ISSN 1691-5402
ISBN 978-9984-44-028-6
Environment. Technology. Resources
Proceedings of the 7th International Scientific and Practical Conference. Volume I1
© Rēzeknes Augstskola, Rēzekne, RA Izdevniecība, 2009

USE OF **"**CATERPILLAR**"** – SSA METHOD FOR ANALYSIS AND **FORECASTING OF INDUSTRIAL AND ECONOMIC INDICATORS**

"CATERPILLAR" – SSA METODES PIELIETOŠANA INDUSTRIĀLU UN EKONOMISKU RĀDĪTĀJU ANALĪZEI UN PROGNOZEI

Svetlana Polukoshko, Janis Hofmanis
Engineering Research Center of Ventspils University College
101 Inzenieru St, Ventspils, LV 3601, Latvia, e-mail: pol.svet@inbox.lv

Abstract. *This paper deals with the Caterpillar»-SSA method, a novel and powerful model-free method of time series analysis and forecasting. Alongside with signal processing this method is successfully used to study the time series in many various areas: in meteorology, hydrology, sociology, economics, traffic analysis, wherever the trend or periodic behavior can present. Examples of application of the Caterpillar"-SSA technique for analysis of one-dimensional time series in Latvian economics are presented in this work. We solve the task of analysis and forecasting of following time series: agricultural crop yield, milk production and purchase, number of road traffic accidents and number of registered road vehicles, electricity consumption. The application of Caterpillar»-SSA approach in geotechnical investigation for processing of data of the static penetration test of soils are offered. This method combines the advantage of many other methods, in particular, Fourier analysis and regressive analysis. At the same time it is noted for simplicity and clearness.*

Keywords: *time series, Singular Spectrum Analysis, sequential algorithm, singular value decomposition, forecast.*

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) [3; 4; 5; 6]. 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 [2; 5; 13; 16; 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, simultaneous extraction of cycles with small and large period, extraction of periodicities with varying amplitudes, finding structure in short time series, detection of structural changes, continuation of the time series [5; 6; 16; 1]. 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 [14; 8]. A method is extensively developing, the variants of the 'Caterpillar'-SSA method for anaslysis of multidimensional time series (MSSA), SSA detection of structural changes are elaborated [7; 5; 11; 14; 15]. The "Caterpillar"-SSA method is useful tool when short and long, one-dimensional and multidimensional, stationary and nonstationary, almost deterministic and noisy time series are to be analyzed.

Mathematical background

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 to reconstruct the time series component on the base of chosen eigentriple numbers.

“Caterpillar” - SSA as a method of analysis performs four steps.

At the first step (called the embedding step) a one-dimensional series $F = (f_1, \dots, f_n)$ is transferred to the L- dimensional series $X_i = (f_i, \dots, f_{i+L-1})^{\mathrm{T}}$, $i = 1, \dots, K$, where $K = N - L + 1$ and L is called by window length. 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: $\mathbf{X} = [X_1 : \dots : X_K]$.

The second step, SVD step, is the singular value decomposition of the trajectory matrix into a sum of rank-one bi-orthogonal elementary matrices and gives the second name of the technique: $\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_L$.

Elementary matrix $\mathbf{X}_i$ is determined by the equality $\mathbf{X}_i = s_i U_i V_i^{\mathrm{T}}$, where s_i (i −th singular value) is the square root of the i −th eigenvalue of the matrix $\mathbf{XX}^{\mathrm{T}}$; U_i and V_i stand for left and right singular vectors of the trajectory matrix. It is assumed that eigenvalues s_i^2 are arranged in the decreasing order of their magnitude. The collection (s_i, U_i, V_i) is called the i −th eigentriple of the matrix $\mathbf{X}$.

The first two steps together are considered as the decomposition stage of ‘Caterpillar’-SSA.

The next two steps form the reconstruction stage. The grouping step corresponds to splitting the elementary matrices into several groups and summing the matrices within each group. The result of the step is a representation of the trajectory matrix as a sum of several resultant matrices: $\mathbf{X} = \mathbf{Y}_1 + \ldots + \mathbf{Y}_m$.

The last step (diagonal averaging) transfers each resultant matrix into a time series, which is an additive component of the initial series F. If y_{ij} stands for an element of a matrix $\mathbf{Y}$, then the k −th term of the resulting series is obtained by averaging of y_{ij} over all i, j such that $i + j = k - 1$. Diagonal averaging is a linear operation and maps the trajectory matrix of the initial series into the initial series itself. In this way we obtain a decomposition of the initial series into several additive components. The result is the expansion of series into simple components $F = F_1 + \dots, F_m$. For a properly made ‘Caterpillar’-SSA expansion, a component F_i in this equation may be identified as a trend of the original series F, an oscillatory series (for example, seasonality) or noise.

