

## A Study of the Regularities in Solar Magnetic Activity by Singular Spectral Analysis

A. Yu. Loskutov<sup>1,\*</sup>, I. A. Istomin<sup>1</sup>, O. L. Kotlyarov<sup>1</sup>, and K. M. Kuzanyan<sup>2</sup>

<sup>1</sup>*Moscow State University,  
Vorob'evy gory, Moscow, 119899 Russia*

<sup>2</sup>*Institute of Terrestrial Magnetism, Ionosphere,  
and Radiowave Propagation,  
Russian Academy of Sciences,  
Troitsk, 142092 Russia*

Received March 12, 2001

**Abstract**—The method of singular spectral analysis (SSA) is described and used to analyze the series of Wolf numbers that characterizes solar activity from 1748 until 1996. Since this method is relatively new, we detail its algorithm as applied to the data under study. We examine the advantages and disadvantages of the SSA method and the conditions for its applicability to an analysis of the solar-activity data. Certain regularities have been found in the dynamics of this series. Both short and long (80–100-year) periodicities have been revealed in the sunspot dynamics. We predict the solar activity until 2014.  
© 2001 MAIK “Nauka/Interperiodica”

Key words: *Sun, nonlinear dynamics, solar cyclicity*

### INTRODUCTION

It has long been noticed that the solar activity is related to the number of sunspots visible on the solar disk. The sunspot number varies greatly within an 11-year interval called the solar cycle. The accompanying change in the solar magnetic-field structure indirectly affects the Earth's climate and has a probable relationship to natural disasters. Since the solar magnetic activity is significant, its analysis is of great practical interest.

Various tracers are used to describe the dynamics of solar magnetic activity, of which the Wolf number (relative sunspot number) is most convenient. The dynamics of this parameter is quasi-periodic in pattern. However, accurate predictions are difficult to make, because simple models of the process disregard many important factors of the solar magnetic activity. During the last 250 years, the duration of the solar cycle has varied by no more than 20%, while its amplitude has varied by more than a factor of 10. Even sophisticated models do not give a detailed description of these variations.

Recently, many methods for predicting and reconstructing the dynamics of the series of Wolf numbers have been proposed (see Schatten 1997, Nagovitsyn 1997, Wilson *et al.* 1998, Hoyt and Schatten

1998, Hathaway *et al.* 1999, and references therein). Since they all have various drawbacks, predicting the sunspot dynamics from the available observational data alone, without constructing a model of the phenomenon, has become very promising. Here, an analysis of time series by the methods of nonlinear dynamics (see Afraïmovich and Reïman 1989, Casdagli 1989, Loskutov and Mikhaïlov 1990, Ruelle 1990, Sauer *et al.* 1991, Malinetskiï and Potapov 2000, and references therein) can give a significant contribution. In this case, however, there are also many difficulties that stem from the fact that the series of Wolf numbers is apparently not a strictly deterministic system and has no well-defined dimensionality (Lawrence *et al.* 1993, 1995); besides, it is relatively short.

As a new method for analyzing and predicting the dynamics of the time series formed by Wolf numbers, we propose to use a singular spectral analysis (SSA). As we show below, it provides highly reliable predictions of the amplitude of the 11-year solar cycle and is suitable for revealing longer cycles. It can also be used to study regularities in series of other astrophysical indices. Since this method is relatively new and covered little in the literature, we detail its algorithm as applied to the formulated problem.

### SINGULAR SPECTRAL ANALYSIS

The SSA method (Broomhead and King 1986a, 1986b; Broomhead and Jones 1989; Vautard *et al.*

\*E-mail: loskutov@moldyn.phys.msu.su

1992; Danilov and Zhiglyavskii 1997) used here allows the following:

–to distinguish the components of a time series obtained from a sequence of values of a quantity taken at equal time intervals;

–to find periodicities in a series that are not known in advance;

–to smooth initial data on the basis of selected components;

–to best separate a component with a period known in advance; and

–to predict the subsequent behavior of the observed dependence.

The SSA method is efficient enough to successfully compete with numerous smoothing techniques (Danilov and Zhiglyavskii 1997, Percival and Walden 1993, Theiler *et al.* 1992, Kaplan and Glass 1992). Moreover, SSA-based predictions in many cases yield more reliable results than do other known algorithms (see Casdagli 1989, Danilov and Zhiglyavskii 1997, Deppish *et al.* 1991, Murray 1993, Cao *et al.* 1995, Keppenne and Ghil 1995, and references therein).

The SSA method is based on the passage from an analysis of the initial linear series  $(x_i)_{i=1}^N$  to an analysis of a multidimensional series composed of its sets, which, apart from  $x_i$  itself, contain a certain number of  $x_{i-j}$ ,  $j = 1, \dots, \tau$ , at preceding times.

Let us briefly describe the main stages of SSA application to the specific series  $(x_i)_{i=1}^N$ .

