EOP Time Series Prediction Using Singular Spectrum Analysis

Grigory Okhotnikov[1] and Nina Golyandina[2]

[1] ITMO University, 49 Kronverksky avenue, St. Petersburg, 197101, Russia
gokhotnikov@itmo.ru
[2] St. Petersburg State University, 7/9 Universitetskaya nab., St. Petersburg, 199034, Russia n.golyandina@spbu.ru

Abstract. Accurate forecasting of Earth orientation parameters (EOP) is important for improving the GPS location accuracy and navigation of Earth satellites. EOP time series include periodic components of complex structure. Singular Spectrum Analysis (SSA) is a nonparametric method that is capable of decomposing and forecasting time series with sine-wave components. In the paper, a unified approach to choosing parameters of the SSA forecasting algorithm for EOP time series prediction is proposed. EOP time series data published by IERS in Bulletin 14 C04 are used for 365-days prediction. The forecasts performed by the proposed techniques are compared with predictions taken from available public sources.

Keywords: singular spectrum analysis · Earth orientation parameters · time series · forecasting.

1 Introduction

Earth orientation parameters (EOP) are a collection of parameters that describe different aspects of Earth rotation. They are changing in time and therefore can be considered as time series. In many applications, such as geolocation or high-precision satellite navigation, not only observed EOP values are needed, but also their predictions for several days in the future. Namely, there are five time series of interest: coordinates of the pole (x, y), length of day LOD, and celestial pole offsets (dX, dY). The coordinates (x, y) of the celestial ephemeris pole determine the position of the celestial pole on the Earth's surface. The pole moves slowly because the axis of the Earth's instantaneous rotation does not stay still. Universal time $(UT1)$ is the time of the Earth clock, which performs one revolution in about 24 hours. The excess revolution time is called length of day (LOD). It is vulnerable to wind and world ocean movements; thus it has a complex structure and cannot be modeled precisely. Celestial pole position is described by IAU (International Astronomical Union) precession and nutation model IAU2000. The approximation accuracy is high; however, some nutations are still not predictable. The observed corrections to the modeled celestial pole coordinates in the celestial reference system are the offsets (dX, dY).

The Earth orientation parameters prediction comparison campaign (EOP PCC, 2005–2008) organized for comparing the predictions of Earth orientation parameters was an attempt to compare existing prediction techniques in the same period of time for various lengths of forecasts: 10, 30, and 500 days. The results of the campaign [7] suggest that smaller prediction errors can be achieved by ensembling various models. A similar initiative was undertaken by the International Earth Rotation and Reference Systems Service (IERS) in 2010–2015; unfortunately, the project's website and its archive are not available at the moment.

Various methods have been applied to the problem of EOP time series forecasting. Least squares interpolation by means of a harmonic model and autoregressive prediction (LS+AR) was applied to x, y and LOD time series in [10, 9]. A seasonal autoregressive model was proposed for the prediction of dX and dY in [11].

Singular spectrum analysis (SSA) was used in several recent works for the prediction of the polar motion (x, y). The applicability of the method to the problem was demonstrated in [12] where the parameters were fixed to manually chosen values. The paper [13] proposes a more flexible approach with a combination of SSA and the copula-based analysis.

In this paper, we introduce a technique for (almost) fully automated procedure for the choice of parameters for each of the five EOP time series predictions, which are performed on the base of historical data by means of singular spectrum analysis. The set of parameters consists of both parameters of SSA and the length of time series to forecast. The used approach to the parameter choice is commonly used in machine learning; therefore, it is particularly interesting to note that the forecasting accuracy appears to be comparable with other public forecasts, as it is demonstrated in this paper.

The structure of the paper is as follows. We start with a brief description of SSA including the SSA forecasting and the approaches to the parameter choice (Section 2). In Section 3, the used EOP data are presented. Section 4 contains a description of the techniques of the automatic parameter choice. In Section 5, the proposed approach is applied to the EOP data and the forecasting accuracy is compared with several prediction results, which were found in the public domain. Section 6 concludes the paper.

2 SSA methodology

SSA is a nonparametric method that is capable of decomposing a time series into additive components: signal (trend plus periodic components) and noise [6]. SSA consists of two stages: decomposition and reconstruction.

