A Brief Introduction to Singular Spectrum Analysis

Hossein Hassani *

1 A Brief Introduction

In recent years a powerful technique known as Singular Spectrum Analysis (SSA) has been developed in the field of time series analysis. SSA is novel and powerful technique applicable to many practical problems such as the study of classical time series, multivariate statistics, multivariate geometry, dynamical systems and signal processing.

The possible application areas of SSA are diverse: from mathematics and physics to economics and financial mathematics. Other areas may include meteorology and oceanology to social sciences, market research and medicine. Any seemingly complex time series with a potential structure of note could provide another example of a successful application of SSA [1].

The basic SSA method consists of two complementary stages: decomposition and reconstruction; both stages include two separate steps. At the first stage we decompose the series and at the second stage we reconstruct the original series and use the reconstructed series for forecasting new data points. The main concept in studying the properties of SSA is 'separability', which characterizes how well different components can be separated from each other. The absence of approximate separability is often observed in series with complex structure. For these series and series with special structure, there are different ways of modifying SSA leading to different versions such as SSA with single and double centering, Toeplitz SSA, and sequential SSA [1].

It is worth noting that although some probabilistic and statistical concepts are employed in the SSA-based methods, we do not have to make any statistical assumptions such as stationarity of the series or normality of the residuals. SSA is a very useful tool which can be used for solving the following problems:

finding trends of different resolution;
smoothing;
extraction of seasonality components;
simultaneous extraction of cycles with small and large periods;
extraction of periodicities with varying amplitudes;
simultaneous extraction of complex trends and periodicities;
finding structure in short time series;

*Cardiff School of Mathematics, Cardiff University, CF24 4AG, UK.
Email: HassaniH@cf.ac.uk

causality test based on the SSA.

Solving all these problems correspond to the so-called basic capabilities of SSA. In addition, the method has several essential extensions. First, the multivariate version of the method permits the simultaneous expansion of several time series; see, for example [3]. Second, the SSA ideas lead to several forecasting procedures for time series; see [1, 3]. Also, the same ideas are used in [1] and [7] for change-point detection in time series. For comparison with classical methods, ARIMA, ARAR algorithm and Holt-Winter, see [8] and [9], and for comparison between multivariate SSA and VAR model see [10]. For automatic methods of identification within the SSA framework see [11] and for recent work in 'Caterpillar'-SSA software as well as new developments see [12]. A family of causality tests based on the SSA technique has also been considered in [13].
In the area of nonlinear time series analysis SSA was considered as a technique that could compete with more standard methods. There are a number of research that considered SSA as a filtering method in (see, for example, [14] and references therein). In another research, the noise information extracted using the SSA technique, has been used as a biomedical diagnostic test [15]. The SSA technique also used as a filtering method for longitudinal measurements. It has been shown that noise reduction is important for curve fitting in growth curve models, and that SSA can be employed as a powerful tool for noise reduction for longitudinal measurements [16].

2 Motivation

We are motivated to use SSA because it is a nonparametric technique that works with arbitrary statistical processes, whether linear or nonlinear, stationary or non-stationary, Gaussian or non-Gaussian. Given that the dynamics of real time series has usually gone through structural changes during the time period under consideration, one needs to make certain that the method of prediction is not sensitive to the dynamical variations. Moreover, contrary to the traditional methods of time series forecasting (both autoregressive or structural models that assume normality and stationarity of the series), SSA method is non-parametric and makes no prior assumptions about the data. The real time series usually has a complex structure of this kind; as a consequence, we found superiority of SSA over classical techniques. Additionally, SSA method decomposes a series into its component parts, and reconstruct the series by leaving the random (noise) component behind.
Another important aspect of the SSA (which can be very useful in economics) is that, unlike many other methods, it works well even for small sample size (see, for example, [5] and [8]).
It should be toted that, although some probabilistic and statistical concepts are employed in the SSA-based methods, no statistical assumptions such as stationarity of the series or normality of the residuals are required and SSA uses the bootstrapping to obtain the confidence intervals for the forecasts.

3 A Short Description of the SSA Algorithm

We consider a time series $Y_T = (y_1, \ldots, y_T)$. Fix L ($L \leq T/2$), the window length, and let $K = T - L + 1$.