(1) At the first stage, the one-dimensional series is transformed to a multidimensional one. For this transformation, it is necessary to take some number of delays  $\tau \leq [(N+1)/2]$ , where  $[\cdot]$  denotes the integer part of a number, and to represent the initial  $\tau$  values of the sequence as the first column of some matrix  $X$ . The sequence values from  $x_2$  to  $x_{\tau+1}$  are chosen for the second column of this matrix, and so on. The last  $\tau$  elements of the sequence  $x_n, \dots, x_N$  correspond to the last column with number  $n = N - \tau + 1$ . Thus, the transformed series takes a matrix form

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_\tau & \dots & x_n \\ x_2 & x_3 & x_4 & \dots & x_{\tau+1} & \dots & x_{n+1} \\ x_3 & x_4 & x_5 & \dots & x_{\tau+2} & \dots & x_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_\tau & x_{\tau+1} & x_{\tau+2} & \dots & x_{2\tau-1} & \dots & x_N \end{pmatrix}.$$

The constructed matrix  $X$  is rectangular, but in the limiting case, i.e., for  $\tau = (N+1)/2$  and odd  $N$ , it degenerates to a square matrix.

(2) Next, the corresponding covariance matrix is constructed for matrix  $X$

$$C = \frac{1}{n} X X^T.$$

(3) Now, the eigenvalues and eigenvectors of matrix  $C$  must be determined. This requires its decomposition into eigenvectors  $C = V \Lambda V^T$ , where we introduced

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_\tau \end{pmatrix}$$

the diagonal matrix of eigenvalues, and

$$V = (V^1, V^2, \dots, V^\tau) = \begin{pmatrix} v_1^1 & v_1^2 & \dots & v_1^\tau \\ v_2^1 & v_2^2 & \dots & v_2^\tau \\ \vdots & \vdots & \ddots & \vdots \\ v_\tau^1 & v_\tau^2 & \dots & v_\tau^\tau \end{pmatrix}$$

the orthogonal matrix of eigenvectors for matrix  $C$ . Clearly,  $\Lambda = V^T C V$ ,  $\det C = \prod_{i=1}^\tau \lambda_i$ , and  $\sum_{i=1}^\tau \lambda_i = \tau$  (the latter equality holds only for the prenormalized rows of matrix  $X$ ).

(4) The matrix of eigenvectors  $V$  is commonly represented as a transition matrix to the principal components  $Y = V^T X = (Y_1, Y_2, \dots, Y_\tau)$  of the initial series, where  $Y_i$ ,  $i = 1, 2, \dots, \tau$ , are the rows of length  $n$ . In this case, the eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_\tau$  may be considered as a contribution of the principal components  $Y_1, Y_2, \dots, Y_\tau$  to the total information content of the time series  $(x_i)_{i=1}^N$ . The initial matrix can then be completely reconstructed from the derived principal components

$$X = (V^1, V^2, \dots, V^\tau) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_\tau \end{pmatrix} = \sum_{i=1}^\tau V^i Y_i;$$

in turn, the time series  $(x_i)_{i=1}^N$  can be reconstructed from it. Note that, in general, not all of the principal components  $Y_1, Y_2, \dots, Y_\tau$  but only some of them that are significant in terms of the information content are used to reconstruct the time series (see Broomhead and King 1986a, 1986b; Broomhead and Jones 1989; Vautard *et al.* 1992). More specifically, each row vector  $Y_i$ ,  $i = 1, 2, \dots, \tau$ , may be considered as the result of projecting a  $\tau$ -dimensional set of points, each of which is specified by the  $\tau$ -coordinate

column vector of matrix  $X$ , onto the direction that corresponds to eigenvector  $V^i$ . Thus, the series turns out to be represented as a set of  $\tau$  components  $Y_i$ ; the weight of component  $Y_i$  in the initial sequence  $(x_i)_{i=1}^N$  can be specified via the corresponding eigenvalue  $\lambda_i$ , which, in turn, corresponds to eigenvector  $V^i$ .

The transformations  $Y_i = (V^i)^T X, i = 1, 2, \dots, \tau$ ,  $Y_i[l] = \sum_{q=1}^r v_q^i x_q^l, l = 1, 2, \dots, n$ , are linear filters. In the case under consideration, the eigenvectors  $V^i$  are the transition functions of these linear filters, while the filters themselves are tuned to the components of the multidimensional series  $X$  and, consequently, to the components of the initial series  $(x_i)_{i=1}^N$ .

Each  $i$ th eigenvector includes  $\tau$  components, i.e.,  $V^i = (v_1^i, v_2^i, \dots, v_\tau^i)^T$ . Let us construct a dependence of components  $v_k^i, k = 1, 2, \dots, \tau$  on their number:  $v^i = v^i(k)$ . Using the orthogonality of eigenvectors, the subsequent analysis of sequence  $(x_i)_{i=1}^N$  can then be performed by studying the diagrams constructed by analogy with Lissajous figures. More specifically, the components  $v_k^i$  and  $v_k^j$  are plotted in pairs along the axes. If the constructed diagrams are nearly circular, then the functions  $v^i = v^i(k)$  and  $v^j = v^j(k)$  will be similar to periodic functions with close amplitudes and with a phase shift of about a quarter of the period.

Thus, a quantity that has the meaning of a period can be calculated for some pairs of eigenvectors  $V^i$  and  $V^j$ . Consequently, a graphical analysis gives an idea of the frequencies of the components that form the initial time series  $(x_i)_{i=1}^N$ .

