

SIGNAL PROCESSING BY MEANS OF “CATERPILLAR” – SSA METHOD

Svetlana N. Polukoshko*¹, Galina I. Hillekica²

¹Ventspils University College, Engineering Research Institute “VSRC”, Ventspils, Latvia
pol.svet@inbox.lv

²Ventspils University College, Engineering Research Institute “VSRC”, Ventspils, Latvia
galinah@venta.lv

Keywords: Singular Spectrum Analysis, Time series, Sequential algorithm, Singular value decomposition, Forecast.

***Abstract.** This paper deals with “Caterpillar”-SSA method - a new powerful method of time series analysis and forecasting, which was independently developed in Russia (under the name of “Caterpillar”) and also in UK and USA (under the name of SSA - Singular Spectrum Analysis). This model-free method intended to study the structure of the time series, combines the advantages of many other methods, such as Fourier analysis and regression analysis and at the same time it is characterized by clarity and controllability.*

The method can be successfully used for processing the signals from the vibrating structure elements and machine parts. Alongside with signal processing this method is successfully used to study the time series in many various areas: in medicine, meteorology, hydrology, sociology, economics, traffic analysis, wherever the trend or periodic behavior can present. This paper shows how SSA method may be realized in MATHCAD program, which is a convenient tool for the engineers.

1 INTRODUCTION

“Caterpillar” method is new powerful method of time series analysis and forecasting, which was independently developed in Russia and also in UK and USA (under the name of SSA - Singular Spectrum Analysis) [1-5]. The “Caterpillar”- SSA is a model-free technique of time series analysis. It combines advantages of other methods, such as Fourier and regression analyses, with simplicity of visual control aids. The “Caterpillar”-SSA approach is used to study the time series in many various spheres wherever the trend or periodic behavior can present, for instance, in meteorology, climatology, hydrology, geophysics, medicine, biology, sociology [6-9]. SSA technique can be applied for various time series arising in industry and economics.

A lot of problems can be solved by means of “Caterpillar”-SSA technique: finding trends of different resolution, smoothing, extraction of the seasonality components, change-point detection, simultaneous extraction of cycles with small and large period, simultaneous extraction of complex trends and periodicities; extraction of periodicities with varying amplitudes, finding structure in short time series, continuation of the time series [5, 6, 10].

A modification of singular spectrum analysis for time series with missing data is developed and successfully tested. This method also can be used to low pass filter incomplete time series [11, 12]. Method is extensively developing, the variants of the “Caterpillar”-SSA method for analysis of multidimensional time series (MSSA), detection of structural changes are elaborated [13-15].

The aim of SSA is to make a decomposition of the original series into the sum of a small number of independent and interpretable components such as a slowly varying trend, oscillatory components and a structureless noise. The “Caterpillar”-SSA method is useful tool for short and long, one-dimensional and multidimensional, stationary and nonstationary, almost deterministic and noisy time series analyzing.

However, in mechanics application this method does not become widely used, but it may be successfully used for periodic motion detection, amplitude and frequency definition, as well as chaotic motion identification.

2 SSA PROCEDURE

The basic version of method consists in transformation of one-dimensional series into multi-dimensional by one-parameter translation procedure, research of the got multidimensional trajectory by means of principal components analysis (Singular Value Decomposition) and reconstruction of the series in accordance with the chosen principal components.

Thus, the result of application of method is expansion of the time series into sum of simple components: slow trends, seasonal and other periodic or oscillating components, and also components of noises. The got decomposition can serve as basis of correct forecasting, both series and its components. The decomposition is followed by the reconstruction, which allows reconstructing the time series component on the base of chosen eigentriple numbers.

Both of two complementary stages of “Caterpillar” - SSA technic (decomposition and reconstruction) include two separate steps, therefore the analysis is performed by four steps.