Step 1. (*Computing the trajectory matrix*): this transfers a one-dimensional time series $Y_T = (y_1, \ldots, y_T)$ into the multi-dimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1})' \in \mathbf{R}^L$, where $K = T - L + 1$. The single parameter of the embedding is the *window length* L, an integer such that $2 \leq L \leq T$. The result of this step is the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K]$:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{pmatrix}.$$

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

Step 2. (*Constructing a matrix for applying SVD*): compute the matrix $\mathbf{XX}^T$.

Step 3. (*SVD of the matrix* $\mathbf{XX}^T$): compute the eigenvalues and eigenvectors of the matrix $\mathbf{XX}^T$ and represent it in the form $\mathbf{XX}^T = P\Lambda P^T$. Here $\Lambda = diag(\lambda_1, \ldots, \lambda_L)$ is the diagonal matrix of eigenvalues of $\mathbf{XX}^T$ ordered so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L \geq 0$ and $P = (P_1, P_2, \ldots, P_L)$ is the corresponding orthogonal matrix of eigen–vectors of $\mathbf{XX}^T$.

Step 4. (*Selection of eigen–vectors*): select a group of l ($1 \leq l \leq L$) eigen–vectors $P_{i_1}, P_{i_2}, \ldots, P_{i_l}$.
The grouping step corresponds to splitting the elementary matrices $\mathbf{X}_i$ into several groups and summing the matrices within each group. Let $I = \{i_1, \ldots, i_l\}$ be a group of indices $i_1, \ldots, i_l$. Then the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_l}$.

Step 5. (*Reconstruction of the one-dimensional series*): compute the matrix $\tilde{\mathbf{X}} = ||\tilde{x}_{i,j}|| = \sum_{k=1}^{l} P_{i_k} P_{i_k}^T \mathbf{X}$ as an approximation to $\mathbf{X}$. Transition to the one–dimensional series can now be achieved by averaging over the diagonals of the matrix $\tilde{\mathbf{X}}$.

4 Some examples

4.1 Economics series

Let us now use the Fabricated metal series for Germany as an example to illustrate the selection of the SSA parameters and to show the reconstruction of the original series in detail (for more information see [9]). To perform the analysis, we have used the SSA software[1]. Fig. 1 presents the series, indicating a complex trend and strong seasonality.

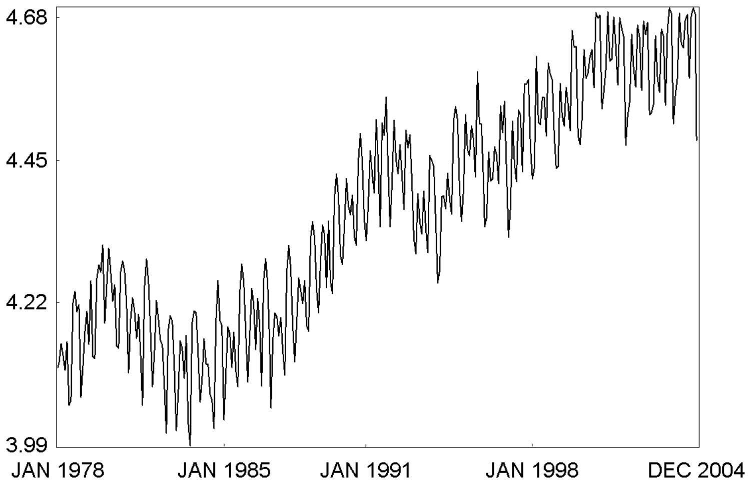

Figure 1: Fabricated metal series in Germany

Selection of the window length L

The window length L is the only parameter in the decomposition stage. Knowing that the time series may have a periodic component with an integer period, to achieve a better separability of this periodic component it is advisable to take the window length proportional to that period. For example, the assumption that there is an annual periodicity in the series suggests that we must pay attention to the frequencies $k/12$ ($k = 1, ..., 12$). As it is advisable to choose L reasonably large (but smaller than $T/2$ which is 162 in this case), we choose $L = 120$.

Selection of r

Auxiliary information can be used to choose the parameters L and r. Below we briefly explain some methods that can be useful in the separation of the signal from noise. Usually a harmonic component produces two eigentriples with close singular values (except for the frequency 0.5 which provides one eigentriple with the saw-tooth singular vector). Another useful insight is provided by checking breaks in the eigenvalue spectra. Additionally, a pure noise series typically produces a slowly decreasing sequence of singular values.

