1 Example of Time Series Analysis by SSA[1]

Let us illustrate the 'Caterpillar'-SSA technique [1] by the example of time series analysis. Consider the time series FORT (monthly volumes of fortified wine sales in Australia from January 1984 till June 1994). We examine the whole time series (of length N=174) as well as its subseries formed from the first 120 points. We use designations FORT174 for the former and FORT120 for the latter. Presented figures illustrate the theoretical concepts and should help to get a better understanding of the method approach. Visual analysis of the time series depicted in Fig. 1 indicates that this series has a trend and this trend can be approximated either by linear function or by exponentially decreasing function. Also it seems that the seasonal component has complicated and changing behavior. The periodogram the centered time series (Fig. 2) confirms this assumption. This periodogram is constructed only on the first 168 points for refining the representation, as 168 is divisible by 12, the number of months per year.

The aim of the 'Caterpillar'-SSA analysis is decomposition of this time series into three components: a trend, a seasonal component and a noise.

The choice of window length. At first, let us guess the possible dimension of the signal (composed of a trend and a periodical component). The trend of the FORT time series is likely described by eigentriples 1–2; the seasonal component consists of harmonics with frequencies 1/12 (annual periodicity), 1/6=2/12 (half-year), 1/4=3/12, 1/3=4/12, 1/2.4=5/12, 1/2=6/12. Dimension of each harmonic with frequency less than 1/2 is equal to 2, whereas the harmonic with frequency 1/2 has the dimension 1 (we assume that changes of the harmonics amplitudes have exponential character, perhaps with different rates for different harmonics). Thus, the expected dimension of the time series should not exceed 13. If the time series didn't include a noise and its components were strongly separable from each other, then it would be sufficient to use window length 13 for time series decomposition into a trend and a seasonal component. However a exact separability doesn't practically exist for real-life data and so we need to use theoretical results about approximate (asymptotic) separability of a slowly varying trend and a harmonic. To obtain better accuracy we need to choose window length L close to a half of the time series length (since a convergence rate of separability error to zero has the order $1/\min(L, K)$, where $K = N - L + 1$, N denotes the length of our time series). We know that for getting better separability of periodical components we should select L and K divisible by period. Moreover it is more important that the smaller of L and K number is divisible by 12, so we choose window length $L = 84$ ($N = 174$, hence $K = 91$).

The method of eigentriples identification. Let us consider the result of Singular Value Decomposition of the trajectory matrix performed with the chosen window length. Fig. 3 represents eigenvectors from the first six eigentriples (recall that eigentriple=(square root of eigenvalue, eigenvector, factor vector)= (singular value, left singular vector, right singular vector)). It should be remarked that the form of factor vectors (so-called right singular vectors) is almost the same as the form of left ones (eigenvectors) because L is close to K. If we took smaller window length, then the eigenvectors would have more regular form in comparison with factor vectors, since the latter ones would reflect the varying of harmonic components amplitudes.

Let us use the conclusions about singular vectors form for identification of eigentriples corresponding to a trend and harmonics under the assumption of their approximate separability.

Trend identification. Let us start with identification of a trend. We know that singular vectors have (in general) the same form as the corresponding components of the initial time series. Thus we should find slowly varying eigenvectors. It can be done by consideration of one-dimensional plots of eigenvectors. In this instance only the leading eigenvector has the

[1] This is the example section of [2]. See also http://www.gistatgroup.com/cat/

required form, i.e., trend is described by one eigentriple. This directly implies that our trend is approximated by an exponential function. Note that the more complicated form a trend has, the larger (approximate) dimension it has and larger amount of eigentriples corresponds to it.

Harmonics identification. Now we will try to identify harmonic components (possibly with varying amplitudes) produced by the seasonal component of the original time series. As shown in Fig. 3, eigentriples 2–6 correspond to some harmonics, since their singular vectors have the regular periodical form. We know that every harmonic with frequency smaller than 0.5 produces two eigentriples. (A harmonic with frequency 0.5 generates one eigentriple with saw-tooth singular vectors as they are also harmonics with period 2; here we do not see such vectors.) One way to find pairs of components corresponding to harmonics is consideration of two-dimensional plots of singular vectors.

It is enough to examine only plots which are built for adjacent eigentriples (ordered by eigenvalues) because it is known that eigenvalues from the corresponding to harmonic pair are close for a sufficiently long time series. In Fig. 4 you can recognize regular two-dimensional graphs which form two-dimensional trajectories with vertices in a spiral-shaped curve. It indicates that these pairs of singular vectors are produced by modulated harmonic components of the initial time series. In that way, eigentriples (ET in abbreviated form) ET2,3 correspond to period 12, ET4,5 to 4, ET6,7 to 6, ET8,9 to 2.4 , ET10,11 to 3. We use a notion 'fractional period' for a harmonic with the frequency inverse to this period.