Decomposition Let us consider a real-valued time series $\mathsf{X}_N = (x_1, \ldots, x_N)$ of length N. Let L $(1 < L < N)$ be some integer (it is called "window length"), and $K = N - L + 1$. On the decomposition stage, K embedding vectors are constructed in the following way:

$$X_i = (x_i, \dots, x_{i+L-1})^{\mathrm{T}} \in \mathbb{R}^L, \quad i = 1 \dots, K. \tag{1}$$

The trajectory matrix of the time series X_N is defined as the Hankel matrix obtained by column-wise stacking the embedding vectors:

$$\mathbf{X} = [X_1 : \ldots : X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_K \\ x_2 & x_3 & x_4 & \ldots & x_{K+1} \\ x_3 & x_4 & x_5 & \ldots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \ldots & x_N \end{pmatrix}. \tag{2}$$

Denote $\lambda_1, \dots, \lambda_L$ the eigenvalues of $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$, which are ordered in non-decreasing order ($\lambda_1 \geq \ldots \geq \lambda_L \geq 0$), and $U_1, \dots, U_L$ an orthonormal basis of eigenvectors of the matrix $\mathbf{S}$, corresponding to these eigenvalues. Let $d = \max\{i : \lambda_i > 0\}$. Denote $V_i = \mathbf{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i}, \quad i = 1, \dots, d$; then the singular value decomposition (SVD) of the matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d, \quad \mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}. \tag{3}$$

The collection $(\sqrt{\lambda_i}, U_i, V_i)$ is called the i-th eigentriple of the SVD (3).

Reconstruction Basing on the decomposition (3), we perform a grouping procedure that divides the whole set of indices $\{1, \dots, d\}$ into m non-intersecting groups $I_1, \dots, I_m$. The grouped matrix $\mathbf{X}_I$ that corresponds to a group $I = \{i_1, \dots, i_p\}$ is defined as follows:

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}.$$

In this notation, the grouped matrix decomposition of (3) has the form

$$\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m}. \tag{4}$$

The procedure of the choice of $I_1, \dots, I_m$ is called eigentriples grouping.

Let $\mathbf{X}_{I_k}$ be a grouped matrix from (4). Then its transformation to a time series is performed by averaging of matrix elements along diagonals $i+j = \mathrm{const}$. After diagonal averaging of $\mathbf{X}_{I_k}$, the k-th reconstructed time series component $\widetilde{\mathsf{X}}^{(k)} = (\widetilde{x}_1^{(k)}, \dots, \widetilde{x}_N^{(k)})$ is obtained. Thus, the original series $(x_1, \dots, x_N)$ can be denoted as a sum of m reconstructed series:

$$x_n = \sum_{k=1}^{m} \widetilde{x}_n^{(k)}, \quad n = 1, \dots, N.$$

If $\mathsf{X}_N = \mathsf{S}_N + \mathsf{R}_N$ and we are interested in the signal S_N, while R_N is noise, the choice $I_1 = \{1, \dots, r\}$ for some $r < d$ is appropriate. Then the signal estimate is $\widetilde{\mathsf{S}}_N = (\widetilde{s}_1, \dots, \widetilde{s}_N) = \widetilde{\mathsf{X}}^{(1)}$.

Forecasting Let us forecast the signal, which was estimated at the reconstruction stage of SSA with the window length L by grouping the r leading components. The recurrent SSA forecasting is based on the r leading eigenvectors $U_1, \ldots, U_r$ and the reconstructed series $\widetilde{\mathsf{S}}_N$.

Denote $\underline{U_i}$ the first $L-1$ coordinates of U_i and π_i the last coordinate of U_i, $\nu^2 = \sum_i \pi_i^2$. Define $R = (a_{L-1}, \ldots, a_1)^{\mathrm{T}}$ as

$$R = \frac{1}{1-\nu^2} \sum_{i=1}^{r} \pi_i \underline{U_i}. \tag{5}$$

The recurrent forecasting algorithm [6] is:

1. The predicted time series $\mathsf{Y}_{N+P} = (y_1, \ldots, y_{N+P})$ is defined as

$$y_i = \begin{cases} \widetilde{s}_i & \text{for } \ i = 1, \ldots, N, \\ \sum\limits_{j=1}^{L-1} a_j y_{i-j} & \text{for } \ i = N+1, \ldots, N+P, \end{cases} \tag{6}$$

2. The values $y_{N+1}, \ldots, y_{N+P}$ are the result of P-step ahead forecasting.

Thus, the recurrent SSA forecasting is performed by applying the linear recurrence relation with the coefficients $\{a_j, j = 1, \ldots, L-1\}$.

