi.org/10.1186/s40623-018-0888-3
Modiri S, Belda S, Hoseini M, Heinkelmann R, Ferrándiz JM, Schuh H (2020) A new hybrid method to improve the ultra-short-term prediction of lod. J Geodesy 94:1–14
Modiri S, Heinkelmann R, Belda S, Malkin Z, Hoseini M, Korte M, Ferrándiz JM, Schuh H (2021) Towards understanding the interconnection between celestial pole motion and earth's magnetic field using space geodetic techniques. Sensors 21(22):7555
Petit G, Luzum B et al (2010) Iers conventions. IERS Tech Note 36(1):2010
Raut S, Modiri S, Heinkelmann R, Balidakis K, Belda S, Kitpracha C, Schuh H (2022) Investigating the relationship between length of day and el-nino using wavelet coherence method. In Geodesy for a sustainable earth: Proceedings of the 2021 Scientific Assembly of the International Association of Geodesy, Beijing, China, June 28–July 2, 2021. Springer, pp 253–258
Rial JA, Oh J, Reischmann E (2013) Synchronization of the climate system to eccentricity forcing and the 100,000-year problem. Nat Geosci 6(4):289–293
Schuh H, Behrend D (2012) Vlbi: a fascinating technique for geodesy and astrometry. J Geodyn 61:68–80
Schuh H, Ulrich M, Egger D, Müller J, Schwegmann W (2002) Prediction of earth orientation parameters by artificial neural networks. J Geodesy 76:247–258
Seitz F, Schuh H (2010) Earth rotation. In Sciences of geodesy-I. Springer, pp 185–227
Shahvandi MK, Adhikari S, Dumberry M, Mishra S, Soja B (2024) The increasingly dominant role of climate change on length of day variations. Proc Natl Acad Sci 121(30):e2406930121
Shen Y, Guo J, Liu X, Kong Q, Guo L, Li W (2018) Long-term prediction of polar motion using a combined SSA and ARMA model. J Geodesy 92:333–343
Śliwińska-Bronowicz J, Kur T, Wińska M, Dobslaw H, Nastula J, Partyka A, Belda S, Bizouard C, Boggs D, Bruni S, Chen L, Chin M, Dhar S, Dill R, Ferrandiz JM, Gou J, Gross R, Guessoum S, Han S, Heinkelmann R, Irrgang C, Kiani Shahvandi M, Li J, Ligas M, Lie L, Lu W, Mayer V, Michalczak M, Modiri S, Otten M, Ratcliff T, Raut S, Saynisch-Wanger J, Schartner M, Shoenemann E, Schuh H, Soja B, Su X, Thaller D, Thomas M, Wang G, We Y, Xu X, Yang X, Zhao X, Zhou Z (2024) Assessment of length-of-day and universal time predictions based on the results of the second earth orientation parameters prediction comparison campaign. J Geodesy 98(3):1–34
Vautard R, Ghil M (1989) Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. Physica D 35(3):395–424
Wu F, Chang G, Deng K (2021) One-step method for predicting lod parameters based on ls+ ar model. J Spat Sci 66(2):317–328
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
Yi S, Sneeuw N (2021) Filling the data gaps within grace missions using singular spectrum analysis. J Geophys Res Solid Earth 126(5):e2020JB021227

Publisher's Note

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