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 [4], [6].

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 choosen; 2) to extract oscillations, oscillating eigenfunctionsare choosen; 3) to extract periodic (harmonic) components with period greater than 2 a pair of eigentriples with the same period should be found and choosen; 4) to aproximate the time series the leading eigentriples are choosen.

Examples of applications of SSA technique for analysis of time series in Latvian economy and agriculture

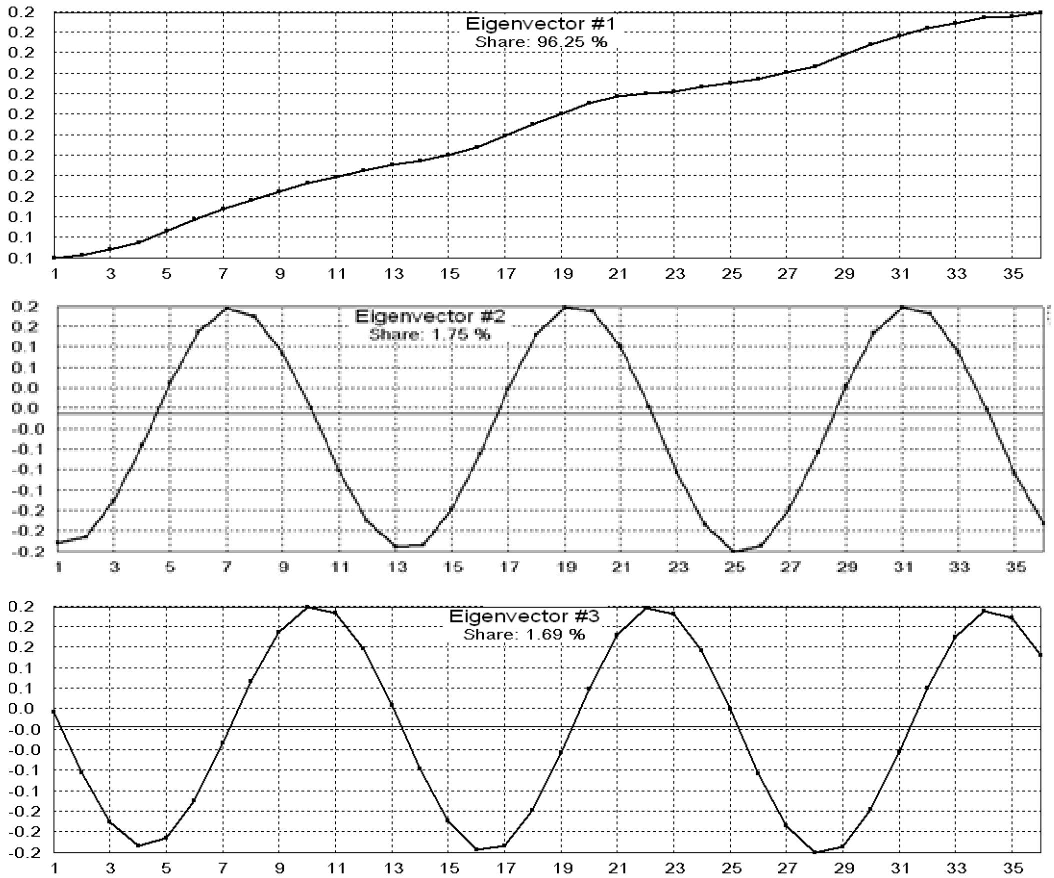
We used Caterpillar'-SSA method to analysis and forecast of industry indicators, available in Latvian statistical databases [7]. The algorithm was applied to diferent data sets and extensively studied; available time series were processed in the interactive mode. Alongside with CatMV 1.1 software, created in Department of Mathematics of St. Petersburg University for analysis of time series which may contain missing values, we realized the SSA forcasting algorithm for class of series govern by the LRF (linear recurrent formulae) on MATLAB program and used it in our work. Some examples are presented below.

In Fig. 1 the results of analysis of milk production and purchase in Latvia are given, the data covers the time from 2002 till 2008 years. Monthly time series of 84 month length was analysed and forecasting was executed for 12 months.