Common consideration about the choice of the SSA parameters Certainly, the forecasting accuracy depends on the proper choice of parameters. SSA can be considered as both a parametric and a non-parametric method. If the trajectory matrix of a signal S_N is rank-deficient, the signal is called a time series of finite rank. Time series of finite rank (under some unrestrictive limitations) can be exactly forecasted by SSA forecasting methods. For example, sine waves and their sums have finite rank. For the SSA forecasting of periodic series, the period should not be known in advance, due to the non-parametric nature of SSA. The choice $L \sim N/2$ is recommended for forecasting finite-rank signals in presence of noise.

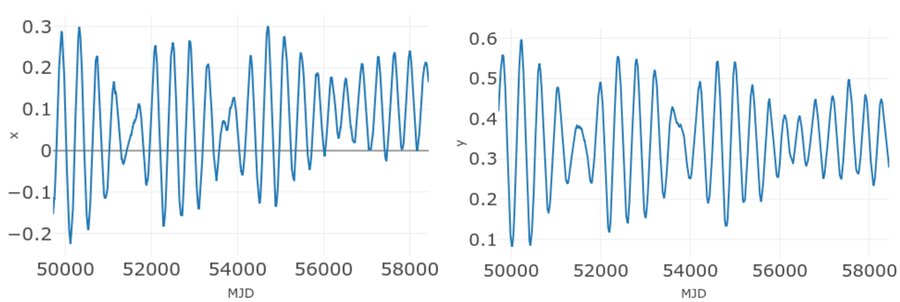
If the signal is an amplitude-modulated sine wave, it is not necessarily exactly of finite rank. However, such a signal can be extracted with good accuracy by an increase of the number r of the chosen components and/or a decrease of the window length L.

In the interactive version of SSA, the identification of signal components is performed by the analysis of the decomposition eigentriples. However, if the time series is long enough, some automatic procedure for the choice of L and r can be applied to minimize forecasting errors on chosen historical data; see e.g. [5, Fragment 3.5.13–3.5.15]. Moreover, if the time series is long and has a changing structure, the question of the choice of the optimal time series length to use for forecasting arises. The problem of the choice of optimal time series length can be solved automatically for a very long time series only.

3 Data sources

As a source data we use IERS 14 C04 bulletin data published by IERS [2]. This source contains daily data for each of the EOP time series beginning from 1962, January 1, for x, y, LOD and from 1984, January 1, for dX, dY until the current date with approximately 30 days delay. The plots of the time series are shown in Figures 1–3.

IERS Bulletin A contains predictions of x, y, LOD time series for 365 days; these data are published weekly. For comparison of the forecasts, we will also use Pulkovo observatory daily predictions of all EOP time series for 365 days [3]. Unfortunately, as of June 2019, the online archive with Pulkovo predictions is not available. Further, in this work, we will use a backup copy of those files that we possess at the moment.



(a) x (b) y

Fig. 1: Examples of celestial pole time series from 1995 to present.

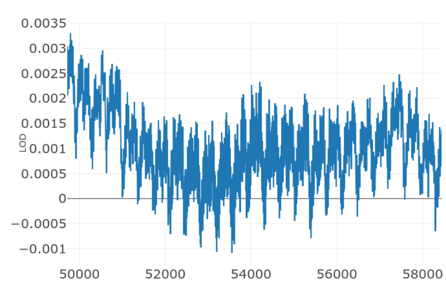
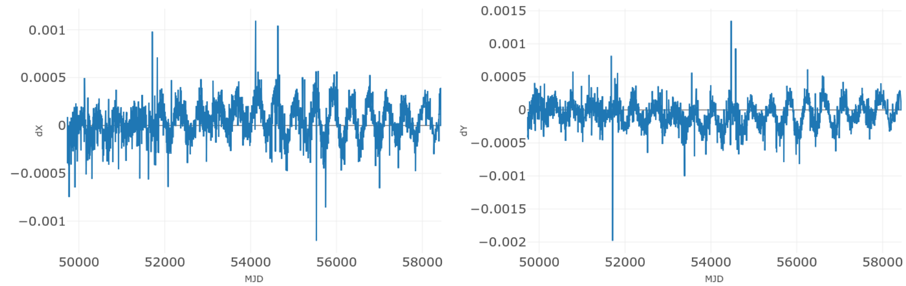


Fig. 2: Example of LOD time series from 1995 to present.



