

Singular Spectrum Analysis for Forecasting of Electric Load Demand

Héctor Briceño^a, Claudio M. Rocco^{a*}, Enrico Zio^{b,c}

^aUniversidad Central de Venezuela, Apartado Postal 47937, Caracas, Venezuela

^bChair on Systems Science and the Energetic Challenge, European Foundation for New Energy-Electricité de France, Ecole Centrale Paris and Supélec, F-92 295 CHÂTENAY-MALABRY Cedex

^cDipartimento di Energia, Politecnico di Milano, Nuclear Section-Cesnef, via Ponzio 34/3, 20133 - Milano, Italy
croccouv@gmail.com

This paper presents the technique of Singular Spectrum Analysis (SSA) and its application for electric load forecasting purposes. SSA is a relatively new non-parametric data-driven technique developed to model non-linear and/or non-stationary, noisy time series. SSA is able to decompose the original time series into the sum of independent components, which represent the trend, oscillatory behavior (periodic or quasi-periodic components) and noise. One of the main advantages of SSA compared to other non-parametric approaches is that only two parameters are required to model the time series under analysis. An example of application is given, with regards to forecasting the monthly electric load demand in a Venezuelan region served by wind power generators. In this case, careful demand estimation is required since the wind generation output could be highly variable and additional conventional generation or transmission links would be required to satisfy the load demand. A comparison with other classical time-series approaches (like exponential smoothing and Auto Regressive Integrated Moving Average models) is presented. The results show that SSA is a powerful approach to model time series, capable of identifying sub-time series (trends along with seasonal periodic components).

1. Introduction

Power system planning and operation are important tasks that all electric utilities face. One of the factors that influence their success is load forecasting. Indeed, this determines the amount of generation needed to satisfy the electric load demand. On the short time range, this is accomplished by scheduling enough generation and transmission capacity (in real time) and by planning new generation allocation for the mid and long ranges. In the case of intermittent renewable generation sources, e.g. wind, the operator is faced not only with the load variation but also with that of the generation output.

In this paper, Singular Spectrum Analysis (SSA) (Elsner and Tsonis, 1996) is applied for forecasting electric power load demand. SSA is a technique developed independently in the USA and UK under the name SSA, and in Russia under the name Caterpillar-SSA (Beneki et al., 2009). It is based on statistics and probability theory, dynamical systems and signal processing concepts and it is considered as a non-parametric or data-driven technique (Beneki et al., 2009) for modeling non-linear, non-stationary and noisy short time series (Hassani, 2007). SSA does not rely on a priori defined functions like, for example, the Fourier approach (based on sine and cosine functions), or statistical assumptions (e.g., stationarity, normality). One of the main advantages of SSA compared to other non-parametric approaches is that only two parameters are required to model the time series under analysis.

The development of the technique and its application for electric load forecasting are illustrated by way of an example related to a Venezuelan region served by power generators.

The remainder of the paper is organized as follows: Section 2 introduces the SSA approach; Section 3 presents the results obtained using SSA and the comparison with other classical time series techniques; finally, Section 4 presents some conclusions.

2. Singular Spectrum Analysis (SSA)

The main idea of SSA is to make a decomposition of the original time series into the sum of independent components, which represent the trend, oscillatory behavior (periodic or quasi-periodic components) and noise. Some of these components are selected and retained for the forecasting task. SSA has been used successfully in several areas like hydrology (Vautard et al., 1992), geophysics (Vautard and Ghil, 1989), economics (Carvalho et al., 2012), medical engineering (Sanei et al., 2011) among others.

Let $\mathbf{y}=[y_1, y_2, \dots, y_T]$ be a time series of length T . The SSA technique consists of two stages: decomposition and reconstruction (Golyandina et al., 2001).

2.1 Decomposition

This stage is subdivided into two steps: Embedding and Singular Value Decomposition

2.1.1 Embedding

The main result of this step is the definition of a *trajectory matrix* or a lagged version of the original time series \mathbf{y} . The matrix is associated with a windows length L ($L \leq T/2$, to be defined by the user). Let $K=T-L+1$; the trajectory matrix is defined as:

$$X=[X_1, \dots, X_K] = \begin{pmatrix} y_1 & y_2 & y_3 & \dots & y_K \\ y_2 & y_3 & y_4 & \dots & y_{K+1} \\ y_3 & y_4 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ y_L & y_{L+1} & y_{L+2} & \dots & y_T \end{pmatrix} \quad (1)$$

The trajectory matrix is a Hankel matrix, that is all the elements along the diagonal $i+j=\text{const}$ are equal (Hassani, 2007).

