Improving Medium- and Long-Term Prediction Accuracy of Polar Motion Using Singular Spectrum Analysis: Postprint

Authors:

Date: 2024-06-03T00:00:00+00:00

Abstract

Due to the time-consuming transmission of space geodetic observation data and the complexity of processing procedures, the acquisition of polar motion measurements suffers from time delays, which cannot satisfy application domains with significant demands for high-precision polar motion predictions. To address the complex time-varying characteristics of polar motion, a prediction method based on singular spectrum analysis (SSA) is proposed. First, SSA is employed to separate and extract the high-frequency and low-frequency components from the polar motion time series; second, a combined prediction model based on least squares (LS) extrapolation and autoregressive (AR) models is established for the high-frequency and low-frequency components of polar motion. The results demonstrate that the SSA method can accurately separate and extract the low-frequency and high-frequency components of polar motion, and that the combined prediction of these components can significantly improve the medium- and long-term (30–365;d) prediction accuracy of polar motion. Compared with the polar motion predictions provided in Bulletin A by the International Earth Rotation and Reference Systems Service (IERS), the SSA method achieves maximum improvements in medium- and long-term prediction accuracy of 45.97% and 62.44% for the PMX component (prime meridian direction) and PMY component (west 90° meridian direction) of polar motion, respectively. The research results validate the effectiveness of the SSA method in improving medium- and long-term polar motion predictions.

Full Text

Preamble

Acta Astronomica Sinica, Vol. 65, No. 3, May 2024
doi: 10.15940/j.cnki.0001-5245.2024.03.002

ChinaRxiv 中X

Improving Medium- and Long-term Prediction Accuracy of Polar Motion Using Singular Spectrum Analysis

ZHAO Dan-ning[1], LEI Yu[2]†, QIAO Hai-hua[3]‡
([1] School of Electrical and Electronic Engineering, Baoji University of Arts and Sciences, Baoji 721016)
([2] School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121)
([3] National Time Service Center, Chinese Academy of Sciences, Xi'an 710600)

Abstract

Due to time-consuming data transmission and complex processing procedures in space geodetic observations, polar motion measurements are subject to delays, which cannot satisfy applications requiring high-precision polar motion predictions. Addressing the complex time-varying characteristics of polar motion, this paper proposes a prediction method based on singular spectrum analysis (SSA). First, SSA is employed to separate and extract the high-frequency and low-frequency components from polar motion time series. Second, a combination of least squares (LS) extrapolation and autoregressive (AR) models is established to predict these components respectively. The results demonstrate that SSA can accurately separate and extract the low-frequency and high-frequency components of polar motion. The combined prediction of these components significantly improves medium- and long-term (30–365 days) prediction accuracy. Compared with the polar motion predictions provided in Bulletin A by the International Earth Rotation and Reference Systems Service (IERS), the SSA method achieves maximum improvements of 45.97% and 62.44% for the PMX component (prime meridian direction) and PMY component (west 90° meridian direction), respectively. These findings validate the effectiveness of SSA in improving medium- and long-term polar motion predictions.

Keywords: astrometry: reference systems; Earth rotation: polar motion; methods: data analysis

1 Introduction

Earth Orientation Parameters (EOP) are essential for transforming between the International Terrestrial Reference System and the International Celestial Reference System, comprising polar motion, Universal Time (UT1), precession, and nutation. Space geodetic techniques such as the Global Positioning System (GPS), Very Long Baseline Interferometry (VLBI), and Doppler Orbitography and Radio-positioning Integrated by Satellite (DORIS) can measure the complete set of EOP parameters, establishing a direct connection between the International Terrestrial Reference Frame and International Celestial Reference Frame. However, due to data transmission delays and complex data processing procedures, EOP data are typically available with a latency of 1–7 days. Short-term EOP predictions are crucial for applications such as deep-space probe tracking and precise satellite orbit determination, while medium- and long-term predictions are important for satellite autonomous navigation and astronomical geodynamics research. Consequently, accurate short-, medium-, and long-term EOP predictions are of paramount importance.