(a) dX (b) dY

Fig. 3: Examples of celestial pole offsets time series from 1995 to present.

4 Automatic choice of parameters

We will perform forecasting of the EOP time series x, y, LOD, dX, dY for one year. For simplicity, we consider one year as consisting of 365 days.

As we mentioned, the SSA forecasting of a signal needs two parameters: the window length L and the number of leading components r. The latter is related to the signal structure, which is determined by the signal rank. The following assertions are related: the signal structure is more complex, a larger rank is used for its approximation, a larger number r of components in the SSA decomposition should be chosen. We will search for an optimal r^* value that minimizes forecast mean squared error (MSE) for historical data. The considered boundaries are $r^* \leq 30$ for the x, y, LOD time series, and $r^* \leq 5$ for the dX, dY time series since they have a simpler structure. These boundaries were checked by the manual SSA analysis of several sample data. The boundaries for L were chosen in a similar manner.

We use the grid-search for the optimal value L^* performing up to 10 steps within the range. The chosen grid for L and r values is shown in Table 1.

Table 1: Search grid for L and r parameter values.

EOP	L **values**	r **values**
x, y	300, 500, 700, 900, 1100, 1300, 1500, 1700, 1900, 2100	[1, 30]
LOD	300, 600, 900, 1200, 1500, 1800, 2100, 2400, 2700, 3000	[1, 30]
dX, dY	250, 300, 350, 400, 450, 500	[1, 5]

In order to choose the parameters L^*, r^* for SSA forecasting we will use the time series cross-validation procedure as described in [5, Section 3.5.7].

For performing the cross-validation, we should fix the training period. Within the training period, we consider moving intervals of length $Q+1$ years, where the interval of Q years ($365 \times Q$ days) is used for the signal estimation and one year

(365 days) is taken for the signal forecasting and error calculation. The training period length is equal to Q years plus the length W of the validation period; the validation period consists of the time series values, which are used for the calculation of the forecasting errors that are involved in the cross-validation procedure. For decreasing the computational costs, we consider M moving intervals of length $Q+1$ years with equal lags; thereby, the lag size is approximately equal to $(W-1)/(M-1)$. For the SSA forecasting, the values L^*, r^* are chosen from the grid to minimize the forecast MSEs averaging by all the moving intervals.

In [9], the values $Q = 5, 10, 15$ years were considered. We compare $Q = 5, 10, 15, 20$ years (if the time series lengths allow this choice) for the validation period length equal to 5 years. The optimal parameters L^*, r^* are chosen automatically using the cross-validation procedure as described above with $M = 10$ folds.

The implementation of the SSA algorithms in R language from the Rssa package [8] was used. The average mean-squared errors of the forecasts are shown in Table 2. Among the considered values, the value $Q = 15$ years results in smaller forecast errors. We will use this setting later for performing forecasts on the test period.

Table 2: Average MSE of 365 days EOP forecasts for different Q in the 2006–2010 years interval.

EOP	**5 years**	**10 years**	**15 years**	**20 years**
x	2.6×10^{-3}	9.5×10^{-4}	8.5×10^{-4}	1.1×10^{-3}
y	1.9×10^{-3}	1.2×10^{-3}	9.0×10^{-4}	1.1×10^{-3}
LOD	2.4×10^{-7}	1.9×10^{-7}	1.0×10^{-7}	1.0×10^{-7}
dX	2.01×10^{-8}	1.93×10^{-8}	1.88×10^{-8}	—
dY	1.63×10^{-8}	1.53×10^{-8}	1.42×10^{-8}	—

Table 3: Average MSE of 365 days EOP forecasts for different validation period lengths in the 2006–2010 years interval.