2.1.2 Singular Value Decomposition (SVD)

From X , define the covariance matrix XX^t . The SVD of XX^t produces a set of L eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$ and the corresponding eigenvectors U_1, U_2, \dots, U_L (often denoted by Empirical Orthogonal Functions (EOF)). Then the SVD of the trajectory matrix can be written as $X = E_1 + E_2 + \dots + E_d$, where $E_i = \sqrt{\lambda_i} U_i V_i^t$, d is the rank of XX^t (that is the number of non-zero eigenvalues) and V_1, V_2, \dots, V_d are the Principal Components (PC), defined as $V_i = X^t U_i / \sqrt{\lambda_i}$. The collection $(\sqrt{\lambda_i}, U_i, V_i)$ is referred to as the i -th

eigentriple of the matrix X . If $\Gamma = \sum_{i=1}^d \lambda_i$, then λ_i/Γ is the proportion of the variance of X explained by E_i : E_1

has the highest contribution while E_d has the lowest contribution (Hassani et al., 2009). SVD could be time consuming if the length of the time series is large (say $T > 1000$).

2.2 Reconstruction

The reconstruction stage consists of two steps: Grouping and Averaging.

2.2.1 Grouping

In this step, r out of d eigentriples are selected by the user. Let $I=\{i_1, i_2, \dots, i_r\}$ be a group of r selected eigentriples and $X_i = X_{i_1} + X_{i_2} + \dots + X_{i_r}$. X_i is related to the "signal" of \mathbf{y} while the rest of the $(d-r)$ eigentriples describe the error term ϵ .

2.2.2 Averaging

The group of r components selected in the previous stage is then used to reconstruct the deterministic components of the time series. The basic idea is to transform each of the terms $X_{i_1}, X_{i_2}, \dots, X_{i_r}$ into reconstructed time series $\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \dots, \mathbf{y}_{i_r}$ through the Hankelization process $\mathbf{H}()$ or diagonal averaging (Golyandina et al., 2001): if z_{ij} is an element of a generic matrix Z , then the k -th term of the reconstructed time series can be obtained by averaging z_{ij} , if $i+j=k+1$. That means that $\mathbf{H}(Z)$ is a time series of length T reconstructed from matrix Z . At the end of the averaging step, the reconstructed time series is an approximation of \mathbf{y} :

$$\mathbf{y} = \mathbf{H}(X_{i_1}) + \mathbf{H}(X_{i_2}) + \dots + \mathbf{H}(X_{i_r}) + \epsilon \quad (2)$$

As pointed by Alexdradov and Golyandina (2005), the reconstruction of a single eigentriple is based on the whole time series. This means that SSA is not a local method and hence robust to outliers.

2.3 Forecasting

The SSA forecasting approach uses the notion of Linear Recurrent Formulae (LRF) defined as:

$$y_{T-j} = \sum_{k=1}^{L-1} a_k y_{T-i-k}, \quad 0 \leq i \leq T-L \text{ and } a_k \neq 0.$$

Broad classes of continuous series are governed by LRF, including harmonics, polynomial and exponential series. The coefficients a_k are determined from the eigenvectors obtained from the SVD stage (see Golyandina et al., 2001 for details).

2.4 Parameter selection

As previously mentioned only two parameters are required in SSA: the window length L and the number of components r . A detailed description of parameter selection is presented in Golyandina et al. (2001). Values for L and r could be defined using information provided by the time series under study or through additional indices:

- **The window length L :** The general rule is to select $L = T/2$. However if the user knows that the time series may include an integer-period component s , then a better *separability* of components could be obtained by defining L proportional to that period ($L=(T-js)/2$, $j=1, 2, \dots$)
- **The number of components r :** The theory of *separability*, that is how well the components can be separated, is the basis for the definition of r . A general criterion is based on the contribution of each

component to the variance of X , evaluated as λ_i/Γ ($\Gamma = \sum_{i=1}^d \lambda_i$): Select r out of d components so that

the sum of their contributions is at least a predefined threshold, for example $\geq 90\%$. In general, noise components have low contribution.