[1] http://www.gistatgroup.com/cat/index.html

Choosing $L = 120$ and performing SVD of the trajectory matrix $\mathbf{X}$, we obtain 120 eigentriples, ordered by their contribution (share) in the decomposition. Fig. 2 depicts the plot of the logarithms of the 120 singular values.

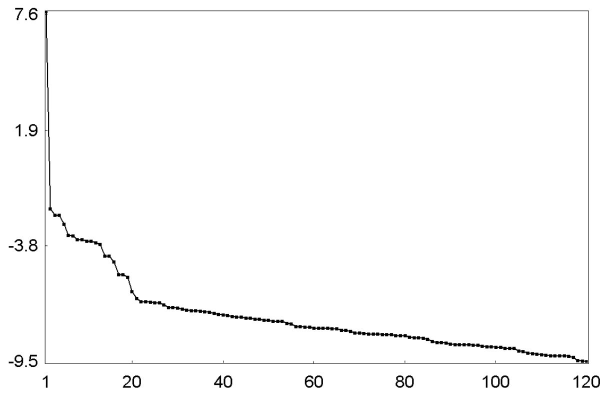

Figure 2: Logarithms of the 120 eigenvalues.

Here a significant drop in values occurs around component 19 which could be interpreted as the start of the noise floor. Six evident pairs, with almost equal leading singular values, correspond to six (almost) harmonic components of the series: eigentriple pairs 3-4, 6-7, 8-9, 10-11, 14-15 and 17-18 are related to the harmonics with specific periods (we show later that they correspond to the periods of 6, 4, 12, 3, 36 and 2.4 months).
Another way of grouping is to examine the matrix of the absolute values of the w-correlations. Fig. 3 shows the w-correlations for the 120 reconstructed components in a 20-grade grey scale from white to black corresponding to the absolute values of correlations from 0 to 1. Based on this information, we select the first 18 eigentriples for the reconstruction of the original series and consider the rest as noise.

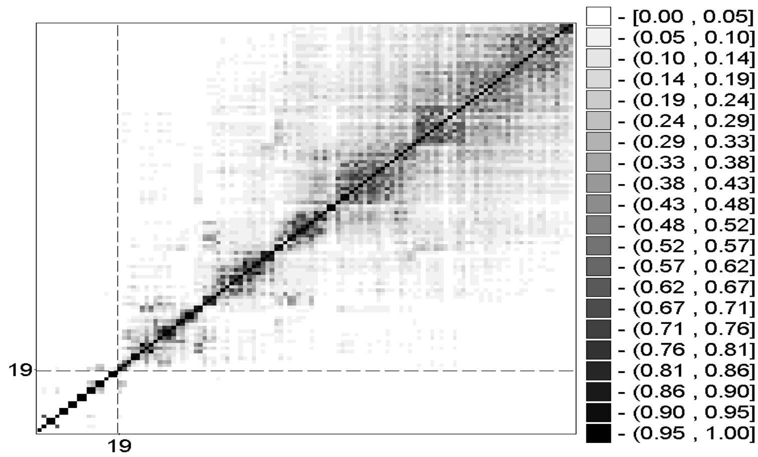

Figure 3: Matrix of w-correlations for the 120 reconstructed components.

The principal components (shown as time series) of the first 18 eigentriples are shown in Fig. 4. Consider a pure harmonic with a frequency w, certain phase, amplitude and the ideal situation where the period $P = 1/w$ is a divisor of both the window length L and

$K = T - L + 1$. In this ideal situation, the left eigenvectors and principal components have the form of sine and cosine sequences with the same period P and the same phase. Thus, the identification of the components that are generated by a harmonic is reduced to the determination of these pairs.