Polar motion, defined as the movement of Earth's instantaneous rotation axis within the Earth's body, consists of two components in the terrestrial pole coordinate system: PMX (prime meridian direction) and PMY (west 90° meridian direction). Modern space geodetic techniques can determine polar motion with an accuracy of approximately 0.1 mas. Influenced by multiple excitation factors including the atmosphere, oceans, and terrestrial hydrosphere, polar motion exhibits irregular variations across different time scales, causing its prediction accuracy to degrade rapidly with increasing forecast length. The prediction accuracy is approximately 4 mas for a 10-day forecast and about 20 mas for a 90-day forecast. These complex time-varying characteristics result in poor predictability, particularly for medium- and long-term forecasts.

To improve polar motion prediction, researchers have proposed various models, including least squares (LS) extrapolation, time series analysis models, neural networks, fuzzy inference systems, autocovariance models, and multi-parameter polar motion models combined with effective angular momentum functions. LS extrapolation, autoregressive (AR) models, and autocovariance models are straightforward and easy to implement. Neural networks and fuzzy inference systems, as machine learning methods, involve more complex modeling and their prediction performance is highly sensitive to model parameters. Multi-parameter models combined with effective angular momentum demonstrate good prediction performance but are typically limited to short-term predictions due to constraints from forecasting surface fluid angular momentum data.

The first Earth Orientation Parameters Prediction Comparison Campaign (EOP PCC) evaluated various EOP prediction models, revealing that no single model achieves optimal prediction accuracy (mean absolute error) for all EOP parameters and prediction spans. Among these methods, the combination of LS extrapolation and AR models yields relatively high prediction accuracy for polar motion and has become a widely applied approach. Currently, the IERS Rapid Service/Prediction Center uses LS extrapolation combined with AR models to generate polar motion predictions in Bulletin A.

Atmospheric, oceanic, and terrestrial hydrospheric excitations cause polar motion to exhibit not only Chandler wobble and annual wobble but also various irregular short-period variations. Chandler and annual wobbles primarily affect medium- and long-term predictions, while short-period variations mainly influence short-term predictions. Therefore, medium- and long-term polar motion forecasting must prioritize Chandler and annual wobbles. To enhance medium- and long-term prediction accuracy, some scholars have employed time-frequency analysis techniques such as wavelet analysis and empirical mode decomposition to decompose polar motion time series, extracting trend, periodic, and irregular components before applying LS extrapolation and AR models for separate prediction. These studies demonstrate that time series decomposition methods can improve polar motion prediction.

Singular spectrum analysis (SSA) is a novel nonlinear and non-stationary signal processing technique capable of extracting useful information from small-sample observations for periodic identification, trend analysis, denoising, and signal reconstruction. SSA is an expansion form of empirical orthogonal functions with the advantage of requiring no parametric model assumptions and being unrestricted by signal stationarity conditions. It possesses the capability to identify and enhance periodic oscillation signals, making it suitable for analyzing signals composed of superimposed periodic oscillations. This paper addresses the complex time-varying characteristics of polar motion by introducing SSA to separate and extract high-frequency and low-frequency components, then applying LS extrapolation and AR models for combined prediction. The results demonstrate that incorporating SSA significantly reduces the mean absolute error of medium- and long-term (30–365 days) polar motion predictions.

2.1 Principle of Singular Spectrum Analysis

Let $x_1, x_2, \cdots, x_N$ be a one-dimensional observed time series of length N. SSA can decompose the series into data subsequences with different characteristics such as periodicity, trend, and quasi-periodicity, enabling identification, denoising, and reconstruction of the original signal. The main computational steps of SSA are as follows:

(1) **Construction of the trajectory matrix.** If the window length is an integer M $(1 < M < N/2)$, the trajectory matrix $\mathbf{X}$ can be expressed as:

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{N-M+1} \\ x_2 & x_3 & \cdots & x_{N-M+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_M & x_{M+1} & \cdots & x_N \end{bmatrix}$$

Each row and column of matrix $\mathbf{X}$ represents a subsequence of the original time series, where $\mathbf{X}_i = [x_i, x_{i+1}, \cdots, x_{i+M-1}]^T$ for $1 \le i \le N-M+1$.