- An index, defined as the weighted correlation *w-correlation* (Hassani, 2007), could also be used to quantify the degree of *separability* among two reconstructed components. Let X_A and X_B be two reconstructed time series: if *w-correlation*(X_A, X_B) is zero then the two series are separable. On the other side, if *w-correlation*(X_A, X_B) is high, then the components could be grouped.

Finally, pairwise scatterplots of eigenvectors are used to detect harmonic components, whose frequencies can be estimated using a periodogram. In general, components with high frequency are associated to noise components. Currently, there are quasi-automatic algorithms to help the user in the determination of r (e.g., AutoSSA (Alexandrov and Golyandina, 2005) or Rssa (Korobeynikov, 2012)).

3. Case study

SSA is used to decompose and forecast the monthly electric load demand in a Venezuelan region served by wind power generators. The time series (in MW) is composed by 240 data. As a usual procedure, the corresponding time series is divided into two sets: the first set (*training data set*), defined by the first $T=228$ observations, is used to “tune” the model, while the second set (*testing data set*) is used to assess the performance of the model. Since data points are recorded monthly, at least a 12-period component is assumed ($s=12$). Results from SSA are compared with other classical time-series approaches (like exponential smoothing and Auto Regressive Integrated Moving Average models, as Briceño (2012)). SSA evaluations are performed using the Rssa procedure in R (Korobeynikov, 2012). For each technique, the following performance indexes related to the residuals are presented: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Box-Pierce statistic Q .

3.1 SSA evaluation

3.1.1: Decomposition stage: Embedding and SVD Decomposition

Figure 1 shows the contribution of each eigenvalue after the SVD stage, using $L = (T-12)/2 = 108$ as the window length required for the embedding step. Note that the first eight eigenvectors account for almost 95% of the variance of the time series.

3.1.2: Reconstruction stage: Grouping and Averaging

Figure 2 shows the plot of the reconstructed components (RC). At a first glance, RC1 and RC2 are linked to slow moving components (i.e., the trend) while the rest of RC are associated to oscillating components. Figure 3 shows the plot associated to pairs of eigenvectors. The geometric figures associated to eigenvectors 5 and 6 and 7 and 8 show the existence of frequencies related to six and twelve months. The

rest of the plots do not suggest any known geometric figure. The raw periodogram of the time series under study shows two significant components at $1/12$ and $2/12=1/6$, indicating the existence of monthly and six-month components, as previously detected. This fact could suggest the use of a new integer L value, proportional to both components detected (e.g., $L=96, 84, \dots$).

In order to evaluate the *separability* of the different eigentriples, Figure 6 shows a graphical representation of the *w-correlation* matrix. Each cell (F_i, F_j) represents the correlation between components i and j , coded from black (*w-correlation* = 1) to white (*w-correlation* = 0). Note that component 1 is clearly separable and components 2-4 and 9-10 are almost separable (*w-correlation* values < 0.50). Components RC5-RC6 and RC7-RC8 have *w-correlations* equals to 1 and, as previously mentioned, could be grouped. Finally, the reconstruction will be based on eigentriples 1 to 10, as shown in Figure 4, along with the forecast for the next 12 months.

3.1.3: Comparison with other approaches

Table 1 shows the comparison of residual indexes between the SSA and two classical approaches, during the training phase. The additive Holt-Winter was the best method selected among exponential smoothing techniques while the best SARIMA model found was a $(2, 1, 0)(2, 0, 2)_{12}$:

$$(1 + 0,28*B + 0,17*B^2)(1 - 0,48*B^{12} - 0,5*B^{24})\nabla X_t = (1 - 0,26*B^{12} + 0,38*B^{24})a_t \quad (3)$$

The SARIMA and SSA residuals are considered as random. Note that SSA performance is better than the other approaches. SSA is able to produce additional information about the decomposition of the time series, especially trend and harmonic components. Table 1 also shows the MSE and MAPE indexes during the testing phase. Again, SSA produce the best results.

