UDC 519.2+551.510.534

APPLICATION OF SINGULAR SPECTRUM ANALYSIS FOR EXTRACTING WEAKLY DEFINED TRENDS

V.B. Kashkin, T.V. Rubleva

Siberian Federal University, Krasnoyarsk
E-mail: kafedra@nifti.krasnoyrsk.ru

The peculiarities of new method of analysis of time series – singular spectrum analysis have been considered, advantages of the method and problems connected with it application have been discussed. To increase the accuracy of trend extraction and elimination of edge effects the prediction of the series in both ends was suggested to be used. The singular spectrum analysis was applied to time series of satellite data of total ozone in latitude rings from 40 to 60° in Northern and Southern hemispheres. The trends were found. It was shown that for 1998–2005 in this ring the total ozone decreases by 0,38 %±0,01 % per year in Northern hemisphere and by 0,10 %±0,01 % per year in Southern hemisphere. It is no wonder because the most part of ozone damaging substances are produced in middle latitudes of Northern hemisphere.

Studying time series the spectral Fourier analysis or wavelet analysis are usually used. The device of singular spectrum analysis (SSA «Gusenitsa») developed and justified at the end of XX c. by the officers of St. Petersburg State University [1, 2] is applied in this paper. The method is based on the analysis of principle components and allows studying stationary and nonstationary time series. A number of useful theorems was proved, computational approaches for SSA implementation were suggested, software was developed. The connection between classical methods of analyzing stationary time series and the method of principle components was examined earlier in [3].

SSA is not widely applied yet, in particular, in technical application. A number of examples of using SSA «Gusenitsa» is given in [4]. In [5, 6] this method is used for analysis and prediction of data on solar activity and data on irregularity in speed of Earth rotation.

Method «Gusenitsa» is based on conversion from unvariate time series with length n with even pitch $(x_1,x_2,x_3,...,x_n)$ to multivariate series constructed of initial unvariate one. «Gusenitsa» allows extracting the components of time series, in particular, linear and nonlinear trend, components with known and unknown before period, smoothing initial data, predicting series, filling in gaps.

Matrix **X** is formed of initial series

$$\mathrm{X}=\begin{pmatrix} x_1 & x_2 & x_3.....x_k.....x_{m+1} \\ x_2 & x_3 & x_4.....x_{k+1}....x_{m+2} \\ x_3 & x_4 & x_5.....x_{k+2}....x_{m+3} \\ ..... & ..... & .... \\ x_k & x_{k+1} & x_{k+2}...x_{2k-1}...x_n \end{pmatrix}.$$

Here $m<n$ is the length of caterpillar, elements $(x_k,x_{k+1},...,x_n)$ are the last line with number $k=n+m-1$ and $x_{ij}=x_{i+j-1}$. This matrix may be considered as m-dimensional sampling of volume k or m-dimensional time series. In some applications a mean values μ_i may be computed by columns then matrix **X** may be centered. Then matrix $\mathrm{R}=(1/k)\mathrm{X}\cdot\mathrm{X}^T$ is computed. If matrix **X** is centered then **R** is sampling correlation matrix:

$$R=\begin{pmatrix} \sigma_1^2 & R_{12} & ...R_{1m} \\ R_{21} & \sigma_2^2 & ...R_{2m} \\ ...... & ...... & .... \\ R_{m1} & R_{m2} & ...\sigma_m^2 \end{pmatrix}.$$

Expressions

$$R_{ij}=\frac{1}{k}\sum_{i=1}^{k}(x_{i+l-1}-\mu_i)\cdot(x_{j+l-1}-\mu_j).$$

serve as elements of matrix **R**.

Proper numbers and proper vectors of matrix **R** are computed by common technique of principal component analysis [3] that is its singular decomposition $\sum_{i=1}^{m}\lambda_i$=**P**, Λ is carried out, is the diagonal matrix of descending ordered proper numbers of matrix **R**:

$$\Lambda=\begin{pmatrix} \lambda_1 & 0 & 0....0 \\ 0 & \lambda_2 & 0....0 \\ ... & .... & .... \\ 0 & 0 & 0...\lambda_m \end{pmatrix},$$

$\sum_{i=1}^{m}\lambda_i$=**P**, **P** is the orthogonal matrix of proper vectors of matrix **R**:

$$P=\begin{pmatrix} p_{11} & p_{21} & ....p_{m1} \\ p_{12} & p_{22} & ....p_{m2} \\ ..... & ...... & .... \\ p_{1m} & p_{2m} & ....p_{mm} \end{pmatrix}.$$

Matrix **P** is the solution of equation$[\mathrm{R}-\lambda I]\mathrm{P}=0$, I is the identity matrix. Matrix **P** fulfills the condition $\mathrm{P}^{-1}=\mathrm{P}^T$ that means retention of series «mean power».