Supplementary characteristics. Let us describe additional information, which can help us to identify eigentriples and to confirm components grouping. Fig. 5 (which depicts logarithms of eigenvalues) provides such information in this way: a pair of eigentriples corresponding to a harmonic produces a plateau in this graph.

Analysis of the matrix of w-correlations between reconstructed components of initial time series is also useful for identification. Fig. 6 depicts the six leading elementary reconstructed components of the original time series, where each component is formally reconstructed by one eigentriple. Recall that w-correlation is a weighted correlation between reconstructed time series. The condition of its equality to zero is necessary for separability of the corresponding time series components. Fig. 7 confirms the performed identification: the w-correlation for components from a pair which corresponds to a harmonic is rather high whereas w-correlations between pairs and a trend are close to zero (it's reflected by white color of corresponding elements of the correlation matrix).

Separation of a signal from noise. Let us dwell upon a question of separation of components corresponding to a signal from noise components. Firstly, irregular behavior of singular vectors can indicate their belonging to a set induced by the noise component. This irregularity should be distinguished from components mixture, which is caused by lack of strong separability of these components. Boundary between signal and noise components can be confirmed by slow (almost without jumps) decreasing of eigenvalues starting from some number. Secondly, the large set of eigentriples generated by reconstructed components which are correlated between each other is quite likely to belong to a noise. Fig. 7 contains such block of eigentriples with numbers 14-84. The question about the pair ET12,13 is still open. On the one hand, this pair is well separated from the residual. On the other hand, period of the component reconstructed by ET12,13 is close to 2.33, as the periodogram indicates. Such period cannot be interpreted in the context of seasonality. It can be caused either by a noise or by a high-modulated harmonic with period 2.4. It isn't possible to come to a reliable conclusion on this specific time series component, so we will classify ET12,13 as a noise.

We apply standard statistical methods to confirm the accuracy of performed separation of a signal from a noise. Fig.8 depicts decomposition of initial time series into three components: the trend (ET1, on the background of the initial time series), the periodic (seasonal) component

(ET2-11), and the noise (ET12-84). The statistical criteria confirm (don't reject) the hypothesis that the third component is a realization of white noise (three different criteria of independence produce p-levels larger than 0.4).

Decomposition of the initial time series into components. Fig. 8 demonstrates the result of grouping of SVD components followed by diagonal averaging. This is a solution of the stated problem of the initial time series decomposition into 'independent' and 'identifiable' additive components. Table 1 contains values of w-correlations between represented in Fig. 8 components.

Таблица 1: w-Correlations for the time series decomposition into the trend, the periodical component, and the noise

	ET1	ET2–11	ET12–84
ET1	1	0	0
ET2–11	0	1	0.016
ET12–84	0	0.016	1

More detailed investigation can be interesting in addition to general time series decomposition. The matter of interest for the considered time series is the seasonal component behavior. Fig. 9 depicts the leading three components of seasonal component decomposition into the sum of harmonics. Several facts are worthy of notice: decreasing amplitude of annual harmonic, more or less invariable behavior of half-year harmonic and increasing amplitude of 4-month harmonic. Note that standard statistical methods generally assume either stationarity of amplitudes (additive models) or amplitude variations similar for all harmonics and proportional to the trend (multiplicative models). It is clear that for 'FORT' time series both models are not appropriate.

The problem of strong separability. The lack of strong separability is the problem of components mixing. It is caused by close to each other eigenvalues (weights) corresponding to different components. For demonstration of this effect, let us consider subseries consisting of the first 120 points of the initial time series. We choose the window length L equal to 60.

Fig. 10 represents the matrix of w-correlations. Just as for matrix depicted in Fig. 7 we see here that eigentriples starting from number 12 can be related to a noise. However dark-colored block formed by ET8-11 reflects probable mixing of two harmonics. Periodogram analysis of eigenvectors confirms this supposition; there is a mixing of harmonics with frequencies 1/3 and 1/2.4=5/12. It is not so essential for extraction of the entire seasonal component; this effect accounts for problems only during the identification. If we wanted to extract, for instance, a quarterly (three-month) harmonic component, then the lack of strong separability would interfere with performing such extraction. The comparison of periodograms of the (almost) entire initial time series (Fig. 2) and of its leading 120 points (Fig. 11) explains why this problem didn't arise for the initial time series: for FORT174 the contributions of frequencies 1/3 and 1/2.4 are slightly different whereas they are practically equal for the subseries FORT120.