(2) **Singular value decomposition.** Let $\mathbf{S} = \mathbf{X}\mathbf{X}^T$. Compute the eigenvalues of $\mathbf{S}$ and sort them in descending order $(\lambda_1 > \lambda_2 > \cdots > \lambda_M > 0)$, with corresponding eigenvectors $\mathbf{U}_1, \mathbf{U}_2, \cdots, \mathbf{U}_M$. If $d = \text{rank}(\mathbf{X})$ and $\mathbf{V}_k = \mathbf{X}^T\mathbf{U}_k/\sqrt{\lambda_k}$ for $k = 1, 2, \cdots, d$, then the trajectory matrix $\mathbf{X}$ can be expressed as:

$$\mathbf{X} = \sum_{k=1}^{d} \sqrt{\lambda_k}\mathbf{U}_k\mathbf{V}_k^T$$

where $\sqrt{\lambda_k}$ are the singular values of the matrix.

(3) **Diagonal averaging.** Each elementary matrix is transformed back into a new time series of length N, called a reconstructed component (RC). The sum of all RCs equals the original time series. Let $\mathbf{Z} = \mathbf{X}_k$ and $z_{i,j}$ be the RC obtained through diagonal averaging, where $i = 1, 2, \cdots, M$ and $j = 1, 2, \cdots, N$. Let $M^* = \min(M, N-M+1)$ and $K^* = \max(M, N-M+1)$. If $M > K^*$, set $z^*_{i,j} = z_{i,j}$; otherwise, set $z^*_{i,j} = z_{j,i}$. The diagonal averaging formula can be expressed as:

$$\tilde{x}_i = \begin{cases} \frac{1}{i}\sum_{m=1}^{i} z^*_{m,i-m+1}, & 1 \le i \le M^* \\ \frac{1}{M^*}\sum_{m=1}^{M^*} z^*_{m,i-m+1}, & M^* < i \le K^* \\ \frac{1}{N-i+1}\sum_{m=i-K^*+1}^{N-K^*+1} z^*_{m,i-m+1}, & K^* < i \le N \end{cases}$$

Polar motion series contain long-term trend variations, Chandler wobble, annual wobble signals, and irregular short-period variations. The w-correlation analysis is used to examine correlations between RCs, grouping those with similar signal characteristics. If reconstructed components are denoted as $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$, the w-correlation between them can be expressed as:

$$\rho_{i,j} = \frac{(\mathbf{Y}^{(i)}, \mathbf{Y}^{(j)})}{\|\mathbf{Y}^{(i)}\| \cdot \|\mathbf{Y}^{(j)}\|}, \quad 1 \le i, j \le M$$

where $(\mathbf{Y}^{(i)}, \mathbf{Y}^{(j)}) = \sum_{k=1}^{N} w_k y_k^{(i)} y_k^{(j)}$, $\|\mathbf{Y}^{(i)}\| = \sqrt{(\mathbf{Y}^{(i)}, \mathbf{Y}^{(i)})}$, and w_k are weight coefficients defined as $w_k = \min(k, M, N-k)$, with k being the index of reconstructed component $\mathbf{Y}^{(i)}$. The absolute value of $\rho_{i,j}$ approaching 1 indicates stronger correlation between $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$, suggesting they should be treated as the same signal component.

2.2 Analysis Results

The EOP 14 C04 series provided by IERS is derived from joint solutions of multiple space geodetic observations including DORIS, GPS, and VLBI, offering extremely high measurement precision. This study analyzes polar motion data from this series with a sampling interval of 1 day. When applying SSA to time series analysis, the critical parameter is the window length M. Generally, when $1 < M < N/3$, the method can effectively identify periodic signals with periods ranging from $M/5$ to M. For polar motion time series with known periodic oscillation frequencies, the Chandler and annual wobbles are prominent and represent the primary factors affecting prediction accuracy. Therefore, with the goal of separating and extracting Chandler and annual wobble signals, this study selects a window length of $M = 730$.