4. Conclusions

This paper presented the use of Singular Spectrum Analysis as a technique able to decompose and forecast a non-stationary and/or non-linear time series into a set of independent components. In the example presented, related to the electric load demand of a Venezuelan region, the performance of SSA was the best, compared with other classical time series approaches. SSA relies only on the definition of two parameters: the windows length L and the number of components r . As a general rule, L could be fixed as $L=(T-s)/2$ while r could be determined automatically from the values of the *W*-correlation matrix. Briceño (2012) shows that the best SSA model is obtained using $L=84$, with performance indexes slightly better than the SSA model presented using $L=118$.

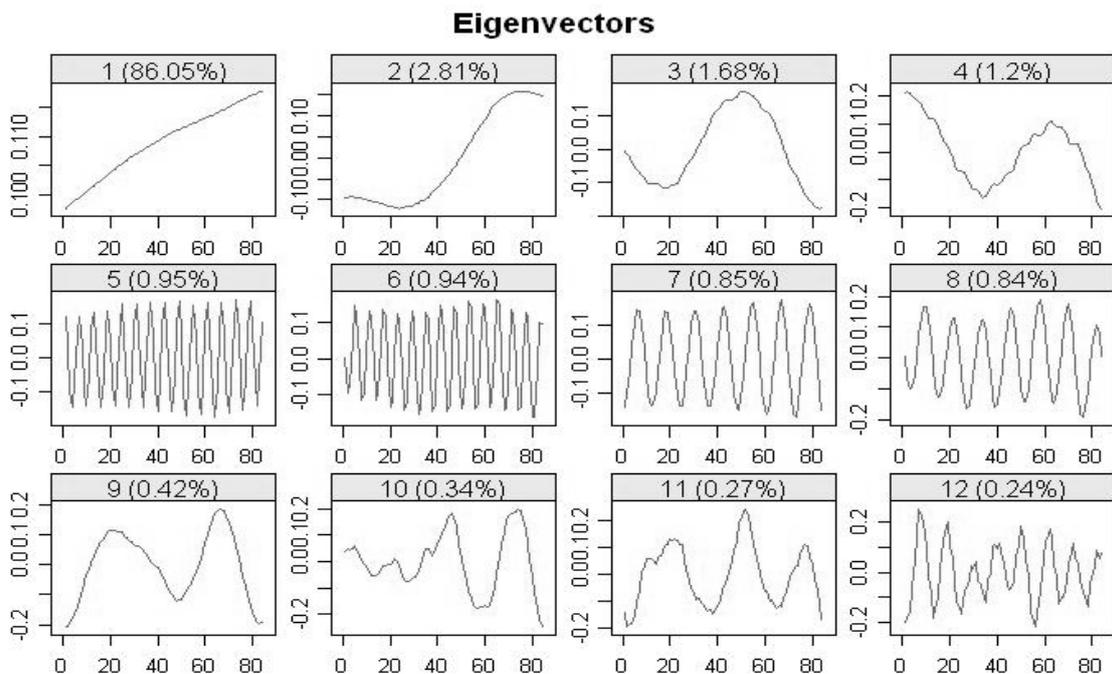


Figure 1 Eigenvector decomposition

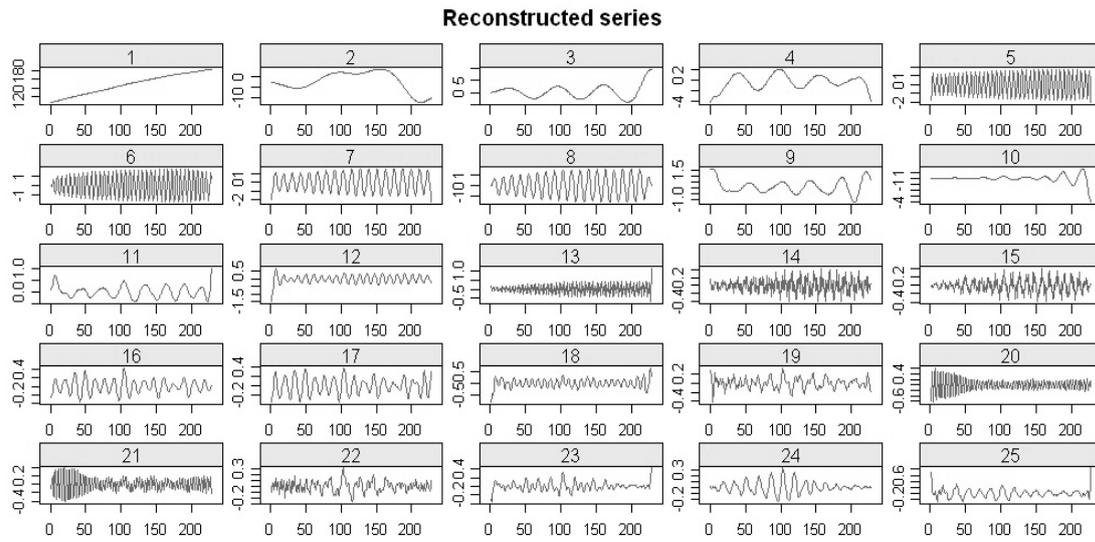


Figure 2: Reconstructed components related to the first 25 eigentriples

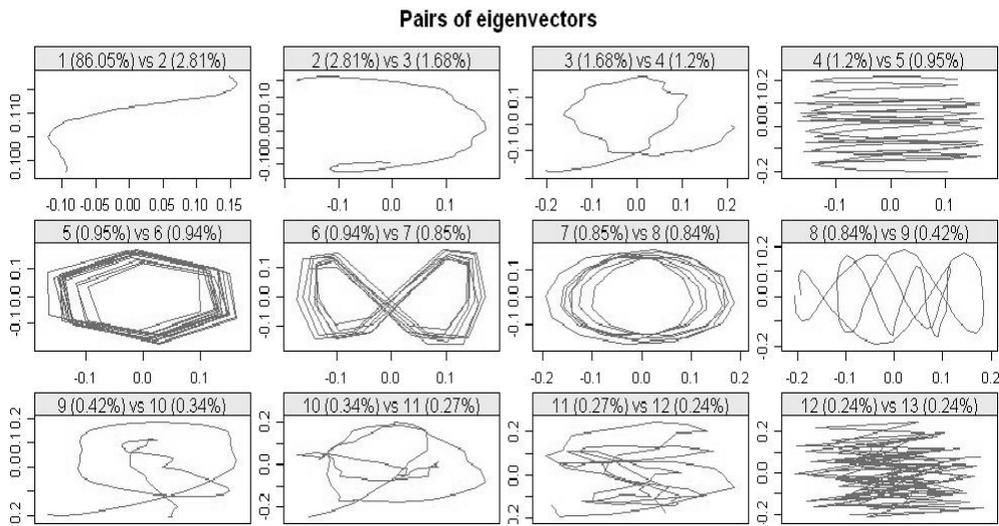


Figure 3: Pair of eigenvectors plot

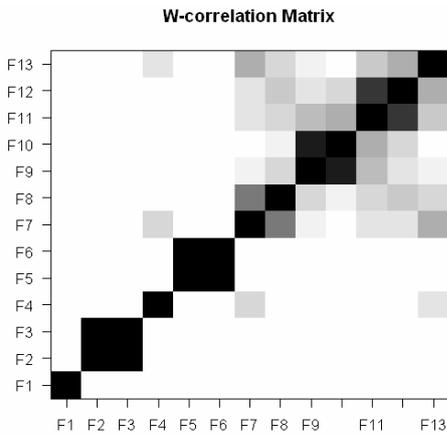


Figure 4: A graphical representation of the W-correlation matrix