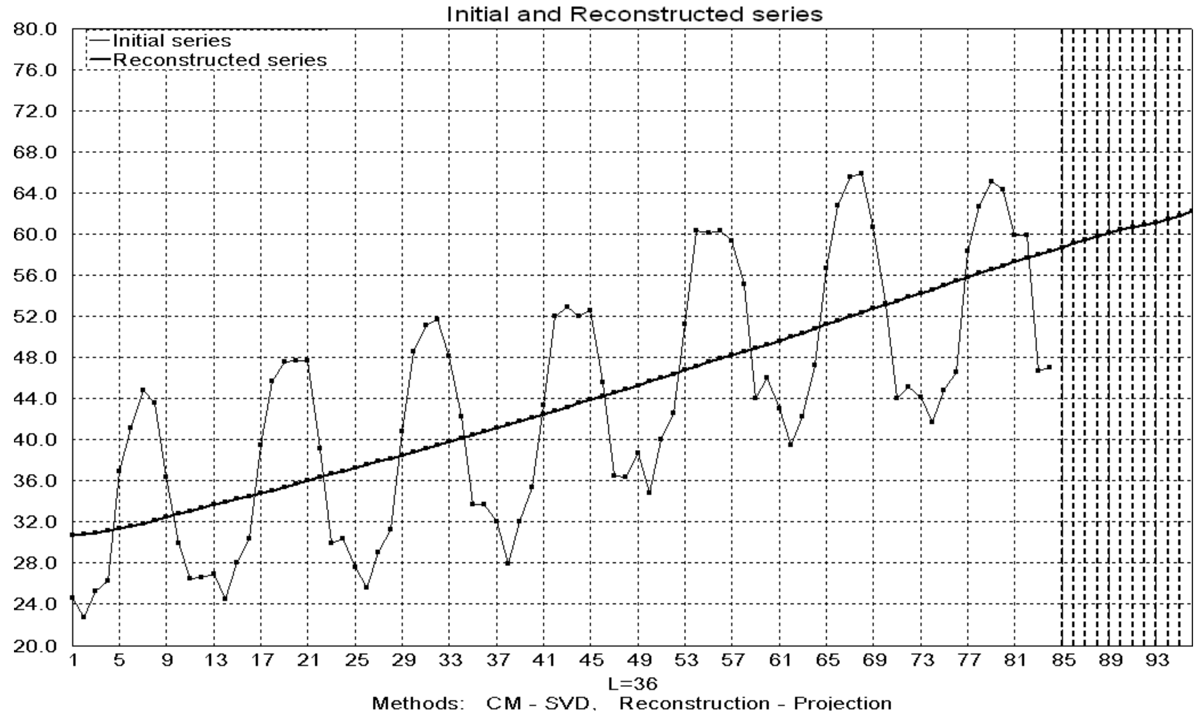
In Fig. 2 the analysis of electricity consumption in Latvia is presented, time series of 36 months length was analysed and prognosis was executed on 12 months.

Analysis and forecast of agricultural crop productivity was made for winter and spring cereals. The Initial time series length was taken of 51 years (from 1957 till 2007) and forecasting was executed for 3 years. In Fig. 3 the analysis of winter cereals crops yield (100 kg per ha) is presented, in Fig.4 - spring cereals crops yield. Length of window L=11, series are reconstructed with 3 eigentriples.