For a given  $\tau$ , the number of all possible pairs of the principal components is  $\sim \tau^2$ . Clearly, all these pairs are very difficult to analyze even at small  $\tau$ . Moreover, since only a few plots are spiral in shape at large  $\tau$ , the range of search should be narrowed before beginning a graphical analysis. This can be easily done if we arrange  $V^i$  and  $Y_i$  in order of decreasing eigenvalues and if we consider only those pairs of eigenvectors that have close values of  $\lambda_i$ . In the  $\lambda = \lambda(i)$  diagram, these pairs at sufficiently large  $\lambda$  appear as steps against the background of a general decrease in  $\lambda(i)$  with increasing  $i$ . By examining these steps, we can determine the minimum value of  $\lambda_{\min}(i)$  below which the function  $\lambda = \lambda(i)$  relaxes to an exponential tail.

(5) Suppose that only the first  $r$  of the  $\tau$  components were retained for the subsequent analysis. The

first  $r$  eigenvectors  $V^i$  are then used to reconstruct the initial matrix  $X$ . In that case,

$$\tilde{X} = (V^1, V^2, \dots, V^r) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_r \end{pmatrix} = \sum_{i=1}^r V^i Y_i,$$

where  $\tilde{X}$  is the reconstructed matrix with  $n$  columns and  $\tau$  rows. The initial time series reconstructed from this matrix is now defined as

$$\tilde{x}_s = \begin{cases} \frac{1}{s} \sum_{i=1}^s \tilde{x}_{i,s-i+1}, & 1 \leq s \leq \tau \\ \frac{1}{\tau} \sum_{i=1}^{\tau} \tilde{x}_{i,s-i+1}, & \tau \leq s \leq n \\ \frac{1}{N-s+1} \sum_{i=1}^{N-s+1} \tilde{x}_{i+s-n,n-i+1}, & n \leq s \leq N. \end{cases}$$

This method of obtaining the sequence  $(\tilde{x}_i)_{i=1}^N$  is called a SSA smoothing of the initial time series  $(x_i)_{i=1}^N$  over the first  $r$  components of  $\tau$ .

(6) At the next stage of SSA application, one may consider a prediction of the initial time sequence (see Keppenne and Ghil 1995, Danilov 1997, Ghil 1997, and references therein), i.e., a construction of series  $(x_i)_{i=1}^{N+p}$ , which is an extension of the known data  $(x_i)_{i=1}^N$ . In turn, the prediction for  $p$  points ahead reduces to applying the operation of prediction for one point  $p$  times.

The basic idea of finding  $x_{N+1}$  is as follows. Let there be a set of  $x_1, x_2, \dots, x_N$ . We now construct a sample in the form of matrix  $X$ . The previously selected eigenvectors  $V^1, V^2, \dots, V^r$  of matrix  $C$  may be taken as a basis of the surface containing this sample.

Let us write the parametric equation for this surface as  $S(P) = \sum_{i=1}^r p_i V^i$ , where the set of  $r$  parameters  $p_i$  corresponds to each value of vector  $S(P)$ , which is a column of height  $\tau$ . In that case, the  $k$ th ( $k = 1, 2, \dots, n$ ) column of the initial matrix  $X$  has its own set of parameters  $P^k = (p_1^k, p_2^k, \dots, p_r^k)$  and, consequently,

$$X^1 = S(P^1), X^2 = S(P^2), \dots, X^n = S(P^n).$$

To predict  $x_{N+1}$ , it is necessary to construct the  $(n+1)$ th column  $X^{n+1}$  of matrix  $X$ , which, in turn, corresponds to some value of parameters  $P^{n+1} = (p_1^{n+1}, p_2^{n+1}, \dots, p_r^{n+1})$ . This set of parameters can be found from the relation  $S(P) = \sum_{i=1}^r p_i V^i$  based on the

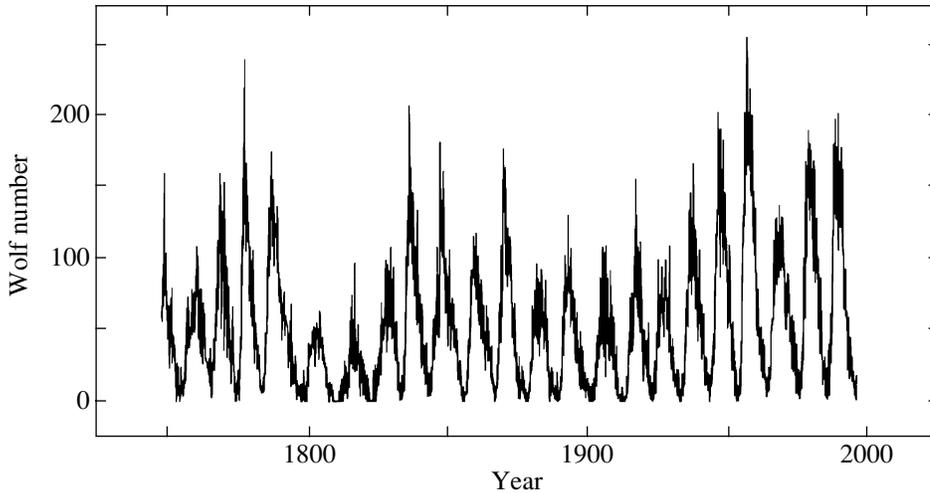


Fig. 1. Monthly mean Wolf numbers.

$(x_i)_{i=1}^N$  data alone. The predicted column is written as  $X^{N+1} = S(P^{n+1})$ .