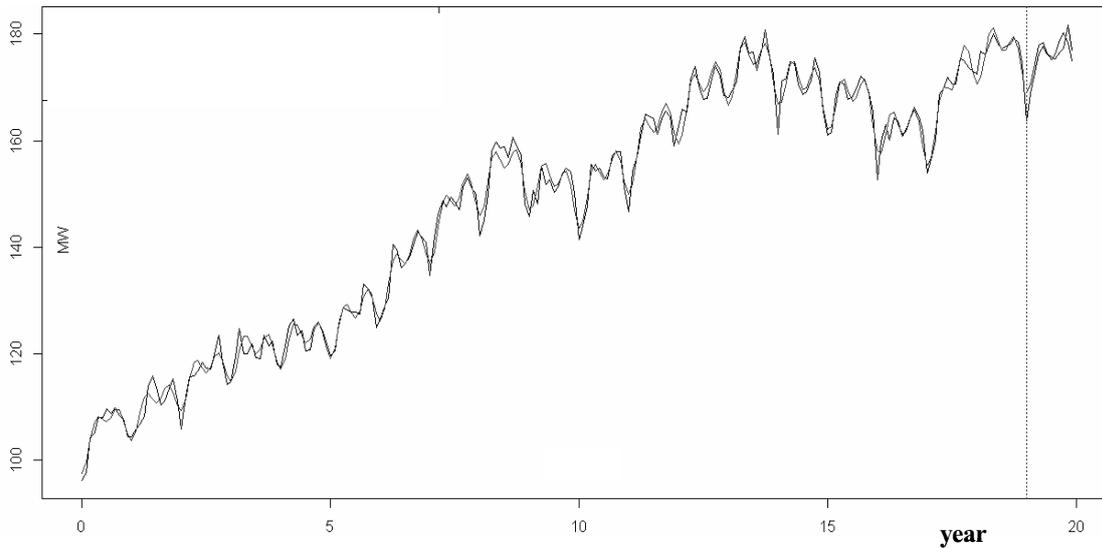


Figure 5: Reconstruction using eigentriples 1 to 10

Table 1: Residual indexes comparison

Method	Training Phase					Testing Phase	
	MSE	MAPE (%)	Theil	Box-Pierce	Randomness	MSE	MAPE (%)
AdDitive Holt-Winter	5.5	1.25	0.80	100.4	No	24.9	2.68
ARIMA (2,1,0)(2,0,2) ₁₂	6.0	1.15	0.45	24.3	Yes	12.6	1.86
SSA	3.6	1.06	0.26	23.1	Yes	7.0	1.26

References

- Alexandrov Th., Golyandina N., 2005, Automatic extraction and forecast of time series cyclic components within the framework of SSA, Proceedings of the Fifth Workshop on Simulation, June 26-July 2, 2005, St. Petersburg State University, St. Petersburg, 45–50.
- Beneki C., Eeckels B., Leon C., 2009, Signal Extraction and Forecasting of the UK Tourism Income Time Series. A Singular Spectrum Analysis Approach, <mpra.ub.uni-muenchen.de/18354/>, MPRA Paper No. 18354, accessed 13.01.2013.
- Briceño H., 2012, Comparing time series techniques for electric load demand, MSc. Thesis, Universidad Central de Venezuela (in Spanish), <www.ing.ucv.ve/Basica.html>, accessed 10.03.2013
- Carvalho M., Rodrigues P.C., Rua A., 2012, Tracking the US business cycle with a singular spectrum analysis, Economics Letters 114, 32–35.