EOP	**3 years**	**5 years**	**7 years**	**10 years**
x	1.0×10^{-3}	8.5×10^{-4}	8.3×10^{-4}	9.1×10^{-4}
y	9.2×10^{-4}	9.0×10^{-4}	8.8×10^{-4}	8.8×10^{-4}
LOD	1.1×10^{-7}	1.0×10^{-7}	1.0×10^{-7}	9.8×10^{-8}
dX	1.91×10^{-8}	1.88×10^{-8}	1.88×10^{-8}	—
dY	1.42×10^{-8}	1.42×10^{-8}	1.39×10^{-8}	—

We also need to choose the validation period length. To do this, Q is fixed to that chosen on the previous step (15 years) and then the forecast errors are compared using the cross-validation with $M = 10$ folds for different lengths of

Table 4: Average MSE of 365 days EOP weekly forecasts from different sources in the 2011–2015 years interval.

EOP	**SSA**	**Pulkovo AM**	**Bulletin A**
x	7.2×10^{-4}	8.6×10^{-4}	7.5×10^{-4}
y	6.1×10^{-4}	7.6×10^{-4}	8.5×10^{-4}
LOD	9.1×10^{-8}	1.0×10^{-7}	—
dX	1.3×10^{-8}	1.1×10^{-8}	—
dY	1.6×10^{-8}	2.2×10^{-8}	—

the validation period: 3, 5, 7, and 10 years. The results of the experiments are presented in Table 3. We also checked (not shown) that the optimal r^* and L^* values fall in the same intervals for different forecasts along the considered training period. This confirms that their choice is not random.

5 Forecasts on the test period

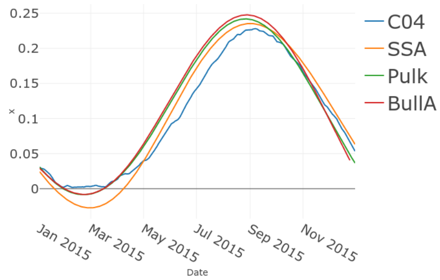
To perform forecasts for 365 days in the test period (2011–2015 years), we used the results of investigations described in Section 4. The lengths of the training and validation periods were chosen with the help of Tables 2 and 3, respectively. For forecasts, the parameters L^*, r^* were chosen automatically using the cross-validation procedure as described in Section 4 (see Figure 4 for examples of forecasts). The forecasting accuracy was compared with that of predictions from two sources available in the public domain: Pulkovo observatory [3] and IERS Bulletin A [1].

Since Bulletin A forecasts are published once a week, we generate forecasts each week starting from the dates of publications for each of the EOP time series. Then we calculate mean squared errors for all the forecasts and take the average MSE in the test period. The results are shown in Table 4. In most cases, the proposed method demonstrates a better average performance, except for the dX time series predictions published by Pulkovo observatory. However, the average MSEs of SSA and Pulkovo forecasts for dX are of the same order.

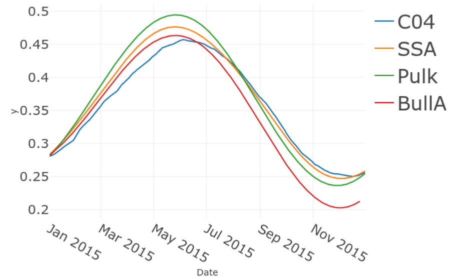
6 Conclusions

In this paper, we applied the SSA recurrent forecasting algorithm to predicting the EOP time series. We stated the problem as a machine learning problem and found a set of learning parameters, including the cross-validation hyper-parameters. The proposed approach can be applied to each of the EOP time series: x, y, LOD, dX and dY in the same manner.

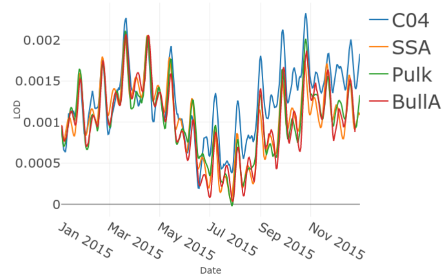
It appears that SSA is well suitable for the prediction of the EOP time series. Since the computational cost of SSA is considerably diminished in the recent implementation [8, 5], the proposed approach is computationally feasible for daily EOP time series predictions. The numerical comparison in the test period since



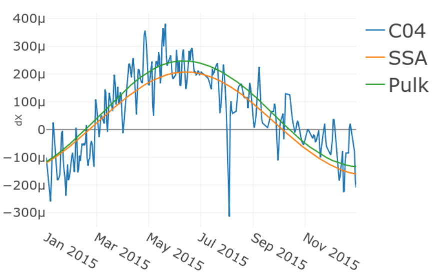
(a) x time series forecast ($L^* = 700, r^* = 10$).