Due to space limitations, this paper presents only the analysis of PMX and PMY component data from January 1, 2003, to December 31, 2014. The SSA decomposition yields the first 100 RCs, whose eigenvalue contribution rates (logarithm base 10) are plotted in Figure 1 [Figure 1: see original paper]. The eigenvalue contribution rate is calculated as:

$$\text{Contribution Rate} = \frac{\lambda_i}{\sum_{j=1}^{M} \lambda_j} \times 100\%$$

Figure 1 shows that RCs with larger eigenvalue contributions are concentrated in the first few orders. To more clearly illustrate the eigenvalue contribution rates for PMX and PMY components, Table 1 lists the contribution rates of the first 8 RCs. The first 8 RCs for PMX and PMY components are shown in Figures 2 [Figure 2: see original paper] and 3 [Figure 3: see original paper], respectively, and their w-correlation analysis results are presented in Figure 4 [Figure 4: see original paper].

Table 1 and Figures 2–3 reveal that the first-order RC has the largest eigenvalue contribution rate for both components: 25.43% for PMX and 62.28% for PMY, corresponding to the polar motion trend signal. The second- and third-order RCs have the next largest contribution rates, which are relatively close to each other: 22.79% and 21.02% for PMX, and 12.09% and 11.29% for PMY. Figure 4 shows that the correlation coefficients between the second- and third-order RCs for both PMX and PMY exceed 0.9.

According to SSA theory, when a periodic signal exists in the series, SSA decomposes it into a pair of nearly equal eigenvalues. When the correlation coefficient between two adjacent RCs exceeds a certain threshold (typically > 0.8), they are considered to belong to the same periodic signal. Therefore, the second- and third-order RCs represent the same periodic signal, which can be identified as the Chandler wobble term based on Figures 2–3. Similarly, the fourth- and fifth-order RCs of both PMX and PMY components have relatively close eigenvalue contribution rates and correlation coefficients greater than 0.8, indicating they belong to the same periodic signal representing the annual wobble term. RCs from the sixth order onward have relatively small contribution rates (below 2%), with decreasing contributions at higher orders. These RCs exhibit quasi-periodic oscillations with less regular patterns compared to the first five orders, representing sub-seasonal quasi-periodic signals. The oscillation characteristics of each RC shown in Figures 2–3 correspond well with the eigenvalue contribution rates listed in Table 1.

Given that the first five RCs have large eigenvalue contribution rates and belong to low-frequency signals, they are linearly superimposed to constitute the low-frequency component of polar motion, while the remaining RCs are superimposed to form the high-frequency component, as shown in Figures 5 [Figure 5: see original paper] and 6 [Figure 6: see original paper]. The low-frequency component constructed from the first five RCs matches the original polar motion series well, capturing the overall variation trend. The high-frequency component formed by the sum of remaining RCs exhibits irregular nonlinear variations, making it suitable for prediction using time series analysis models.

To detect periodic oscillation signals in both the low-frequency component and original polar motion series, Fast Fourier Transform (FFT) spectral analysis was performed. The results, also shown in Figures 5–6, reveal that both the original series and low-frequency component contain Chandler wobble with a period of 438 days and annual wobble with a period of 365 days, confirming that the main periodic oscillations in the low-frequency component match those of the original series. The spectra also show smaller-amplitude interannual signals with periods around 500 days. These results demonstrate that SSA effectively separates and extracts low-frequency and high-frequency components from polar motion series.

3 Prediction Model

Unlike direct SSA-based polar motion prediction, this study fully considers the influence of low-frequency and high-frequency components. By leveraging SSA's capability to separate and identify periodic components of complex signals—particularly its excellent performance in improving the resolution of periodic components—we construct an LS+AR combined model to predict low-frequency and high-frequency components separately, thereby further improving polar motion prediction accuracy. The SSA-based polar motion prediction workflow is as follows: First, SSA separates and extracts low-frequency components (including Chandler wobble, annual wobble, and long-term trend) and high-frequency components from the polar motion series. Next, an LS extrapolation model is established for the low-frequency component, where a segment of the low-frequency component series is used to solve for prior model parameters via least squares fitting, yielding extrapolated values and LS residuals. Then, an AR model predicts the sum of high-frequency components and LS residuals. The final polar motion prediction is the sum of the LS extrapolation and AR prediction values. Figure 7 [Figure 7: see original paper] illustrates this workflow.