Let us introduce the following notation:

$$V_* = \begin{pmatrix} v_1^1 & v_1^2 & \dots & v_1^r \\ v_2^1 & v_2^2 & \dots & v_2^r \\ \vdots & \vdots & \ddots & \vdots \\ v_{\tau-1}^1 & v_{\tau-1}^2 & \dots & v_{\tau-1}^r \end{pmatrix},$$

$$\tilde{p} = \begin{pmatrix} \tilde{p}_1^{n+1} \\ \tilde{p}_2^{n+1} \\ \vdots \\ \tilde{p}_r^{n+1} \end{pmatrix}, \quad Q = \begin{pmatrix} x_{N-\tau+2} \\ x_{N-\tau+3} \\ \vdots \\ x_N \end{pmatrix},$$

$$V_\tau = (v_\tau^1, v_\tau^2, \dots, v_\tau^r).$$

The parameters  $(p_1^{n+1}, p_2^{n+1}, \dots, p_r^{n+1})$  can be determined from the system of equations  $V_* \tilde{P} = Q$  for  $\tilde{P}$ . Thus, the final expression for the predicted value reads

$$x_{N+1} = \frac{V_\tau V_*^T Q}{1 - V_\tau V_\tau^T}.$$

In the simplest case, to predict the next values requires only appropriately changing the matrix  $Q$  and again multiplying it by  $V_\tau V_*^T / (1 - V_\tau V_\tau^T)$ . Additionally, however, the entire SSA algorithm can be partially or completely repeated for each next point. In this case, the matrices  $V_\tau$  and  $V_*$  also change.

(7) At the final stage of SSA application, one chooses the main parameter, the number of delays  $\tau$  used to construct the multidimensional sample  $X$ .

As with the selection of principal components, the choice of  $\tau$  significantly depends on the problem being studied.

Let the problem consist in smoothing a series by the SSA method, i.e., in reconstructing the series using known periodicities. In that case, as was already noted above, separating the principal component is filtering the series with the filter transition function in the form of an eigenvector of this principal component. The larger  $\tau$ , the larger the number of parallel filters, the narrower the passband of each of them, and the better the series smoothing.

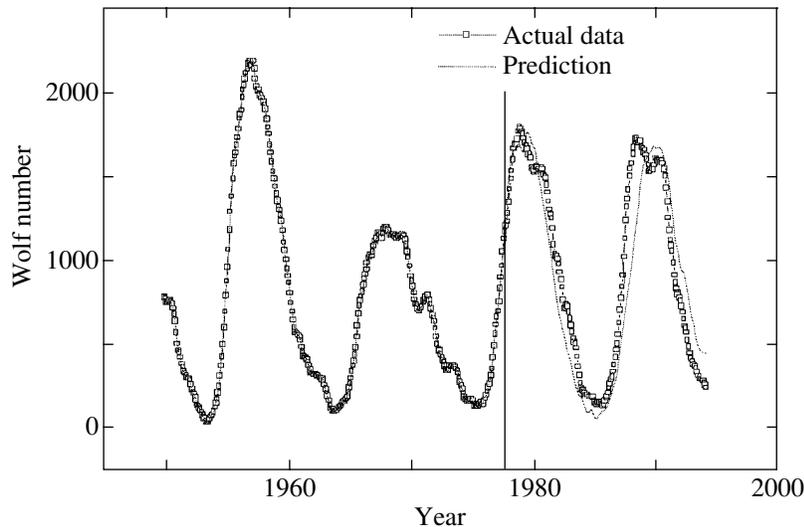
If unknown (hidden) periodicities must be determined in the observed sequence, then we should first take the largest possible value of  $\tau$ . After rejecting nearly zero eigenvalues, the delay must then be reduced.

Suppose that it is necessary to separate one known periodicity. In this case, we should choose  $\tau$  to be equal to the sought-for period.

Finally, let the problem consist in extending the series under study by a specified value (i.e., in predicting the evolution of the observed process). We should then take the maximum admissible value of  $\tau$  and then choose  $r$ .

## USING THE SSA METHOD TO ANALYZE SOLAR MAGNETIC ACTIVITY

Here, we apply the SSA method to observational data on solar activity. Wolf proposed to use the sunspot number as a measure of solar activity in 1848. To this end, he considered the sum of the total number of spots seen on the solar disk and ten times the number of regions in which these spots were located.



**Fig. 2.** Prediction of the solar activity for 216 points (18 years ahead) from monthly mean Wolf numbers. The vertical line corresponds to the boundary of the rejected values. We used 500 and 150 components for the decomposition and reconstruction, respectively. The computations were performed in three stages. After predicting another 72 points, the components of the series were recomputed.

The latter summand is intended to reconcile the measurements made under different conditions. By comparing the previous observations obtained from various sources, Wolf reconstructed the solar-activity data until 1818 with several small gaps and with acceptable accuracy. Later, the monthly means were reconstructed until 1749 (this series was used here) and the yearly means until 1700. In the latter case, however, the error in the data can be several tens of percent.