In the method «Gusenitsa» columns of matrix **P** play a part of transition functions of filters adjusted to components of initial process. Thus, filters are generated by the studied process itself; «Gusenitsa» adjusts itself to those spectral components which exist in this process. Series spectral components are extracted by linear conversion of initial process by discrete convolution operator:

$$y_j[l]=\sum_{q=1}^{m}x_{lq}p_{jq}=\sum_{q=1}^{m}x_{l+q-1}p_{jq}.$$

Selection of several principle components is similar to parallel connection of several filters. Width of filter pass band is determined by proper vector type and length of averaging interval («caterpillar» length m). Width of pass band is inversely proportional to m. Peak value m equals to a half of length of series n in this case at even n matrix X is square. At small m up to m=2 series is smoothed. Type of proper vectors and principle components obtained due to conversion of matrix **R** gives information on structure of studied process and properties of its summands. In particular, among principle components it is possible to observe and select interactively those ones that refer to trend (slowly varying), periodical, noise.

Comparison of method «Gusenitsa» and Fourier analysis is of interest. If a series consists of a set of strictly harmonic components then «Gusenitsa» decomposes to Fourier series extracting those very components. «Gusenitsa» allows extracting spectral components taking into account changes of amplitude and phase of quasiperiodic signal that is impossible using common Fourier series. «Gusenitsa» extracts also pairs of quasiperiodic constituents of series principle components which have length $m \leq n/2$. These components are phase-shifted by $\pi/2$ that is orthogonal. Arc tangent of these components relation gives the dependence of total phase Φ of quasiperiodic component on time but this series length is shorter than initial series length n and equals to $m \leq n/2$. Though phase Φ is determined with accuracy up to 2π it is not difficult to eliminate phase jumps of 2π in time series of total phase. Sometimes total phase may be approximated by linear function: $\Phi(t_i)=A \cdot t_i+B$, $i=1...m$. In this case coefficient $A=\omega$ is the frequency that is phase derivative. More compound phase approximation is possible, for example, $\Phi(t_i)=A \cdot t_i+C \cdot t_i^2+B$. Phase derivative gives trend of frequency: $\omega(t_i)=A+C \cdot t_i$..

Using SSA it is necessary to improve spectral resolution which depends on series length n. Authors of this paper increased n extending series in both ends using prediction made by the «Gusenitsa». «Inverse decimation» may be carried out that is series is extended inserting intermediate interpolated values between initial values using properties of «Gusenitsa» to recover missing data. These procedures are compound while discrete Fourier transform allows improving spectral resolution by an ordinary zero addition [7]. Let us note as well that in comparison with rapid discrete Fourier transform the «Gusenitsa» is the slow method in principle which does not allow in modern interpretation processing time series in real time as it is necessary to work with matrix of large dimension.

Evidently, one of complex tasks of time series analysis is accurate extraction of subdelirium trend against the background of noise and periodic signal of high intensity. Computing experiment with a series of length 3000 pitches (left upper corner in Fig. 1) was carried out where trend is linear and periodic signal contains a component of sinusoidal form. In Fig. 1 the original trend is shown by Figure 1, the result of extracting trend by a polynomial of the first degree by the least squares method (LSM) (2) and the result of extracting trend by «Gusenitsa» (3). Use of Hanna window and MNAT-wavelet showed significantly worth results. Edge effects are shown by arrows when using. Although accuracy of extracting trend by «Gusenitsa» is significantly higher than by the method of LSM it should be increased eliminating edge effects.