Список литературы

[1] Golyandina N., Nekrutkin V., and Zhigljavsky A. *Analysis of Time Series Structure: SSA and Related Techniques*, London: Chapman & Hall/CRC, 2001, 305 P.

[2] Golyandina N. *The 'Caterpillar'-SSA Method for Time Series Analysis: Training Aids*, St.Petersburg, St.Petersburg State University, 2004, 87 P. (In Russian).

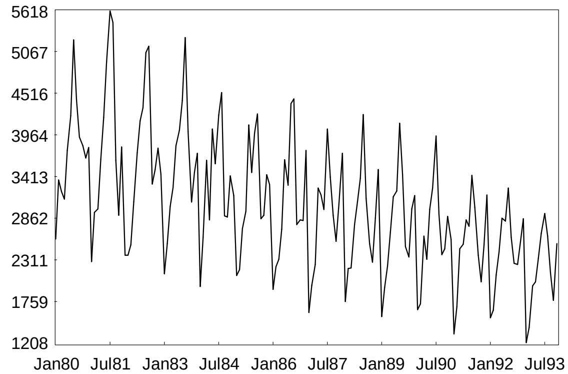

Рис. 1: FORT174: initial time series

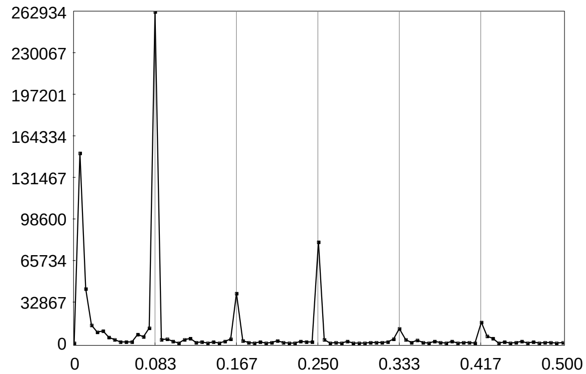

Рис. 2: FORT174: periodogram of centered time series (the first 168 points)