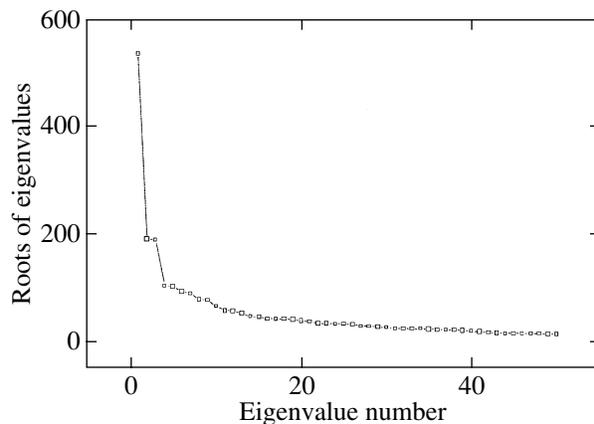
The sequence chosen for our analysis spans a wide period, from January 1749 until December 1996, without gaps and with good time resolution (see Fig. 1, where the time at intervals of one month and the corresponding Wolf number are plotted along the horizontal and vertical axes, respectively). Thus, there is a total of 2976 values.

At the first stage of SSA application, the maximum admissible value of  $\tau$  should be taken. For our studies, we chose  $\tau = 500$ , which allowed the periodicities up to a period of 42 years to be covered. Using larger  $\tau$  significantly complicated numerical calculation. Moreover, a slight increase in  $\tau$  (to 600) did not cause any significant changes in the results of the first principal-component decomposition but considerably reduced the computer resources.

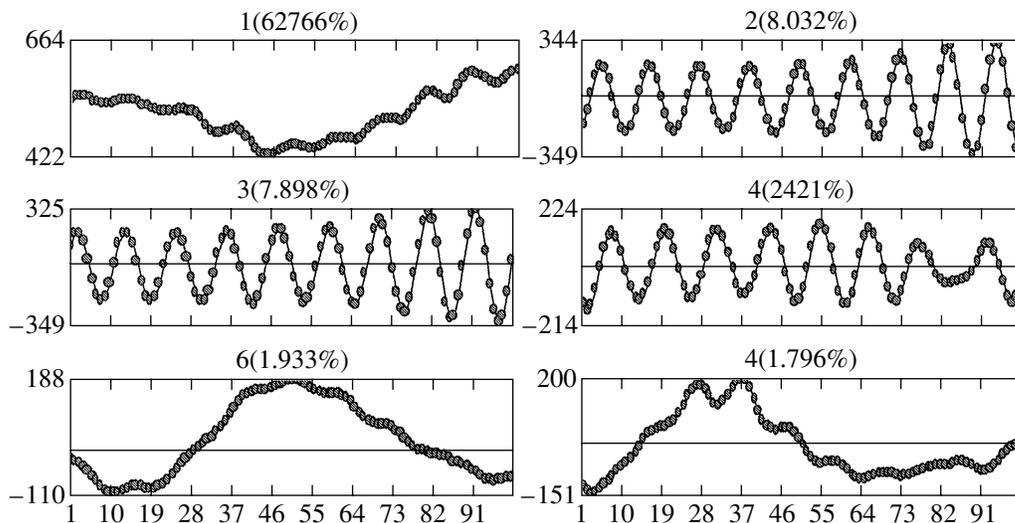
Because of the large  $\tau$ , the sequence of the roots of eigenvalues for the matrix of the second moments arranged in decreasing order rapidly relaxed to an exponential tail. In combination with a large number of initial points, this leads to the fact that even the first principal component represents only a slight smoothing of the initial series, and it is almost completely

reconstructed from the first four or five components (the sum of the first five eigenvalues exceeds 99% of their total). Moreover, the form of the first principal component changes little at small  $\tau$ , for example, at  $\tau = 5$ , which is attributable to the SSA stability for this parameter. Therefore, using a large  $\tau$  is justifiable only from the viewpoint of prediction.

To test the prediction by the SSA method, let us truncate the sequence of *monthly* mean Wolf numbers on the right by 216 points (18 years) and try to reconstruct it according to the following scheme. Determine the optimum parameters of the algorithm



**Fig. 3.** The first 50 eigenvalues of the matrix of the second moments when decomposing the series of yearly mean Wolf numbers into 123 components. The eigenvalues are arranged in decreasing order.



**Fig. 4.** Some components corresponding to the eigenvalues in Fig. 3. Their percentage contribution to the initial series is given in parentheses.

for reconstructing this series by an additional truncation of the derived series and decompose it into  $\tau = 500$  components. In this case, we must choose such a number of the first components for which the agreement between the predicted values and these additionally rejected data is best. Subsequently, we reconstruct the initially truncated part of 216 points by using the parameters found.

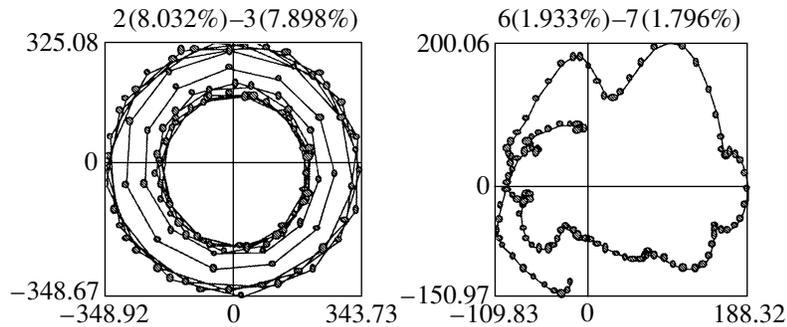
It can be established by a direct exhaustive search that the best results are obtained for  $r = 150$  (the number of selected components). Let us again take the initial series truncated only by 216 points and use the  $r$  chosen for its prediction. The prediction quality can be further increased if we break down the predicted interval into segments and recompute the principal components after predicting each of these segments. Ideally, such a recomputation must be done after predicting each point, but this increases the computational time. Figure 2 shows our prediction for which the components were recomputed three times at intervals of 72 points (which is almost identical to a prediction for 216 points with no breakdown into intervals).