3.1 Least Squares Extrapolation Model

For the low-frequency component of polar motion, the LS extrapolation model includes linear trend, Chandler wobble, and annual wobble terms. The prior model is mathematically expressed as:

$$f(t) = a + bt + c_1 \cos(\varpi_1 t) + d_1 \sin(\varpi_1 t) + c_2 \cos(\varpi_2 t) + d_2 \sin(\varpi_2 t)$$

where $t = 1, 2, \cdots, N$, N represents the length of the polar motion observation series, ϖ_1 and ϖ_2 denote the angular frequencies of Chandler and annual wobbles, respectively; a and b are trend parameters; c_1 and d_1 are Chandler wobble parameters; and c_2 and d_2 are annual wobble parameters. These six unknown parameters are determined through least squares estimation.

3.2 Autoregressive Model

For a stationary random time series $\{u_t, t = 1, 2, \cdots, N\}$, an autoregressive model describes the linear relationship between the current value and the previous p values plus an error term. The p-order AR model is expressed as:

$$u_t = \sum_{i=1}^{p} \phi_i u_{t-i} + \xi_t$$

where ξ_t represents Gaussian white noise and ϕ_i are AR model coefficients ($i = 1, 2, \cdots, p$), determined by solving the Yule-Walker equations in this study.

The order p of the AR model affects its generalization capability. This study employs Akaike's Information Criterion (AIC) to optimize the selection of the AR model order:

$$\mathrm{AIC}(p) = \ln \sigma_p^2 + \frac{2p}{N}$$

where σ_p represents the standard deviation of the AR model fitting residuals. The optimal order p corresponds to the minimum AIC value across $p = 1, 2, \cdots, N$.

4.1 Accuracy Evaluation Metrics

Prediction accuracy is assessed using absolute error (AE) and mean absolute error (MAE). The calculation methods are expressed as:

$$\mathrm{AE}_{h,c} = |P_c(h) - O_c(h)|$$

$$\mathrm{MAE}_h = \frac{1}{n} \sum_{c=1}^{n} \mathrm{AE}_{h,c}$$

where h represents the prediction length, n represents the number of prediction epochs, and $P_c(h)$ and $O_c(h)$ denote the predicted and observed values at epoch c, respectively.

4.2 Comparison with LS+AR Model

Polar motion data from January 1, 2003, to October 25, 2019, were used for prediction experiments, with the period from January 1, 2015, to October 25, 2019, serving as the prediction interval. Predictions were made every 7 days using a 12-year baseline for model construction, with the modeling data updated after each prediction, totaling 200 prediction epochs.

To verify the improvement of SSA on the LS+AR combined model, both LS+AR and SSA+LS+AR models were used for 1–365 day predictions. Figure 8 [Figure 8: see original paper] shows the prediction AE for both models during 2015–2019. The maximum AE for both methods occurs in 2017, significantly larger than other years, possibly related to the La Niña event that year. The LS+AR model's maximum AE exceeds 60 mas for PMX and 90 mas for PMY, while SSA+LS+AR reduces these to below 50 mas and 80 mas, respectively. The AE of LS+AR exhibits periodic effects, indicating that direct LS fitting of the polar motion series cannot fully capture Chandler and annual wobble characteristics, leaving residual periodic fitting and extrapolation errors. In contrast, SSA+LS+AR uses different models for low- and high-frequency components, reducing prediction AE.

To quantitatively analyze the improvement, Table 2 presents MAE for both models across different prediction lengths. SSA+LS+AR improves accuracy at all spans, with improvements within 10% for ultra-short-term predictions (1–10 days). Starting from 30 days, improvements become more pronounced, reaching maximum improvements of 32.89% and 23.14% for PMX and PMY, respectively. PMX accuracy improvements remain above 20%, while PMY improvements stay above 15%. The more significant improvement in medium- and long-term predictions confirms that Chandler and annual wobbles primarily affect medium- and long-term forecasts, validating SSA's effectiveness in extracting these signals.