- Elsner, J. B. and Tsonis, A. A., 1996, Singular Spectrum Analysis. A New Tool in Time Series Analysis, Plenum Press, New York, London.
- Golyandina N., Nekrutkin V., Zhigljavsky A., 2001, Analysis of Time Series Structure—SSA and Related Techniques, Chapman & Hall/ CRC, Boca Raton, FL, USA,
- Hassani H, Heravi S, Zhigljavsky A., 2009, Forecasting European industrial production with singular spectrum analysis, International Journal of Forecasting 25 103–118.
- Hassani H., 2007, Singular Spectrum Analysis: Methodology and Comparison, Journal of Data Science 5, 239-257
- Korobeynikov A., 2012, Rssa package, <cran.r-project.org/web/packages/Rssa/index.html>, accessed 12.12.2012
- Sanei S, Ghodsi M., Hassani H., 2011, An adaptive singular spectrum analysis approach to murmur detection from heart sounds, Medical Engineering & Physics, 33, 362–367.
- Vautard R., Ghil M., 1989, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Physica D, Nonlinear Phenomena, 35, 395–424.
- Vautard R., Yiou P., Ghil M., 1992, Singular spectrum analysis: a toolkit for short noisy chaotic signals, Physica D, Nonlinear Phenomena, 58, 95–126.