At the first step (called the embedding step) a one-dimensional series $F = (f_0, \dots, f_{N-1})$ with length $N \geq 3$ is transferred to the L - dimensional series $X_i = (f_{i-1}, \dots, f_{i+L-2})^T$, $i=1, \dots, K$, where $K = N - L + 1$, and L is called by window length, $L \in N$, $1 < L < N$. This delay procedure gives the first name to the whole technique. Very important parameter of the embedding step is the window length L . It should be big enough but not greater than a half of series length. Vectors X_i form columns of the trajectory matrix $X = [X_1 \dots X_K]$, or:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} f_0 & f_1 & f_2 & \cdots & f_{K-1} \\ f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \cdots & f_{N-1} \end{pmatrix} \quad (1)$$

Trajectory matrix \mathbf{X} is a Hankel matrix, which mean that all the elements along the diagonal $i+j = \text{const}$ are equal.

The second step, which gives the second name of technique, SVD step, is the singular value decomposition of the trajectory matrix into a sum of rank-one bi-orthogonal elementary matrices:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d. \quad (2)$$

Elementary matrix \mathbf{X}_i is determined by the equality $\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T$, where $\sqrt{\lambda_i}$ is the square root of the i -th eigenvalue (i -th singular value) of the matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, $\mathbf{U}_1, \dots, \mathbf{U}_L$ - orthonormal set of eigenvectors of matrix \mathbf{S} , $d = \max(i; \lambda_i > 0)$, $i = 1, \dots, d$, $\mathbf{V}_i = \mathbf{X}^T \mathbf{U}_i / \sqrt{\lambda_i}$. \mathbf{U}_i and \mathbf{V}_i are called as left and right singular vectors of the trajectory matrix. It is assumed that eigenvalues λ_i are arranged in the decreasing order of their magnitude ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$). The collection $(\sqrt{\lambda_i}, \mathbf{U}_i, \mathbf{V}_i^T)$ is called the i -th eigentriple of the matrix \mathbf{X} .

The third step (grouping step) corresponds to splitting the elementary matrices \mathbf{X}_i into several groups and summing the matrices within each group. Let $I = \{i_1, \dots, i_p\}$ be a group of indices i_1, \dots, i_p . Then the matrix \mathbf{X}_I corresponding to group I is defined: $\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}$. The spilt of the set of indices $J = 1, \dots, d$ into the disjoint subsets I_1, \dots, I_m corresponds to the representation $\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}$. The procedure of choosing the sets $I = I_1, \dots, I_m$ is called the eigentriple grouping. For given group I the contribution of the component \mathbf{X}_I into the expansion is measured by the share of the corresponding eigenvalues.

The last step (diagonal averaging) transfers each resultant matrix into a time series, which is an additive component of the initial series F . Let y_{ij} stands for an element of $L \times K$ matrix \mathbf{Y} , where $1 \leq i \leq L$, $1 \leq j \leq K$, $L^* = \min(L, K)$, $K^* = \max(L, K)$, $N = L + K - 1$; $y_{ij}^* = y_{ij}$ if $L < K$ and $y_{ij}^* = y_{ji}$ otherwise. Diagonal averaging maps the matrix \mathbf{Y} into series g_0, \dots, g_{N-1} by formula:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m, k-m+2}^* & \text{for } 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+2}^* & \text{for } L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m, k-m+2}^* & \text{for } K^* \leq k < N, \end{cases} \quad (3)$$

In this way we obtain a decomposition of the initial series into several additive components.

The result is the expansion of series X_{I_k} into simple components $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$ and initial series (f_0, \dots, f_{N-1}) is decomposed into the sum of m series: $f_n = \sum_{k=1}^m \tilde{f}_n^{(k)}$. then k -th term of the resulting series is obtained by averaging of y_{ij} over all i, j such that $i + j = k + 2$.

There are two parameters in ‘‘Caterpillar’’ - SSA: the first is an integer L , the window length, and the second parameter is structural: it is the way of grouping of elementary matrices (or grouping of the eigentriples; since each matrix component of the SVD is completely determined by the corresponding eigentriple). The problem of selection of ‘‘Caterpillar’’- SSA parameters is thoroughly discussed (from theoretical and practical viewpoints) in [1, 3].

As for the way of grouping, it is useful to mention that under the proper choice of window length L singular vectors in a sense ‘repeat’ the behavior of the corresponding time series components. In particular, trend of the series corresponds to slowly varying singular vectors. Harmonic component produces a pair of left (and right) harmonic singular vectors with the same frequency, etc.