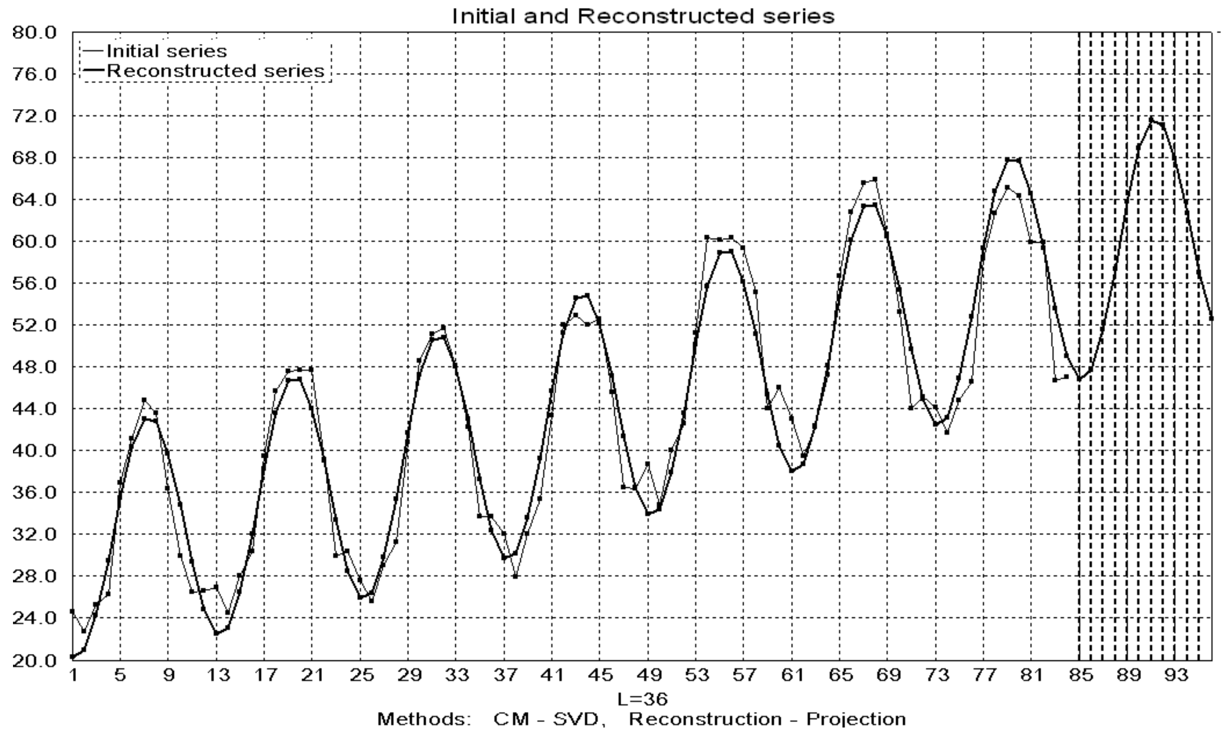
In Fig. 5 and 6 the results of number of road traffic accidents and number of registered road vehicles (lorries, passenger cars, buses and motorcycles) analysis are presented, time series length is 27 quarters, forecasting period length is 1 year (4 quarters). It is obviously that nomber of road accidents does not depend of number of registered vehicles.