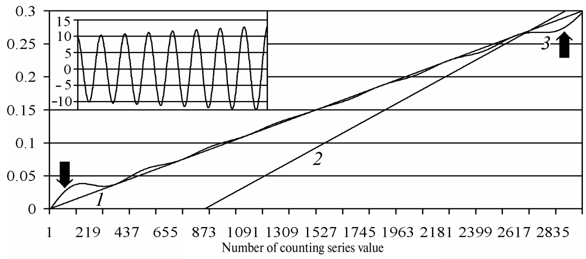

Fig. 1. *Extraction of trend (1) by the LSM (2) by the «Gusenitsa» (3)*

The method mentioned before was used for this purpose – in this case series prolongation by 1000 pitches forward using prediction by «Gusenitsa». Prediction allowed decreasing significantly the edge effect – shifting to the right inexact estimate region of a trend out of investigated series with the length of 3000 pitches (Fig. 2). The surge owing to the influence of edge effect is shown by an arrow in Fig. 2.

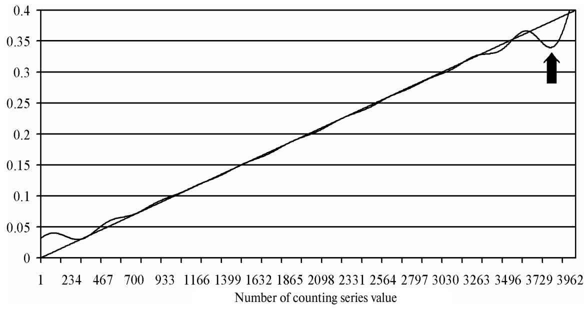

Fig. 2. *Displacement of trend inexact estimate region (shown by an arrow) to the right out of initial series with the length of 3000 pitches*

Let us consider processing of time series of total ozone (TO) in Earth atmosphere as an example. Ozone layer performs important ecological task – protects every living thing on Earth from destructive ultraviolet solar emission. Gradual global exhaustion of ozone layer during decades is observed [8]. Extracting trends of time series of TO it is possible to determine the rate of exhaustion.

TO that is ozone layer thickness reduced to the temperature 273 K and normal atmospheric pressure is determined by surface and satellite methods. Artificial Earth satellites estimate TO recording solar radiation scattered «backward» in ultraviolet spectrum. In this paper daily global data on TO (series with length n=2920 day) obtained by spacecraft EP/TOMS (USA) for the period from the beginning of 1998 to the end of 2005 and placed on site NASA are used [9].

TO in the region of circumpolar vortices in latitudes in the circle 40...60° in Northern and Southern hemispheres is studied. Circumpolar vertex that is large cyclone where TO is always rather high surrounds Antarctic ozone layer in Southern hemisphere. Selection of latitude interval is stipulated by the fact that in latitudes higher than 40° TO changes considerably that exceeds changes in tropics and subtropics. In latitudes more 60° these changes are more considerable but continuous series of TO for these latitudes are absent by satellite data as satellite equipment of optical range can not estimate TO in winter months on polar atmosphere regions unlit by the Sun.

Dependence of average TO in a circle of latitudes from 40 to 60° on time for Northern and Southern hemispheres in Dobson units (D.u.) is given in Fig. 3.

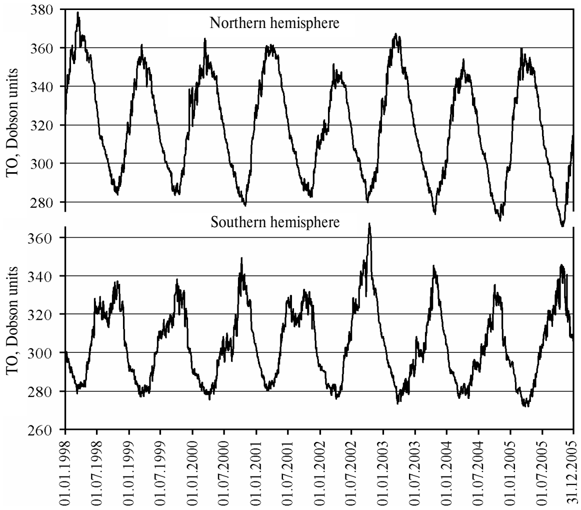

Fig. 3. *Diagrams of time series of average TO in circle of latitudes from to for Northern and Southern hemispheres*

In the period when in Northern hemisphere TO value is maximal in Southern hemisphere minimum is observed. TO maximums are timed to spring, minimums – to autumn of each hemisphere. Let us note as well that in the period of maximum of average TO in Southern hemisphere the Antarctic ozone hole has the largest area. Each time series presented in Fig. 3 contains trend, periodical and random components.