The main principles for identification of eigentriples corresponding to the time series component of interest are as follows: 1) to extract the trend or to smooth the time series, slowly changing eigenfunctions are chosen; 2) to extract oscillations, oscillating eigenfunctions are chosen; 3) to extract periodic (harmonic) components with period greater than 2 a pair of eigentriples with the same period should be found and chosen; 4) to approximate the time series the leading eigentriples are chosen.

3 NUMERICAL EXAMPLES OF USING OF ‘‘CATERPILLAR’’-SSA METHOD

3.1 Software for ‘‘Caterpillar’’-SSA technic

In recent years owing to SSA method developing, corresponding computer programmes are elaborated for time series analysis: ‘‘Caterpillar’’- SSA in St-Petersburg University [3,5], SSA-MTM Toolkit for Spectral Analysis, kSpectra Toolkit [4]. MATLAB program is also widely used. In this work Mathcad Programme are applied as easy-to-use environment for engineers and scientists research. Mathcad has a large number of the built-in functions, among them matrix functions, and programming tools. Examples some of them is presented bellow. Sub-programme for trajectory matrix composition in accordance with Eqs. (1):

$$\text{TrajektMat}(X, L) := \left| \begin{array}{l} N \leftarrow \text{length}(X) \\ n \leftarrow N - L \\ \text{for } k \in 1..L \\ \quad c^{(k)} \leftarrow \text{submatrix}[X, k, (n + k), 1, 1] \\ c^T \end{array} \right.$$

Matccad built-in function for covariance matrix (KovMat) and for eigenvalues of covariance matrix (VX) are used here.

$$\begin{aligned} \text{TrM} &:= \text{TrajektMat}(X, L) \\ \text{KovMat} &:= \text{TrM} \cdot \text{TrM}^T \\ \text{VX} &:= \text{svds}(\text{KovMat}) \end{aligned}$$

In the same way the sub-programmes for all step of SSA method realization are composed.

3.2 Electricity consumption in Latvia time series analysis

The analysis of monthly electricity consumption in Latvia from 01. 2006 till 05. 2013 is performed by means of “Caterpillar” – SSA method using Mathcad. The original statistic data (mln. kWh per month) are taken from [16]. Data matrix is a column with length of 89 months ($N = 89$). Initial time series and its trend are presented in Fig. 1. Window length is set $L=36$.

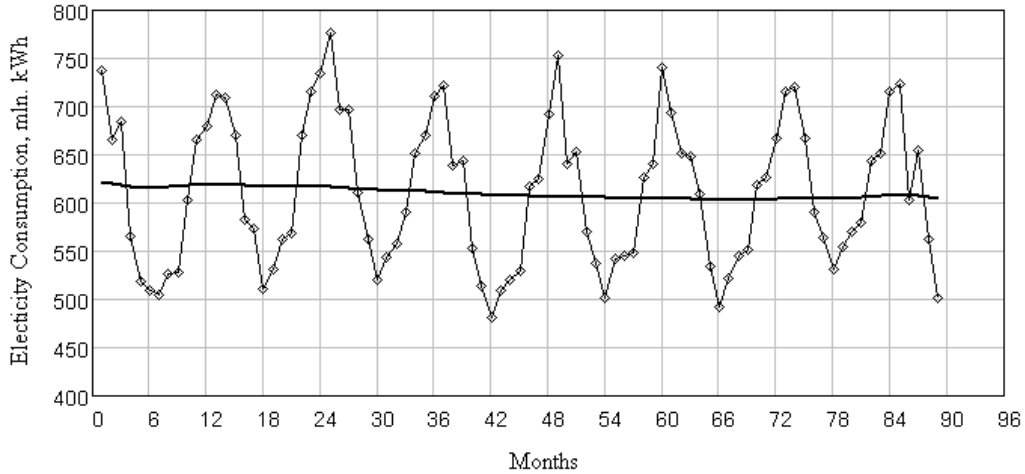


Figure 1: Initial series and its trend (1 eigentile): \diamond initial series — reconstructed series.

In Figure 2 one-dimensional plots of the first six eigenvectors are shown. Eigenvector analysis shows that the first eigenvector is a trend component (slowly changing component), second and third eigenvectors relates to low frequency components, the fourth and fifth do not carry useful information about the process.