a) the first three eigenvectors

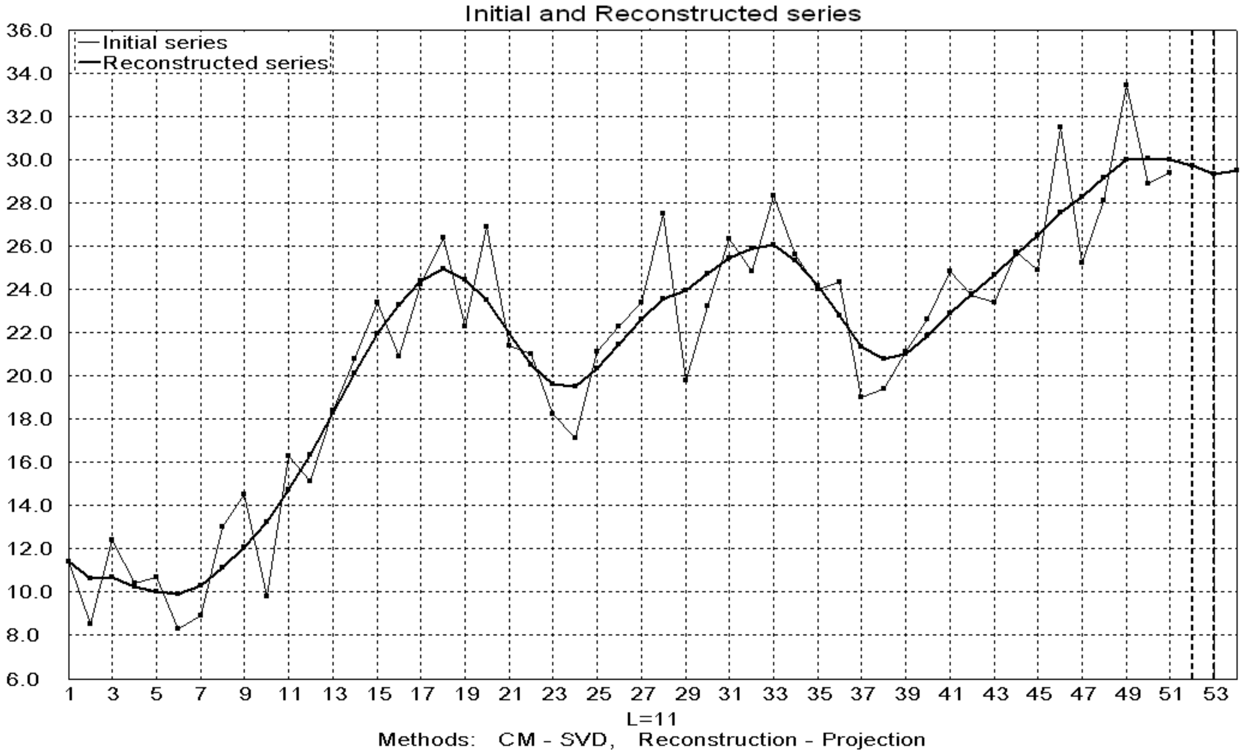


b) initial series and its trend



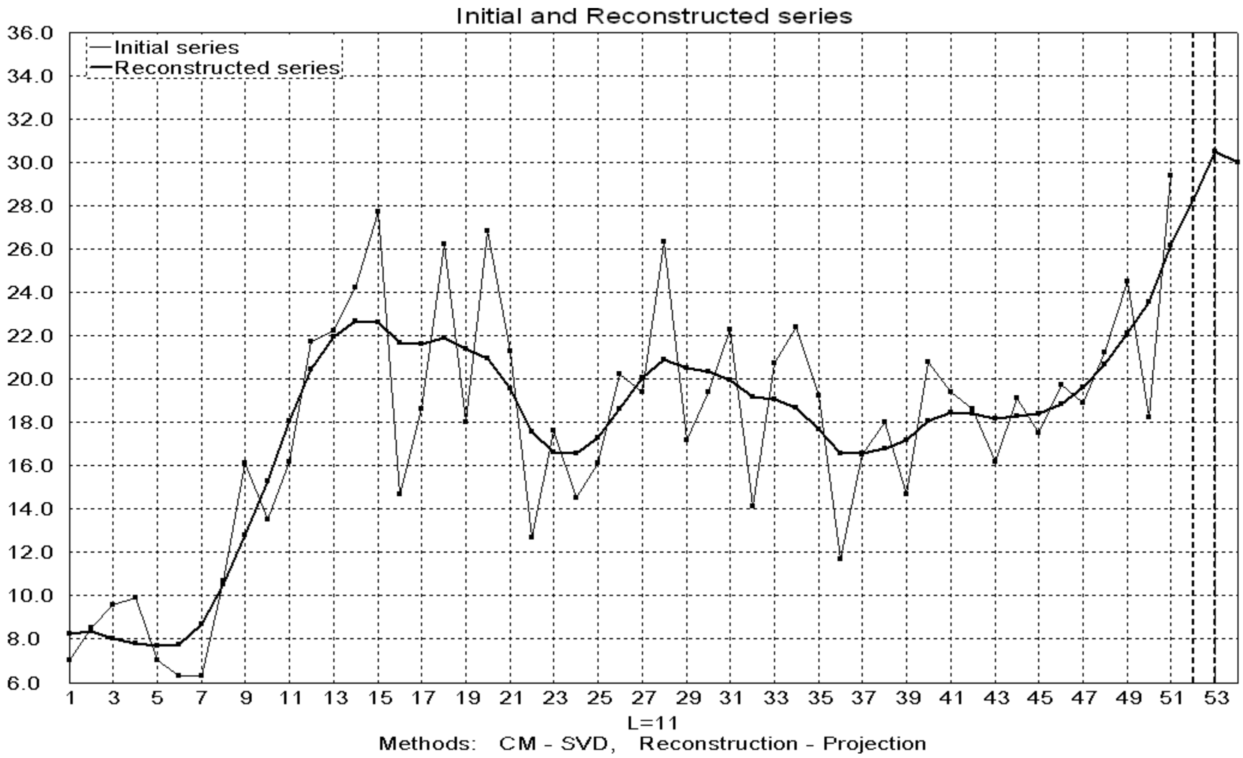
c) initial and reconstructed (with three eigentriles) time series and forecasting for 12 months

Fig.1. **Milk production and purchase - thousand of tones per month, from 2002 year**