Linear trends of series presented in Fig. 3 were originally found by the LSM in the form $x_i=a\cdot t_i+b,i$, $i=1...n$. Parameter a carries important information and characterizes the rate of changing average TO in the circle. For Northern hemisphere estimation $a=-0,00726$ D.u./day, standard deviation of estimation $\sigma_a=0,00056$ D.u./day. Minus sign means TO decrease. TO decrease amounts to 2,6 D.u. or 0,83 % per year relative to mean value.

For Southern hemisphere LSM-estimation $a=-0,00044$ D.u./day, standard deviation of estimation $\sigma_a=0,00042$ D.u./day that is estimation a and its standard deviation are almost equal, therefore, estimation a can not be considered as significant one.

Then analysis of TO series was carried out by the method of SSA «Gusenitsa» using proper software [4]. To increase spectral resolution and eliminate edge effects the series were prolonged in both ends by 1000 days using prediction. It was found by simulation that in this case the error due to edge effects and noise at estimation a amounted to $\Delta a\leq 0,0001$ D.u./day. Application of singular spectrum analysis allowed increasing considerably reliability of estimating TO trend, degradation rate of ozone layer and trend type should not be specified beforehand. The obtained trends of average TO for Northern and Southern hemispheres in circle of latitudes from 40 to 60° are given in Fig. 4.

It follows from Fig. 4 that TO trend in Northern hemisphere is well described by linear dependence; trend in Southern hemisphere at some approximation may be approximated by a straight line as well. For Northern hemisphere estimation of parameter $a=-0,00334$ D.u./day, estimation standard deviation $\sigma_a=2,3\cdot 10^{-6}$ D.u./day, for Southern hemisphere estimation $a=-0,00811$ D.u./day, estimation standard deviation $\sigma_a=5,2\cdot 10^{-6}$ D.u./day. In both cases $\sigma_a<<\Delta a$, therefore, error of estimation a should be considered as equal to Δa.

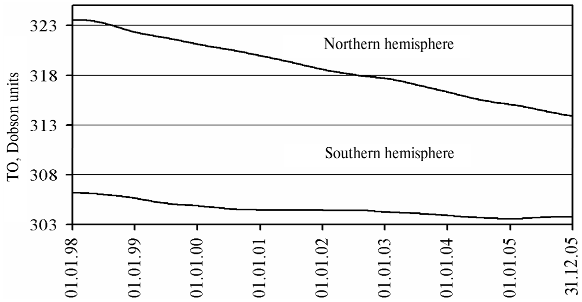

Fig. 4. *Trend of average TO for Northern and Southern hemispheres in circle of latitudes from 40 to 60°*

Thus, total ozone in Northern hemisphere has decreased for 1998–2005 on average by 1,222 D.u./year that is by 0,38 %±0,01 % relative to a mean value. In Southern hemisphere there is a tendency to decrease of TO by 0,296 D.u./year that is by relative to a mean value. Application of singular spectrum analysis allowed increasing significantly accuracy of trend extraction.

Work is performed at support of RFBR, grant № 07-01-00326

REFERENCES

1. Principle components of time series: method «Gusenitsa» / Ed. by D.L. Danilov and A.A. Zhiglyavskiy. – St. Petersburg: Presskom, 1997. – 308 p.
2. Golyandina N.E. Method «Gusenitsa» SSA: time series analysis. – St. Petersburg: St. Petersburg University, 2004. – 74 p.
3. Brillindger L. Time series. Data processing and theory. – Moscow: Mir, 1980. – 536 p.
4. http://www.gistatgroup.com.
5. Loskutov A., Istomin I.A., Kuzanyan K.M., Kotlyarov O.L. Testing and Forecasting the Time Series of the Solar Activity by Singular Spectrum Analysis // Nonlinear Phenomena in Complex Systems. – 2001. – V. 4. – № 1. – P. 47–51.
6. Kashkin V.B., Baskova A.A. Studying irregularity in speed of Earth rotation by singular spectrum analysis // Vestnik Krasnoyarskogo Gosudarstvennogo Universiteta. Fis.-mat. nauki. – 2006. – № 7. – P. 53–60.
7. Marpl-Jr. S.L. Digital spectrum analysis and its applications. – Moscow: Mir, 1990. – 586 p.
8. Aleksandrov A.L., Israel Yu.A., Karol I.L., Khrgian A.Kh. Earth's ozone sheet and its changes. – St. Petersburg: Gidrometizdat, 1992. – 287 p.
9. http://jwocky.toms.gsfc.nasa.gov.

Received on 05.07.2007