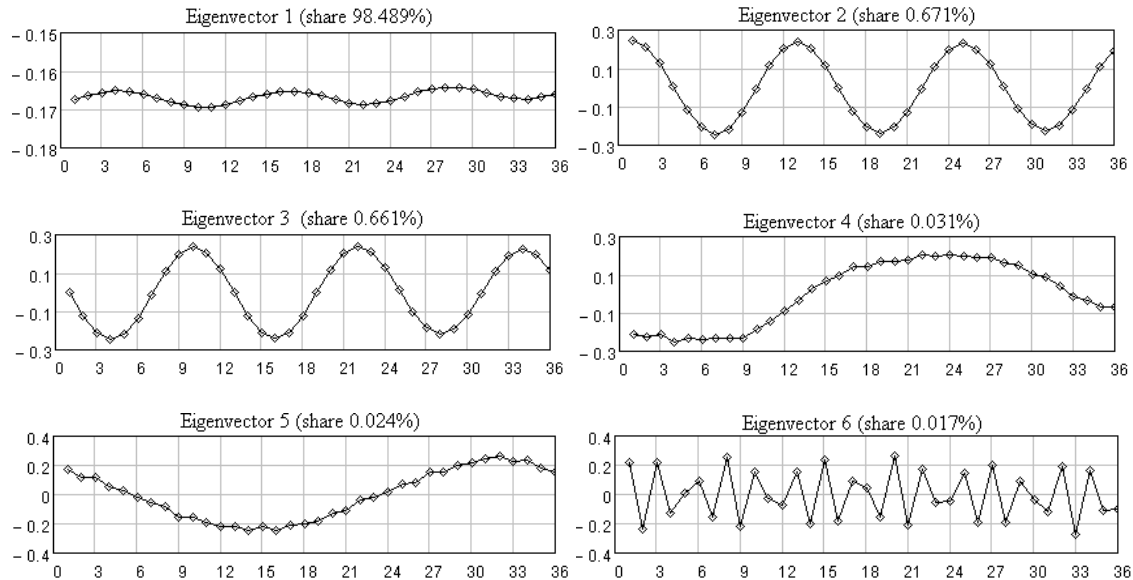


Figure 2: One-dimensional plots of the first six eigenvectors.

Figure 3 shows the two-dimensional projection of the eigenvectors; vectors number are indicated in superscript of function and argument.