4.3 Comparison with IERS Bulletin A

To further validate the SSA model's improvement, SSA+LS+AR predictions are compared with IERS Bulletin A predictions from weeks 1–52 of 2019 for the period January 1, 2019–December 25, 2020. Using IERS EOP 14 C04 series as reference, MAE values are calculated for both methods, as shown in Figure 9 [Figure 9: see original paper]. For short-term predictions (1–30 days), the proposed method's accuracy is lower than IERS Bulletin A. However, for medium- and long-term predictions beyond 30 days, IERS Bulletin A errors increase rapidly with prediction length, while the SSA+LS+AR method maintains higher accuracy, significantly outperforming IERS Bulletin A. Maximum improvements reach 45.97% and 62.44% for PMX and PMY, respectively. The better performance for PMY is primarily due to stronger regularity in PMY data, though prediction results also depend on the specific prediction period.

5 Summary and Outlook

Addressing the complex time-varying characteristics of polar motion, this study introduces SSA to separate and extract low-frequency and high-frequency components. The singular spectrum analysis demonstrates that SSA effectively separates these components. By establishing an LS+AR combined model that accounts for the distinct influences of low- and high-frequency components, prediction accuracy is improved, particularly for medium- and long-term forecasts. Compared with IERS Bulletin A predictions, the proposed method shows lower accuracy for short-term predictions (1–30 days) but significantly superior performance for medium- and long-term predictions beyond 30 days, with maximum improvements of 45.97% and 62.44% for PMX and PMY components, respectively. These results validate the effectiveness of SSA in improving medium- and long-term polar motion predictions.

Acknowledgments
We thank the reviewers for their valuable suggestions that significantly improved the quality of this paper. We also thank IERS for providing polar motion data.

References

[1] Kehm A, Hellmers H, Blobfeld M, et al. JGeod, 2023, 97:
[2] Dill R, Dobslaw H. JGeod, 2010, 84: 529
[3] Su X Q, Liu L T, Houtse H, et al. JGeod, 2014, 88: 145
[4] Bizouard C, Lambert S, Gattano C, et al. JGeod, 2019, 93: 621
[5] Zhao D N, Lei Y. StGG, 2020, 64: 293
[6] Wu F, Chang G B, Deng K Z, et al. AcGG, 2019, 54: 557
[7] Wang G C, Liu L T, Tu Y, et al. StGG, 2018, 62: 243
[8] Akyilmaz O, Kutterer H, Shum C K, et al. Applied Soft Computing, 2011, 11: 837
[9] Kosek W, McCarthy D D, Luzum B J. JGeod, 1998, 72:
[10] 赵鑫, 吴元伟, 杨新宇, 等. 天文学报, 2022, 63: 47
[11] Zhao X, Wu Y W, Yang X Y, et al. ChA&A, 2022, 46:
[12] Kalarus M, Schuh H, Kosek W, et al. JGeod, 2010, 84:
[13] Zhao D N, Lei Y. StGG, 2019, 63: 83
[14] Lei Y, Cai H B, Zhao D N. G&G, 2017, 8: 141
[15] Vautard R, Yiou P, Ghil M. PhyD, 1992, 58: 95
[16] 周茂盛, 郭金运, 沈毅, 等. 地球物理学报, 2018, 61: 4383
[17] Chen Q, van Dam T, Sneeuw N, et al. JGeo, 2013, 72:
[18] 郭金运, 高文宗, 于红娟, 等. 地球物理学报, 2018, 61: 3889
[19] Biz H. JGeod, 2008, 82: 871
[20] Ding H, Shen W B. NPGeo, 2015, 22: 473
[21] Cryer J D, Chan K S. Time Series Analysis with Application with R. New York: Springer, 2009: 249-276
[22] Akaike H. Annals of the Institute of Statistical Mathematics, 1971, 23: 163

Note: Figure translations are in progress. See original paper for figures.