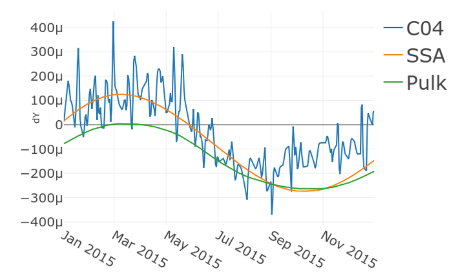
(b) y time series forecast ($L^* = 500, r^* = 10$).



(c) LOD time series forecast ($L^* = 900, r^* = 19$).



(d) dX time series forecast ($L^* = 350, r^* = 4$).



(e) dY time series forecast ($L^* = 300, r^* = 5$).

Fig. 4: EOP time series forecasts for 365 days starting from January 1 2015 from different sources.

January 1, 2011, till December 31, 2015, shows that SSA together with the proposed method of automatization of the parameter choice provides forecasting accuracy, which is comparable to those that are published by international services; in many cases, the proposed method's errors are smaller on average.

The resulting forecasts and corresponding parameter values are published as a web-application and can be downloaded via http://eoppredict.ru. The interface allows one to compare predictions from different sources on historical data.

In future work, the prediction accuracy can be improved by a more sophisticated grouping of the SSA decomposition components using specific time series properties, in an automated way. Also, the multivariate versions of SSA [4] can be applied to improve accuracy.

References

1. Bulletin A — all available versions, https://www.iers.org/IERS/EN/Publications/Bulletins/bulletins.html
2. International Earth rotation and reference systems service earth orientation parameters EOP (IERS) 14 C04, https://hpiers.obspm.fr/iers/eop/eopc04/eopc04_IAU2000.62-now
3. Pulkovo EOP and reference systems analysis center (PERSAC), http://www.gao.spb.ru/english/as/persac/eopcppp/
4. Golyandina, N., Korobeynikov, A., Shlemov, A., Usevich, K.: Multivariate and 2D extensions of singular spectrum analysis with the Rssa package. Journal of Statistical Software **67**, 1–78 (2015)
5. Golyandina, N., Korobeynikov, A., Zhigljavsky, A.: Singular Spectrum Analysis with R. Springer-Verlag Berlin Heidelberg (2018), https://ssa-with-r-book.github.io
6. Golyandina, N., Nekrutkin, V., Zhigljavsky, A.A.: Analysis of time series structure: SSA and related techniques. Chapman and Hall/CRC (2001)
7. Kalarus, M., Schuh, H., Kosek, W., Akyilmaz, O., Bizouard, C., Gambis, D., Gross, R., Jovanović, B., Kumakshev, S., Kutterer, H., et al.: Achievements of the earth orientation parameters prediction comparison campaign. Journal of Geodesy **84**, 587–596 (2010)
8. Korobeynikov, A., Shlemov, A., Usevich, K., Golyandina, N.: Rssa: A collection of methods for singular spectrum analysis (2017), https://cran.r-project.org/package=Rssa, R package version 1.0
9. Kosek, W.: Future improvements in EOP prediction. In: Geodesy for Planet Earth, vol. 136, pp. 513–520. Springer (2012)
10. Malkin, Z.: Employing combination procedures to short-time EOP prediction. Artificial Satellites **45**(2), 87–93 (2010)
11. Malkin, Z.: Free core nutation and geomagnetic jerks. Journal of Geodynamics **72**, 53–58 (2013)
12. Miller, N.: Forecasting pole motion with SSA (in Russian). vol. 223, pp. 119–123 (2016)
13. Modiri, S., Belda, S., Heinkelmann, R., Hoseini, M., Ferrándiz, J.M., Schuh, H.: Polar motion prediction using the combination of SSA and copula-based analysis. Earth, Planets and Space **70**(1), 115 (2018)