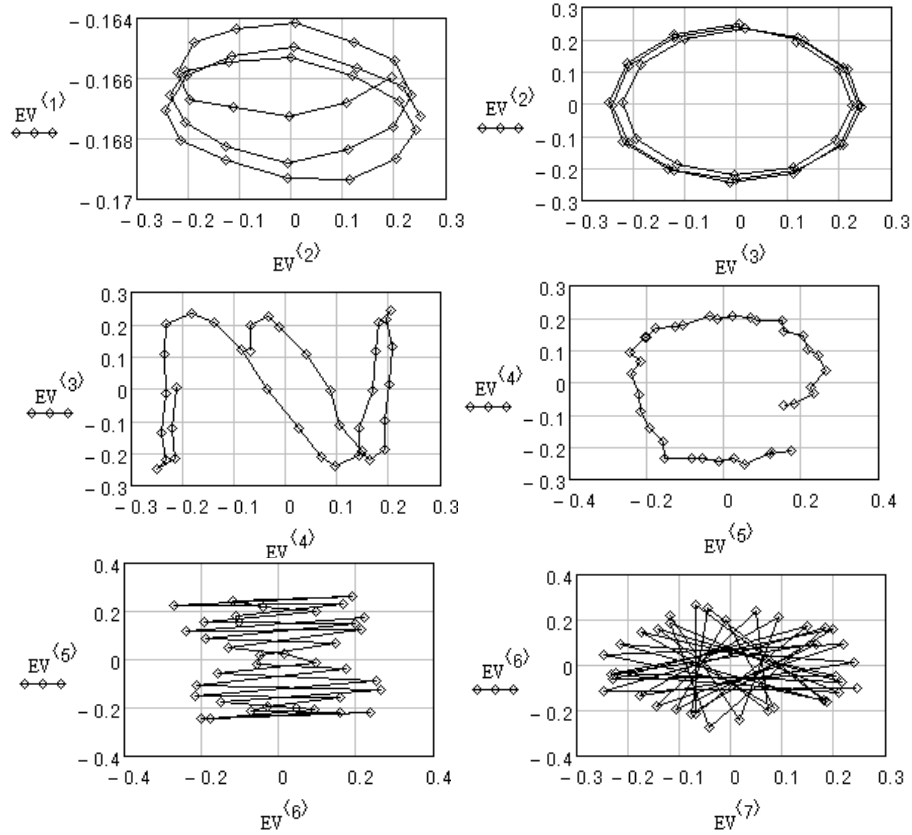


Figure 3: Two-dimensional plots of eigenvectors.

Analysis of plots in Figure 3 leads to the conclusion that the second and third eigenvectors reflect the annual cycle, the remaining two-dimensional projection do not give the useful information.

Initial and reconstructed time series with 3 eigentriple combination is presented in Figure 4 .

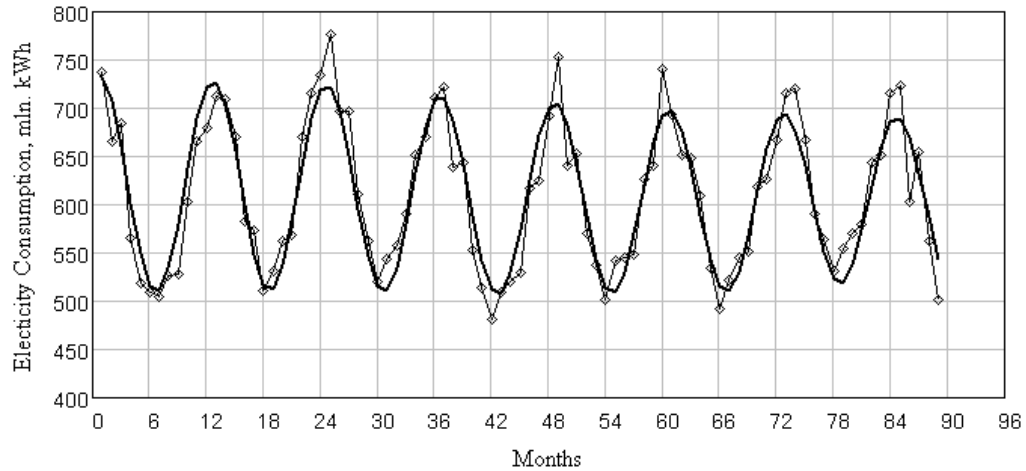


Figure 4: Initial and reconstructed series (3 eigentriple): $\diamond\diamond$ initial series — reconstructed series.

The recursive forecasting on the depth of 12 months is performed and results are shown in Figure 5.

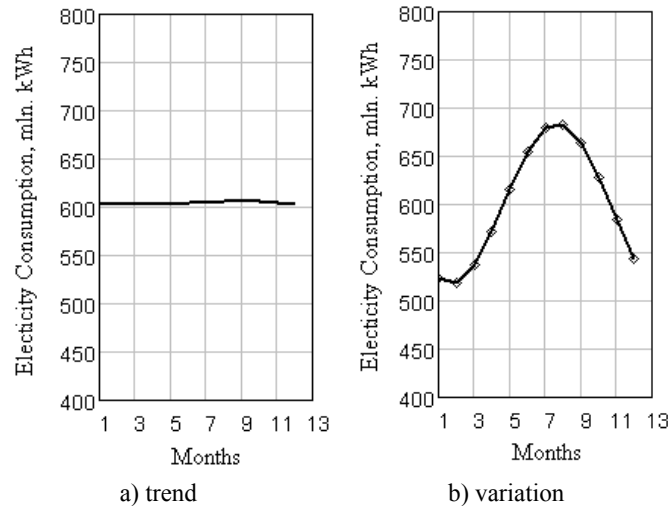


Figure 5: Electricity consumption prognosis for 12 months, received by recurrent forecasting.

To determine the size of informative basis the next features of eigenvalue is used: the larger the share of eigenvalue in the sum of elements of the eigenvalues vector, the more background information includes a projection on the corresponding eigenvector.