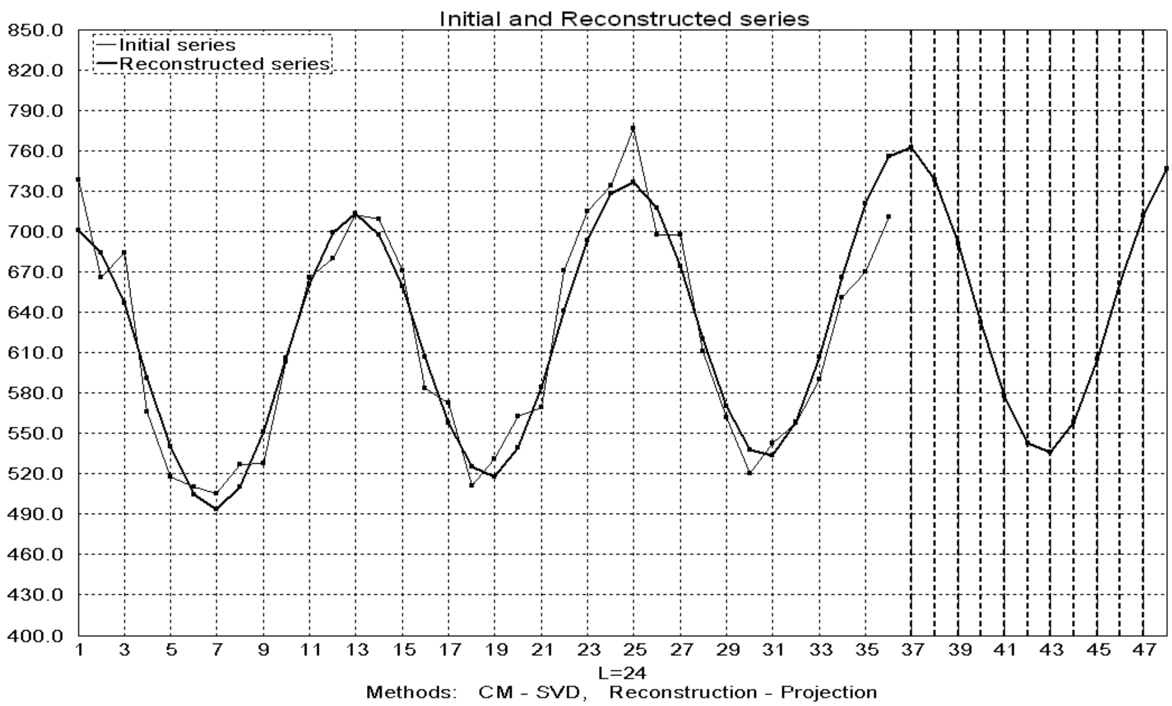
initial and reconstructed (with three eigentriles) time series and forecasting on 4 years

Fig.2. **Yield of winter cereals crops - hundreds kilograms per hectare, from 1957 till 2007 year**



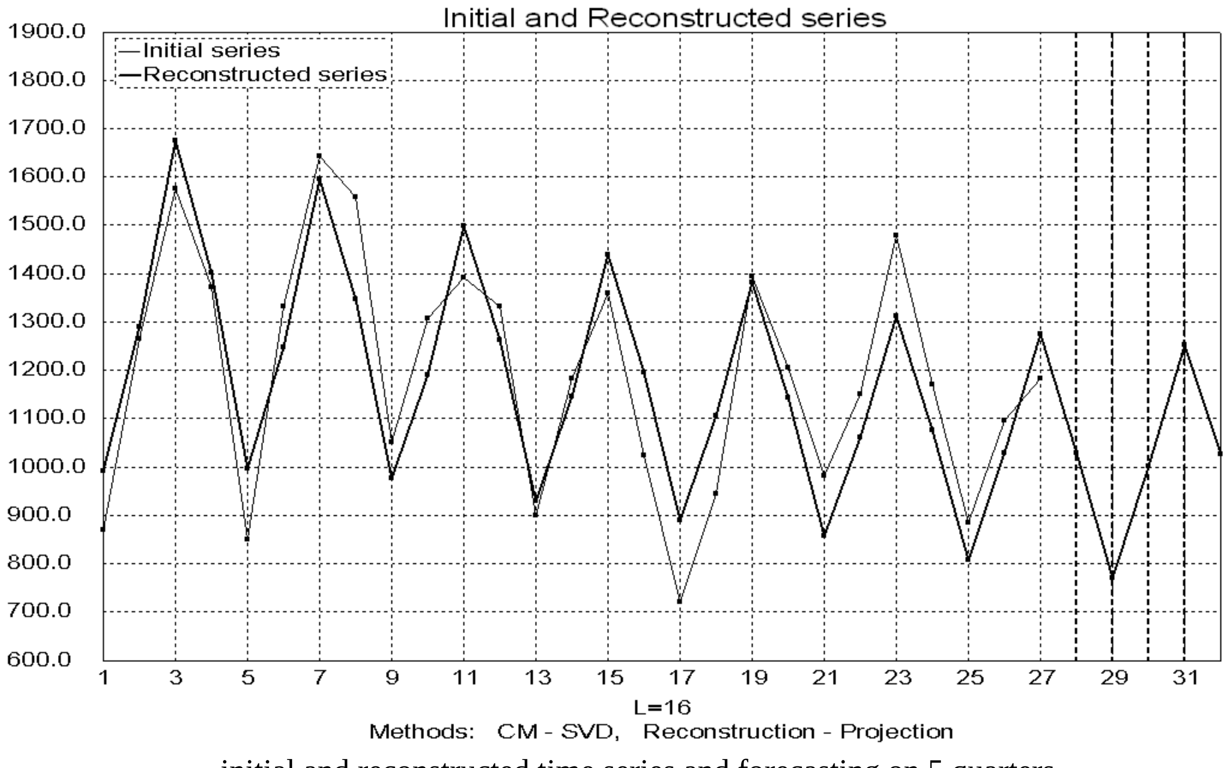
initial and reconstructed (with three eigentriles) time series and forecasting on four years