We could try to analyze the components of the series for the presence of particular periods or for the separation of known periods. However, their large number and the associated similarity between the components  $Y_i$  of the initial series makes this problem very complex (although quite solvable); i.e., the information contained in the series of monthly mean Wolf numbers is, in a sense, redundant. In addition, since something certain can hardly be said about periodicities of several months, even in the case of their separation, it is easier to take a series with a large time step.

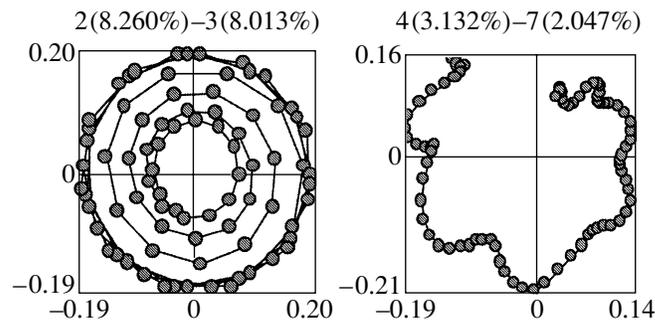
Let us now consider the series of *yearly* mean Wolf numbers. Since the series contains a mere 248 points, the maximum possible delay is  $\tau = 123$ . Let us choose it as the initial one. The first 50 eigenvalues are shown in Fig. 3. The first value gives the principal component responsible for the trend; the steps form the pairs of components with numbers 2–3, 4–5, 6–7, 8–9, and 11–12; and the dependence relaxes to an exponential tail starting from number 14. The eigenvectors for pairs 2–3, 4–5, 8–9, and 11–12 (Fig. 4) correspond to 11-year periodicities (the spiral two-dimensional plot for components 2 and 3 is shown in the left panel of Fig. 5). However, in addition to this (obvious) 11-year cycle, Gleisberg's presumed 80-year cycle shows up (see the pair of eigenvectors 6 and 7 in Fig. 5). Since the corresponding eigenvalues are not quite equal (the step is skewed; see Fig. 3) and since the phase shift differs from  $\pi/2$ , the plot for them is not spiral in shape. Nevertheless, the periodicity is traceable, although very small eigenvalues correspond to it.

For a better separation of the 80-year periodicity, we can try to adjust  $\tau$  for it. For the decomposition of the series with  $\tau = 80$  of eigenvectors, vectors 4 and 7 correspond to this period. The eigenvalues that correspond to these vectors are close, but the total contribution of these components at the given  $\tau$  exceeds 5%. The two-dimensional plot for components 4 and 7 resembles a circumference (Fig. 6). Figure 7 shows the 80-year cycle obtained by reconstructing the series from these two components alone.

Let us now use the SSA method to predict the series of yearly mean Wolf numbers. Let us truncate it on the right by 18 years (i.e., by 18 points) and



**Fig. 5.** Diagrams for the vector components 2–3 (left) and 6–7 (right) shown in Fig. 4. The left and right panels correspond to the first and third steps of eigenvalues in Fig. 3, respectively.

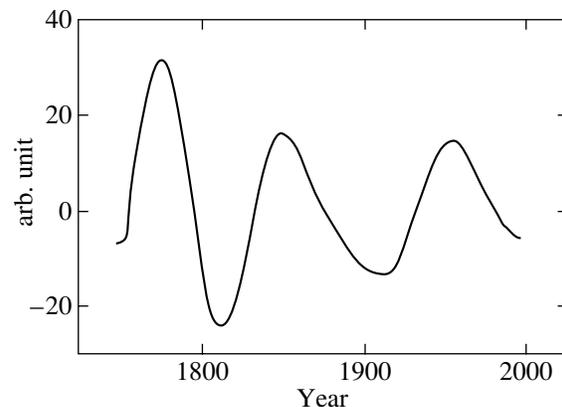


**Fig. 6.** Plots for the vector components 2–3 (left) и 4–7 (right) at  $\tau = 80$ .

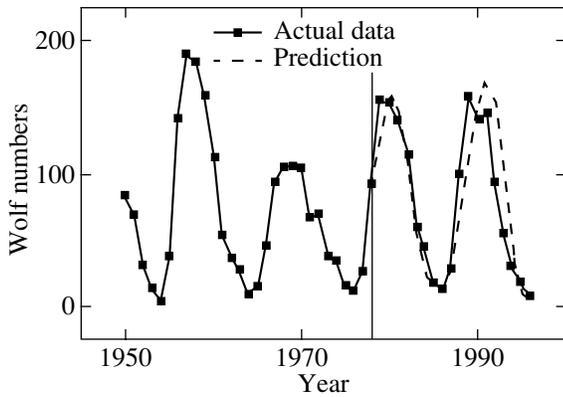
try to reconstruct them. To this end, we additionally remove an 11-point-long segment and attempt to reconstruct them with the maximum possible accuracy. We perform this procedure in different parts of the series by choosing the best number of eigenvectors for the reconstruction.