Window length enlargement to $L=48$ gives close results.

4 CONCLUSIONS

- Simple example of application of the “Caterpillar”-SSA technique for analysis of one-dimensional time series in are presented in this work: task of analysis and forecasting of electricity consumption time series is solved in Matchcad environment. Elaborated code is universal for one-dimensional series of any length.
- Because of a large number of the built-in matrix functions, programming simplicity, computational efficiency and visualization capabilities Mathcad is excellent instrument for “Caterpillar” - SSA technic application.
- Easy-to-use Mathcad environment is ideal for iterative learning, research and analysis of time-series.

REFERENCES

- [1] N. Golyandina, A. Zhigljavsky, *Singular Spectrum Analysis for time series*. Springer Briefs in Statistics, Springer, 2013.
- [2] I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.
- [3] D. Danilov and A. Zhigljavsky, *Principal Components of Time Series: "Caterpillar" Method*. St. Petersburg University, 1997. <http://www.gistatgroup.com/gus/>
- [4] SSA –MTM Toolkit for Spectral Analysis. <http://www.atmos.ucla.edu/tcd/ssa/>
- [5] Th. Alexandrov, N. Golyandina, Automatic extraction and forecast of time series cyclic components within the framework of SSA. *Proceedings of the 5th St.Petersburg Workshop on Simulation*, St. Petersburg State University, St. Petersburg, 45-50, 2005.

- [6] M Ghil et al., Advanced Spectral Methods for Climatic Time Series. *Reviews of Geophysics*, 40, 1, 1-1-1-41, 2002.
- [7] V. V. Antoniou et al., Principal Component Analysis of Network Traffic: the “Caterpillar” - SSA Approach, VIII Int. Workshop on “*Advanced Computing and Analysis Techniques in Physics Research*”, ACAT’2002, Book of abstracts, Moscow, 176, 2002.
- [8] H. Hassani, S. Heravi, A. Zhigljavsky, Forecasting UK Industrial Production with Multivariate Singular Spectrum Analysis, *The 2012 International Conference on the Singular Spectrum Analysis and its applications, Nay 2012, China*. <http://www.cefs.ac.cn/express/images/SSA%20China-2012%20Hassani%20et%20.pdf>
- [9] S. Polukoshko, J. Hofmanis, Use of “Caterpillar” - SSA method Analysis and Forecasting of Industrial and Economic Indicators. *Environment. Technology. Resources Proceedings of the 7th International Scientific and Practical Conference*. Volume II, Rēzeknes Augstskola, Rēzekne, RA Izdevniecība, 241-248, 2009. http://zdb.ru.lv/conferences/2/II_241-248_Polukosko2.pdf
- [10] R.Vautard, P. Yiou, M. Ghil, Singular Spectrum Analysis: A Toolkit for Short, Noisy Chaotic Signals, *Physica D.*, V. 58, 95- 126, 1992.
- [11] D. H. Schoellhamer, Singular spectrum analysis for time series with missing data. *Geophysical Research Letters*, V. 28, no. 16, 3187 – 3190, 2001.
- [12] N. Golyandina, E. Osipov, The "Caterpillar"-SSA method for analysis of time series with missing values, *Journal of Statistical Planning and Inference*, Volume 137, Issue 8, 2642-2653, 2007.
- [13] N. Golyandina, K. Usevich, I. Florinsky, Filtering of Digital Terrain Models by two-dimensional Singular Spectrum Analysis. *International Journal of Ecology & Development*, Vol. 8, No. F07, 81-94, 2007.
- [14] N. E. Golyandina, V.V. Nekrutkin and D. V. Stepanov, *Variants of the “Caterpillar”-SSA method for analysis of multidimensional time series*. <http://www.gistatgroup.com/cat/>
- [15] C. Keppenne, U. Lall, Complex singular spectrum analysis and multivariate adaptive regression splines applied to forecasting the southern oscillation. *Experimental Long-Lead Bulletin*, Volume 1, Number 2 of the On-line version. <http://www.cpc.ncep.noaa.gov/products/predictions/experimental/bulletin/Mar96/article13.html>
- [16] LR Centrālā Statistiskās Pārvalde. *Datubāzes*. <http://www.csb.gov.lv>