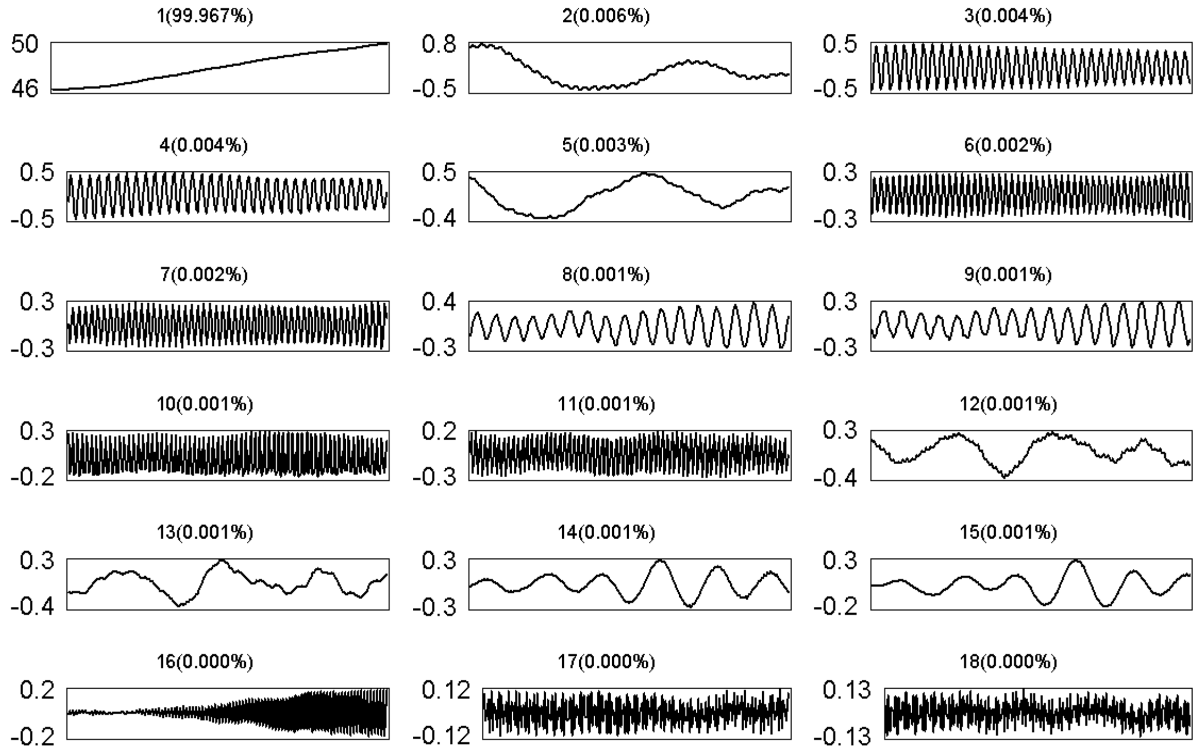

Figure 4: The first 18 principal components plotted as time series

Fig. 5 depicts the scatterplots of the paired principal components in the series, corresponding to the harmonics with periods 6, 4, 12, 3, 36 and 2.4 months. They are ordered by their contribution (share) in the SVD step (from left to right).

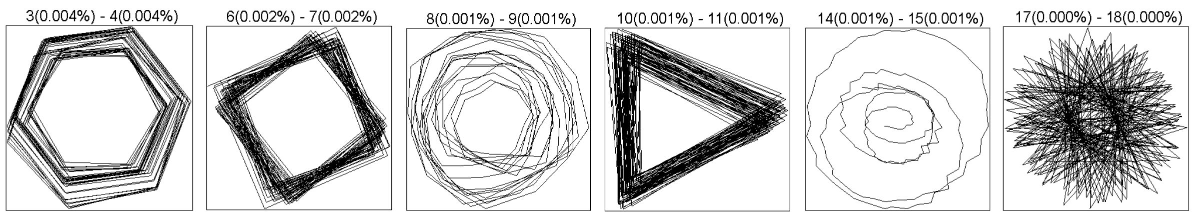

Figure 5: Scatterplots (with lines connecting consecutive points) corresponding to the paired harmonic principal components.

The periodograms of the paired eigentriples (3-4 , 6-7, 8-9, 10-11 and 17-18) also confirm that the eigentriples correspond to the periods of 6, 4, 12, 3, 36 and 2.4 months.