Fig.3. **Yield of spring cereals crops - hundreds kilograms per hectare, from 1957 till 2007 year**



initial and reconstructed (3 eigentriles) time series and forecasting for 12 monthes

Fig. 4. **Electricity consumption - mln. kWh per month, from year 2006**



the initial and reconstructed time series and forecasting on 5 quarters

Fig. 5. **Road traffic accidents - number per quater, from 1.01. 2002**

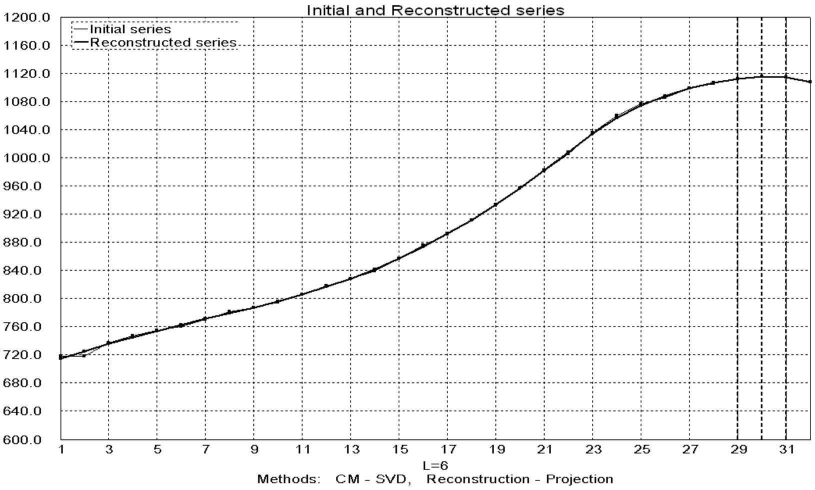


Fig. 6. **Registreted road vehicles - number per quater, from 1.01. 2002**

Example of using of «Caterpillar»-SSA method in geotechnical investigations

Geotechnical investigation is executed to determine the ground physical and mechanical characteristics. One of the methods of soils investigation is the static penetration test which is used for the determination of the boundary lines between the soil layers of different soil conditions, constitution, heterogeneity, and also as an indirect method for estimation of the ground physical and mechanical characteristics. We proposed to use «Caterpillar»-SSA approach to the processing of results of static probing and dividing the ground into engineer-geological elements. The results of the static penetration test are recorded by the field computer point resistance q_c (MPa) of penetration into the soil of standard metal cone. Indicated value are written down through equal intervals a depth (from 2 till 20 cm), the depth of penetration may reache 30 m. In such series there is generally no periodic components and its processing with «Caterpillar»-SSA method, in which trend and random component is selected, simplifies finding of identical layers. Point resistance to standard cone penetration depenging on the depth and the results of its SSA processing are given in Fig.7, where the time series component reconstructed from the leading eigentriple and the initial time series, L=10.