For the derived 219-point-long series ( $219 = 248 - 18 - 11$ , where 248 is the total number of points; see above), the maximum admissible  $\tau$  is 109. As numerical analyses show, the prediction for 11 points strongly depends on the choice of the components used for the reconstruction. However, the qualitative picture is satisfactory for a wide range of choices. Nevertheless, this value of  $\tau$  is clearly large for a quantitative prediction. At its lower values, it is much easier to select the required number of components. For example, at  $\tau = 33$ , the prediction is satisfactory when choosing the first 11 components. Therefore, we use these parameters to reconstruct the 18 distant points. In contrast to the case with monthly mean initial data and  $\tau = 500$ , it is now possible, for  $\tau = 33$ , to recompute the eigenvectors every time by taking into account the last predicted point (both at the stage of vector selection for the prediction and during the prediction itself). The prediction corrected at each step is shown in Fig. 8.

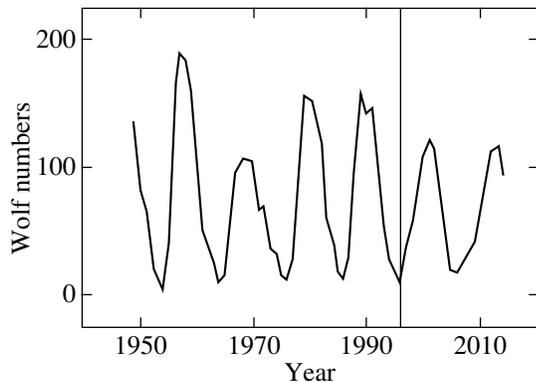
In addition, we considered a series of natural logarithms of the initial series. Taking the logarithm is commonly used to analyze time series (e.g., in a correlation analysis) and occasionally yields better results. As with yearly mean Wolf numbers, we use  $\tau = 123$ . In this case, the general form of the eigenvalues and eigenvectors did not change fundamentally. Thus, it was unnecessary to use the logarithmic series



**Fig. 7.** Reconstruction of components 4 and 7, which correspond to the 80-year solar cycle.



**Fig. 8.** Reconstruction of yearly mean Wolf numbers corrected at each step. We chose 11 components from 33. The boundary of the rejected 18 points is marked by the vertical straight line.



**Fig. 9.** Prediction of solar activity until 2014.

in this case, because all the main periodicities and the prediction were obtained from the initial series.

At the final stage, we used the SSA method for the actual prediction of solar activity. To this end, we took the series of yearly mean Wolf numbers from 1748 until 1996. Since this yearly mean series ends at a minimum of solar activity, it is of interest to describe its two subsequent maxima. This requires extending the yearly mean sequence by 18 points. We decomposed the series into 33 components and selected the first 11 of them for the prediction.

The prediction until 2014 is shown in Fig. 9. It can be seen from this figure that the maximum of solar cycle 23 occurs in the first half of 2000 and its amplitude is 122. The error in the maximum amplitude is typically of the order of 5–10. This is in excellent agreement with the maximum of Wolf numbers that apparently occurred in the middle of 2000 and that had a smoothed value of 121. As we see from the figure, in the immediate future, the Sun will be relatively quiet (compared to the two previous 11-year periods), while the maximum of solar cycle 24

(presumably in 2011) will be comparatively low. Its estimation yields a Wolf number of 117 (Fig. 9).

For comparison and testing the stability of our result, we also analyzed the series of monthly mean Wolf numbers. However, this analysis introduced no significant additions in the prediction.

It follows from our experience of SSA application that the prediction for more than one and a half 11-year cycles is not informative. At the same time, the Wolf numbers in Fig. 9 (prediction for 18 years ahead) may be considered quite reliable.

## CONCLUSIONS

An analysis of time series by the SSA method will probably rank high among the various techniques used to analyze and predict experimental data. Since the initial series is decomposed into components whose analytic form is not fixed in advance, the SSA method allows us to satisfactorily separate components with specified periods from the series and to predict its dynamics. The only criterion for SSA applicability is the information content of the sequence under study. In this case, the constraints on the number of points and characteristic periods are generally much weaker than in other methods (e.g., in a correlation or Fourier analysis).

Here, we have considered the possibility of analyzing the sequence of Wolf numbers that characterize solar activity by the SSA method. Despite the relatively small length of this sequence, the method makes it possible to reveal its components that correspond to the already known solar cycles and allows for a reconstruction using only some of its components. We also found that even short series of observations could be predicted with acceptable accuracy by the SSA method.

Like any other method, SSA has its own drawbacks. First, there is a problem of accurately determining the unknown frequencies in the sequence under study (for a sufficiently long series, this problem can be solved by a Fourier analysis). Second, SSA does not include well-defined component selection rules for a reconstruction, particularly for a prediction. Finally, when applied to an analysis of solar activity, it does not allow the occurrence times of activity peaks (maximum Wolf numbers) to be accurately estimated, although it gives an accurate estimate of their amplitude. As a result, a systematic phase shift is accumulated in the predicted series. Nevertheless, as follows from our analysis, the SSA method described above serves as a good supplement to the available experimental data reduction techniques, particularly for analyzing and predicting fairly short time series to which the series of Wolf numbers belongs (there are only data on 23 11-year cycles).