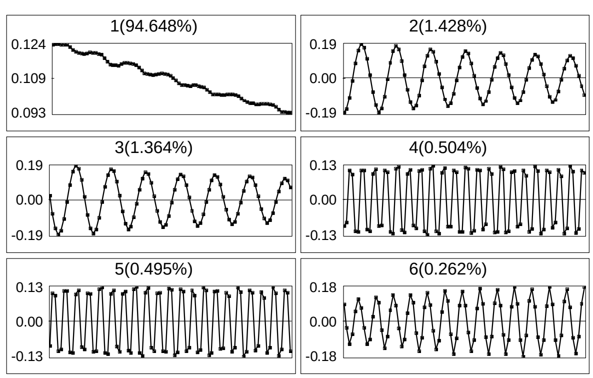

Рис. 3: FORT174: one-dimensional plots of eigenvectors

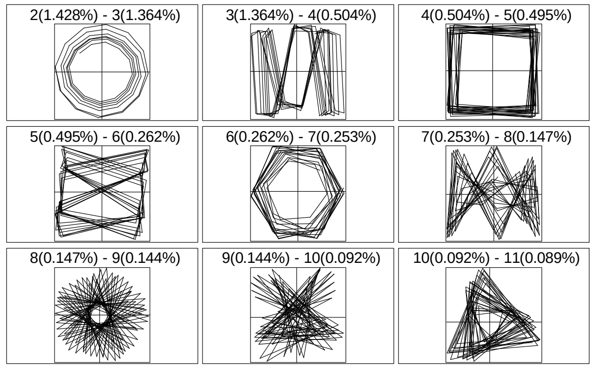

Рис. 4: FORT174: two-dimensional plots of eigenvectors

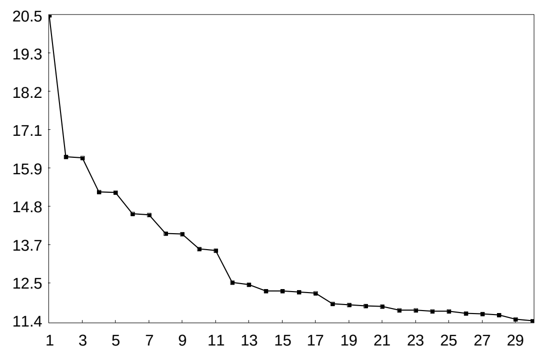

Рис. 5: FORT174: logarithms of the leading 30 eigenvalues

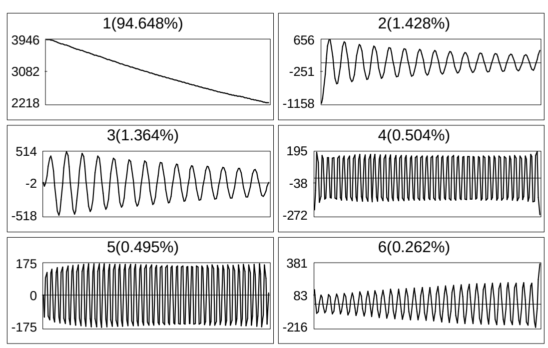

Рис. 6: FORT174: elementary reconstructed components

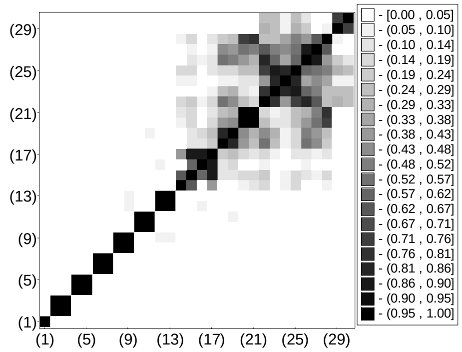

Рис. 7: FORT174: matrix of w-correlations between elementary reconstructed components

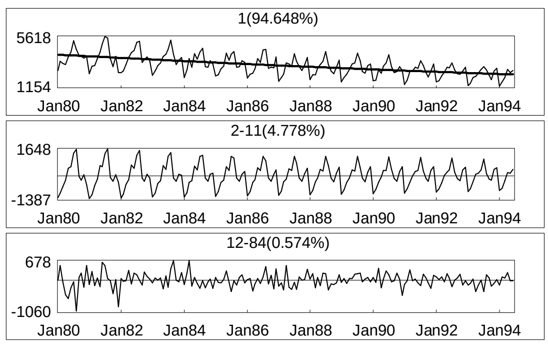

Рис. 8: FORT174: decomposition of the time series into the trend (on the background of the initial time series, see the top graph), the seasonal component (in the middle) and the noise (the bottom graph)

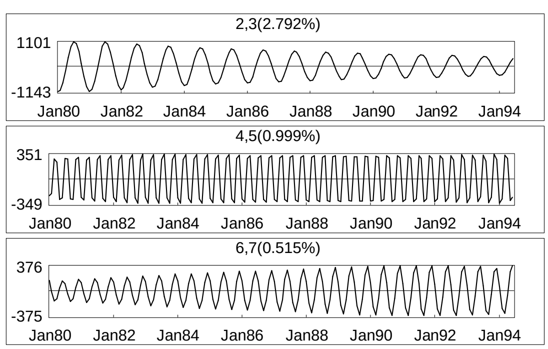

Рис. 9: FORT174: decomposition of the seasonal component into the sum of harmonics

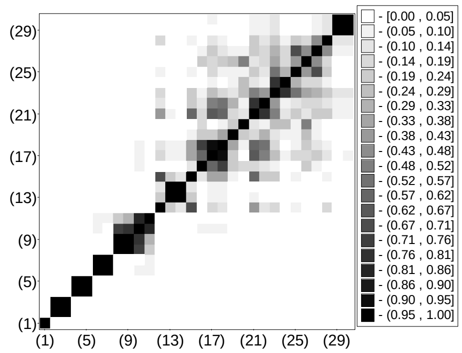

Рис. 10: FORT120: matrix of w-correlations between elementary reconstructed components

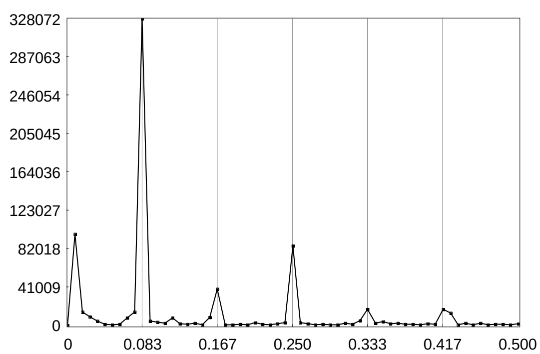

Рис. 11: FORT120: periodogram of the centered time series