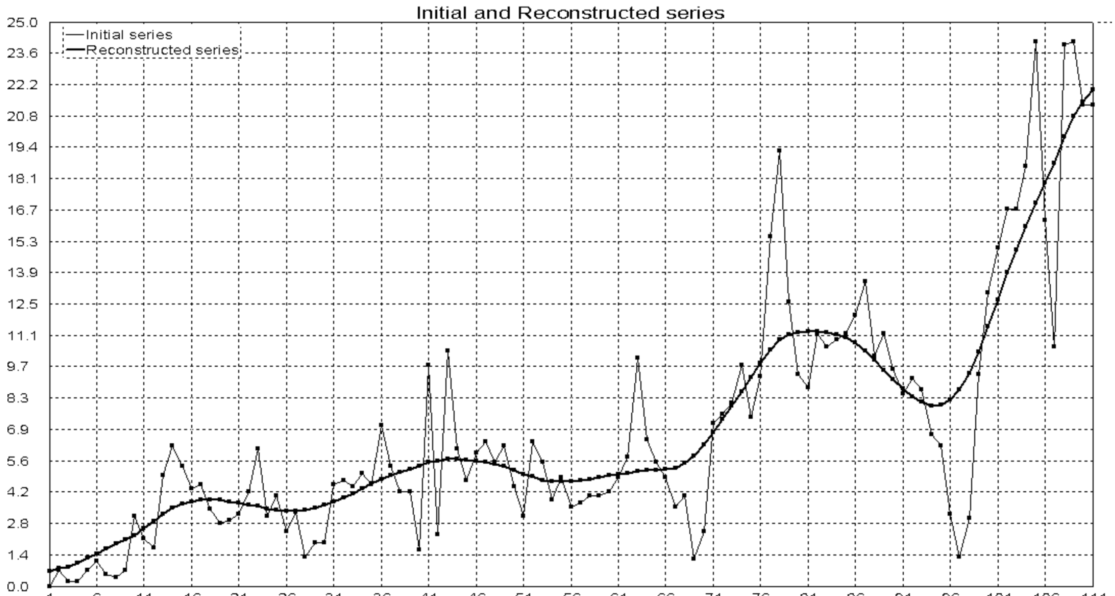


Fig. 7. **Point resistance to standard cone penetration qc (MPa) depenging on depth (m)**

Conclusion

Several example of application of the Caterpillar"-SSA technique for analysis of one-dimensional time series in Latvian economics are presented in this work. We solve the task of analysis and forecasting of following time series: milk production and purchase, electricity consumption, yield of winter and spring cereals crops, number of road traffic accidents and number of registered road vehicles.

The results allow improving quality of prediction of industrial and agricultural indictors characterizing complex non-stationary process, to increase the efficiency of strategic decisions making in the conditions of existing economic situation, to minimize the losses of capital.

In geotechnical investigation the using of SSA method make easier the solution of the problem of division of soil column on geotechnical units.

Future investigation is to create the MATLAB program for analysis of multidimensional time series and for solving the forecasting problem with the confidence limits.

References

1. Alexandrov Th., Golyandina N. Automatic extraction and forecast of time series cyclic components within the framework of SSA. In: Proceedings of the 5th St.Petersburg Workshop on Simulation, June 26-July 2, 2005, St. Petersburg State University, St. Petersburg, pp. 45-50.
2. Antoniou V.V. 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, 24-28 June 2002, Moscow, Russia, Book of abstracts, p. 176.
3. Cadzow J.A. Signal Enhancement - A Composite Property Mapping Algorithm. IEEE Transactions on Acoustics, Speech and Signal Processing, 1988. V. 36, pp. 49-62.
4. Danilov D., Zhigljavsky A. eds. Principal Components of Time Series: "Caterpillar" Method. St. Petersburg University, 1997. http://www.gistatgroup.com/gus/
5. Ghil M. et al.: Climatic time seriers analysis: Reviews of Geophysics, 40, 1 / Month 2002 pages 1-1–1-41
6. Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques. London: Chapman and Hall, 2001.
7. Golyandina N., Usevich K., Florinsky I. Filtering of Digital Terrain Models by two-dimensional Singular Spectrum Analysis. International Journal of Ecology & Development, 2007. Vol. 8, No. F07, P.81-94.
8. Golyandina N.E., Nekrutkin V.V., and Stepanov D.V. Variants of the "Caterpillar"-SSA method for analysis of multidimensional time series. http://www.gistatgroup.com/cat/
9. Golyandina N., Osipov E. The "Caterpillar"- SSA method for analysis of time series with missing values Journal of Statistical Planning and Inference, Volume 137, Issue 8, 1 August 2007. pp. 2642-2653.
10. Jolliffe I.T. Principal Component Analysis. Springer-Verlag, 1986.
11. Keppenne C., Lall U. Complex singular spectrum analysis and multivariate adaptive regression splines applied to forecasting the southern oscillation // Exp.Long-Lead Forcst. Bull. 1996. http://www.cpc.ncep.noaa.gov/products/predictions/experimental/bulletin/Mar96/article13.html
12. LR Centrālā statistiskas pārvalde. Datubāzes. http://www.csb.gov.lv
13. Preizendorfer R.W. Principal Component Analysis in Meteorology and Oceanography. Elsevier, 1988.
14. Schoellhamer D.H. Singular spectrum analysis for time series with missing data. Geophysical Research Letters, V. 28, no. 16, 2001. pp. 3187-3190.
15. Stepanov D., Golyandina N. SSA-based approaches to analysis and forecast of multidimensional time series. In: Proceedings of the 5th St.Petersburg Workshop on Simulation, June 26-July 2, 2005, St. Petersburg State University, St. Petersburg, pp. 293-298.
16. Vautard R., Yiou P., Ghil M. Singular Spectrum Analysis: A Toolkit for Short, Noisy Chaotic Signals, Physica D. 1992. V. 58, pp. 95-126.