In the next papers on solar-activity analysis, we will explore the possibility of correcting the cycle phase by using other prediction methods and solar-cycle regularities. It is also possible to improve the governing parameters of this method and/or to use additional information, for example, an empirical relationship between solar-cycle amplitude and phase (Dmitrieva *et al.* 2000). Note also that the decomposition components in the SSA method are generally harmonic functions. However, because of the nonlinearity in the behavior of the solar magnetic-field generation mechanism, significant anharmonicity emerges. As a result, the rise phase of the solar cycle is appreciably shorter than its decline phase. Although the maximum amplitude in the 11-year cycle calculated by SSA is close to its true value, its occurrence time is determined inaccurately. Hence, significant errors are possible in the current Wolf numbers at the rise and decline phases immediately before and after the maximum of the 11-year cycle.

In general, it can be said that SSA is fairly efficient and very promising method for predicting the dynamics of solar magnetic activity. It can also be used to study regularities in series of other astrophysical indices.

#### ACKNOWLEDGMENTS

This study was supported in part by the Russian Foundation for Basic Research (project no. 00-02-17854) and a youth grant of the Russian Academy of Sciences. K.M. Kuzanyan wishes to thank the NATO grant PST.CLS.976557.

#### REFERENCES

1. V. S. Afraïmovich and A. M. Reïman, in *Nonlinear Waves. Dynamics and Evolution* (Nauka, Moscow, 1989), p. 238.
2. D. S. Broomhead and R. Jones, *Proc. R. Soc. London* **423**, 103 (1989).
3. D. S. Broomhead and G. P. King, *Physica D* (Amsterdam) **20**, 217 (1986).
4. D. S. Broomhead and G. P. King, in *Nonlinear Phenomena and Chaos*, Ed. by S. Sarkar (Adam Hilger, Bristol, 1986), p. 113.
5. L. Cao, Y. Hong, H. Fang, and G. He, *Physica D* (Amsterdam) **85**, 225 (1995).
6. M. Casdagli, *Physica D* (Amsterdam) **35**, 335 (1989).
7. D. L. Danilov, *Principal Components of Time Series: The Caterpillar Method* (St. Petersburg State Univ., St. Petersburg, 1997).
8. D. L. Danilov and A. A. Zhiglyavskii, *Principal Components in Time Series: Caterpillar Method* (St. Petersburg State Univ., St. Petersburg, 1997).
9. J. Deppish, H.-U. Bauer, and T. Geisel, *Phys. Lett. A* **158**, 57 (1991).
10. I. V. Dmitrieva, K. M. Kuzanyan, and V. N. Obridko, *Sol. Phys.* **195** (1), 209 (2000).
11. M. Ghil, *Proc. SPIE* **3165**, 216 (1997).
12. D. H. Hathaway, R. M. Wilson, and E. J. Reichmann, *J. Geophys. Res.* **104** (A10), 22375 (1999).
13. D. V. Hoyt and K. H. Schatten, *Sol. Phys.* **181**, 491 (1998).
14. D. T. Kaplan and L. Glass, *Phys. Rev. Lett.* **68**, 427 (1992).
15. C. L. Keppenne and M. Ghil, *Exp. Long-Lead Forecast Bulletin* (National Meteorological Center, NOAA, US Department of Commerce, 1992–1995), Vol. 1, Nos. 1–4; Vol. 2, Nos. 1–4; Vol. 3, Nos. 1–4; Vol. 4, Nos. 1–2.
16. J. K. Lawrence, A. A. Ruzmaikin, and A. C. Cadavid, *Astrophys. J.* **417**, 805 (1993).
17. J. K. Lawrence, A. A. Ruzmaikin, and A. C. Cadavid, *Astrophys. J.* **455**, 366 (1995).
18. A. Yu. Loskutov and A. S. Mikhāilov, *An Introduction to Synergetics* (Nauka, Moscow, 1990).
19. G. G. Malinetskii and A. B. Potapov, *Modern Applied Nonlinear Dynamics* (URSS, Moscow, 2000).
20. D. B. Murray, *Physica D* (Amsterdam) **68**, 318 (1993).
21. Yu. A. Nagovitsyn, *Pis'ma Astron. Zh.* **23**, 851 (1997) [*Astron. Lett.* **23**, 742 (1997)].
22. D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications. Multitaper and Conventional Univariate Techniques* (Cambridge Univ. Press, Cambridge, 1993).
23. D. Ruelle, *Proc. R. Soc. London* **427** (1873), 241 (1990).
24. T. Sauer, Y. A. Yorke, and M. Casdagli, *J. Stat. Phys.* **65**, 579 (1991).
25. J. Theiler, S. Eubank, A. Longtin, *et al.*, *Physica D* (Amsterdam) **58**, 77 (1992).
26. K. Schatten, *Astron. Soc. Pac. Conf., Ser.* **154**, 1315 (1997).
27. R. Vautard, P. Yiou, and M. Ghil, *Physica D* (Amsterdam) **58**, 95 (1992).
28. R. M. Wilson, D. H. Hathaway, and E. J. Reichmann, *J. Geophys. Res.* **103**, 17411 (1998).

*Translated by V. Astakhov*