Identification of trend, harmonics and noise components

Trend is a slowly varying component of a time series which does not contain oscillatory components. Henceto capture the trend in the series, we should look for slowly varying eigenvectors. Fig. 6 (top) shows the extracted trend which is obtained from the eigentriples 1, 2, 5, and 12–13. It clearly follows the main tendency in the series.
Fig. 6 (middle) represents the selected harmonic components (3,4, 6–11, 14–18) and clearly shows the same pattern of seasonality as in the original series. Thus, we can classify the rest of the eigentriples components (19–120) as noise. Fig. 6 (bottom) shows the residuals which are obtained from these eigentriples. The w-correlation between the reconstructed series (the eigentriples 1-18) and the residuals (the eigentriples 19-120) is equal to 0.0006, which can be considered as a confirmation that this grouping is very reasonable. The p-value of Anderson-Darling test for testing normality is 0.6 suggesting that the residual series is close to the normal distribution.

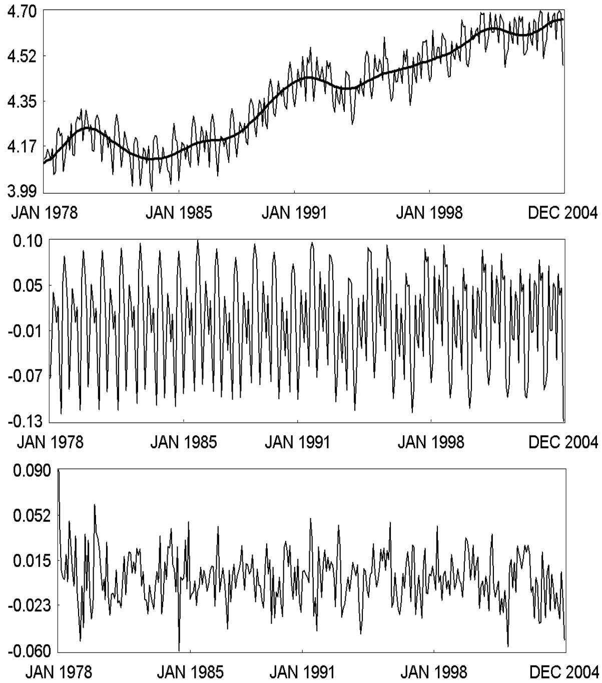

Figure 6: Reconstructed trend (top), harmonic (middle) and noise (bottom).

4.2 Biomedical data

Temporomandibular disorders (TMDs) occur as a result of problems with jaw, temporomandibular joint (TMJ), and surrounding facial muscles that control chewing and moving the jaw [18]. TMJ is the hinge joint that connects the lower jaw (mandible) to the temporal bone of the skull, which is immediately in front of the ear on each side of the head. The joints are flexible, allowing the jaw to move smoothly up and down and side to side. If you place your fingers just in front of your ears and open your mouth, you can feel the joints on each side of your head (illustrated in Fig. 7).

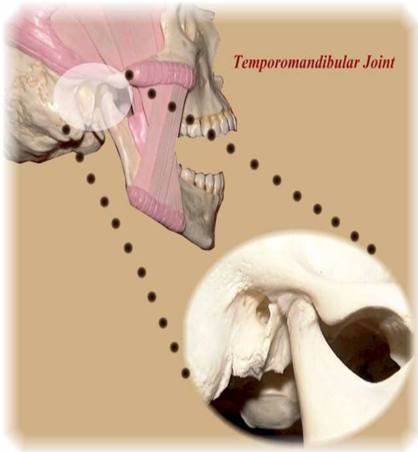

Figure 7: An illustration of the temporomandibular joint and its location adopted from [17].

Symptoms of TMD include headaches, tenderness of the chewing muscles, and clicking or locking of the joints [19]. More than 40% of the general population has at least one sign of TMD, yet only one in four of such people is actually aware of, or reports any symptoms [20]. One of the most popular areas of TMD research is developing clear guidelines for diagnosing these disorders. Automatic measurement and classification of TMDs before and during the treatment can assist in early diagnosis, accurate monitoring of treatment, and enhance the efficacy of the treatment.
Here we present an alternative method for detection of TMD based on visual analysis of facial movement. For this purpose we attach a number of markers to the points of interest on the individuals' faces and track their positions over a large number of frames in the video sequences. Afterwards, we analyse the motion patterns of the markers and extract their main signal using the SSA technique. The SSA technique has been used in [15] as a new method for detection of temporomandibular disorder. In this method the motion data of markers placed on the points of special interest on the faces of several subjects is extracted and analysed. The individuals are classified into a group of healthy subjects and a group of those with temporomandibular disorder by extracting the signal components of the original time series and separating the noise using the proposed technique.

Fig. 8 shows the noise series for an individual with TMD (left side) and a heathy subject (right side) after extracting the signal from the original series. Large peaks are clearly visible in Fig. 8 for the individual with TMD (for more information see [15]).

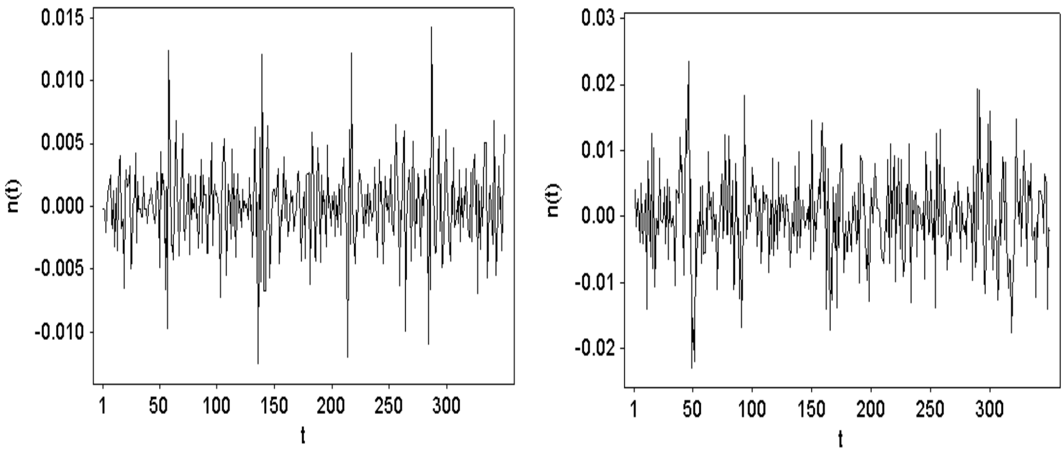

Figure 8: Noise series (eigentriples 6–40) for an individual with TMD (in the left) and that of a healthy individual (in the right).

The analysis of chewing velocity pattern suggested that opening and closing patterns with obvious peak velocity was significantly more frequent in patients with TMD than healthy subjects. Here the behavior of the peaks is dynamic, confirming that the series with this structure in the noise series relates to the individuals with TMD. Note that the peaks appear in Fig. 8 are the points in the TMD signal when the individual has a problem during closing of the mouth. Therefore, we conclude that the noise with such a structure is related to individual with TMD. In fact, our detection method is based on the separated noise, not the signal, and the peaks are important observations.

4.3 Image Processing

A methodology of image processing based on application of the Singular Spectrum Analysis is proposed and studied in [21]. The first stage of the procedure is the transformation of the data into another matrix which is a version of the trajectory matrix in Basic SSA. Similarly to the standard SSA, we need to define the window which will be moved over the image. Unlike the one-dimensional case, the window length has not only width, but height as well. The window is then placed at all possible positions in the image from left to right and top to bottom. The following figures show standard Lena image of size 128×128 pixels in a grey scale (coded with numbers from 0 to 255) with different noise levels. A reconstructed image is then shown in Fig. 10 (for more information see [21]).

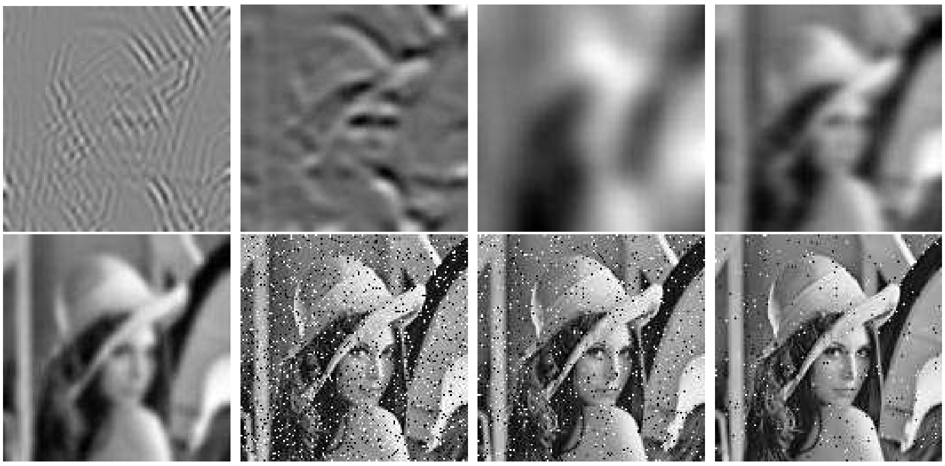

Figure 9: SSA reconstruction of the Lena image using different eigenvalues and different window sizes.

Figure 10: Lena image.

References

[1] Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*, Chapman & Hall/CRC, New York - London.

[2] Broomhead, D. S., and King, G. (1986). Extracting qualitative dynamics from experimental data, *Physica* D, **20**, pp. 217–236.

[3] Danilov, D., and Zhigljavsky, A. (1997) (Eds.). *Principal Components of Time Series: the 'Caterpillar' method*, University of St. Petersburg, St. Petersburg. (In Russian.).

[4] Elsner, J. B., and Tsonis, A. A. (1996). *Singular Spectrum Analysis, A New Tool in Time Series Analysis*, Plenum Press, New York and London.

[5] Hassani, H., and Zhigljavsky, A. (2009). Singular Spectrum Analysis: Methodology and Application to Economics Data, *Journal of System Science and Complexity*, **22**(3), pp. 372–394.

[6] Vautard, R., Yiou, P., and Ghil, M. (1992). Singular spectrum analysis: a toolkit for short, noisy and chaotic series, *Physica D*, **58**, pp. 95–126.

[7] Moskvina, V. G., and Zhigljavsky, A. (2003). An algorithm based on singular spectrum analysis for change-point detection, *Communication in Statistics - Simulation and Computation*, **32**(4), pp. 319–352.

[8] Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison, *Journal of Data Science*, **5**(2), pp. 239–257.

[9] Hassani, H., Heravi, S., and Zhigljavsky, A. (2009). Forecasting European Industrial Production with Singular Spectrum Analysis, *Internation Journal of Forecasting*, **25**(1), pp. 103–118.

[10] Patterson, K; Hassani, H; Heravi, S; Zhigljavsky, A. (2010). Forecasting the Final Vintage of the Index of Industrial Production, *Journal of Applied Statistics*, Forthcoming.

[11] Alexandrov, Th., and Golyandina, N. (2004). The automatic extraction of time series trend and periodical components with the help of the Caterpillar-SSA approach. *Exponenta Pro 3-4* (In Russian.), pp. 54–61.

[12] www.gistatgroup.com

[13] Hassani, H; Zhigljavsky, A; Patterson, K; Soofi, A. (2010). A Comprehensive Causality Test Based on the Singular Spectrum Analysis, *Causality in Science*, Oxford University press, Forthcoming.

[14] Hassani, H., Dionisio, A., and Ghodsi, M. (2010). The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets, *Nonlinear Analysis: Real World Applications*, **11**(1), pp. 492–502.

[15] Ghodsi, M., Hassani, H., Sanei, S., and Hick, Y. (2009). The use of noise information for detecting temporomandibular disorder, *Biomedical Signal Processing and Control*, **4**, pp. 79–85.

[16] Hassani, H., Zokaei, M., von Rosen, D., Amiri, S., and Ghodsi, M. (2009). Does Noise Reduction Matter for Curve Fitting in Growth Curve Models?, *Computer Methods and Program in Biomedicine*, **96**(3), pp. 173–181.

[17] http://www.daemen.edu/offices/grants/Images/TMJ.JPG

[18] Okeson, J. P. (1996). Temporomandibular disorders in the medical practice, *J. Family Practice*, **43**(4), pp. 347.

[19] Dworkin, S. F., and LeResche, L. (1992). Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique, *J. craniomandib. Disord. Facial and Oral Pain*, **6**(4), pp. 301–355.

[20] Okeson, J. P.(ed.). (1996). Orofacial pain. Guidelines for assessment, diagnosis, and management, Chicago, Quintessence Publishing Co, Inc, pp. 116–117.

[21] Rodrígez-Aragón, L. J. and Zhigljavsky, A. (2010). Image processing by means of the Singular Spectrum Analysis, *Statistics and its interface*, Forthcoming.