Windowed SSA (Singular Spectral Analysis) for Geophysical Time Series Analysis

Rajesh Rekapalli and Ram Krishna Tiwari
Theoretical Modeling Group, CSIR-National Geophysical Research Institute, Hyderabad 500007, India

Abstract: Although the SSA (singular spectral analysis) is a potential tool for analysing time series of different physical processes, the processing of large geophysical data set requires more time and is found to be computationally expansive. In particular for the SVD (singular value decomposition) of large trajectory matrix, the processing units require huge memory and high performance computing system. In the present work, we propose an alternative scheme based on WSSA (windowed singular spectral analysis), which is robust for analysing long data sets without losing any valuable low-frequency information contained in the data. The underlying scheme reduces the floating point operations in SVD computations as the size of the trajectory matrix is small in windowed processing. In order to test the efficiency, the authors applied the proposed method on two geophysical data sets i.e., the climatic record with 30,000 data points and seismic reflection trace with 8,000 data points. The authors have shown that without distorting any physical information, the low-frequency contents of the data are well preserved after the windowed processing in both the cases.

Key words: Singular value decomposition, singular spectral analysis, trajectory matrix.

1. Introduction

The time-series analysis has been widely used to detect hidden fundamental pattern and basic properties of the physical processes that engender it. To analyse the periodicities of any earth system processes with high resolution, it is essential to process the data with available sampling interval of long data sets. SSA (singular spectral analysis) [1–5] is one of the most popular and effective methods for analysing evolutionary time series structures and inherited dynamics. The applicability of SSA extends over a wide range time series from astronomical records to seismic data sets. As it is designed to extract information from short and noisy time series, it provides insight into the unknown or partially known dynamics of the underlying systems that generated the series [1]. The method has been extensively used in trend extraction [6], principle component analysis [7, 8], reconstruction of missing data in series [9–11], and signal extraction from complex noise [12] of different types of atmospheric and geophysical signals. The quantitative interpretation of SSA results in terms of attractor dimensions has been given by number of authors [1, 13]. Recently, several researchers have applied the SSA method to various fields like signal reconstruction and forecasting of time series [14], improving Nakamura technique [15], eliminating the random noise in the seismic data using FX Singular Spectrum Analysis [16] and multichannel SSA [17] and for filtering the geophysical fields and astronomical images [18], etc.

The computation time and cost are two crucial parameters that decide the success and wider applicability of the data processing methods. Even though the SSA method is well proven for variety of applications in signal analysis, its applicability for large data is limited due to the large time and memory requirements for the SVD (singular value decomposition) of Hankel matrix (Trajectory matrix). The SVD scheme is the main computational process of SSA and for a $n \times m$ matrix, it requires $O(nm^2)$ floating point operations [19]. As the length of the data record increases, the authors are bound to calculate SVD for
large trajectory data matrix. For example, the processing of seismic reflection/refraction data using SSA method consumes more time and requires huge memory as the size of the data (seismic trace) becomes high. Consequently, this limits the applicability of SSA for the long data sets. In the present work, the authors introduce a skilled algorithm “WSSA (windowed singular spectral analysis)”, in which the entire dataset of seismic trace is divided into short windows and each part is independently processed using SSA and finally, the outputs obtained from different windows are merged together to get the SSA processed time series of original length. The high variance Eigen components representing the trends are also considered in the reconstruction as they represent the part of signal with high periodicities. Thus, the application of WSSA method reduces the computational cost and time as the size of the trajectory matrix formed using windowed data is small.

2. Methodology

The linear trend in the data must be removed before proceeding to WSSA application. The computational steps of WSSA algorithm are almost similar to that of singular spectral analysis [19, 20]. The crucial steps involved in WSSA methodology are briefly mentioned below:

Step (i): divide the time series \( D (t) = \{ y_1, y_2, ..., y_N \} \) of length \( N \) into \( d \) equal windows and apply the following steps on each window.

Step (ii): embed the phase space portrait or trajectory matrix \( T \) of the data using a proper window length \( L \). While dealing the data of unknown periodicity (i.e., if the authors do not have any prior knowledge on the periodicity of the data), then the authors can choose half of the lag difference of successive same phased point as the window length from the auto correlation plot. Thus, the authors engender the trajectory matrix of size \( L \cdot K \) given by:

\[
T_{L \cdot K} = [Y_1: \ldots : Y_K]
\]  

(1)

where, \( Y_i = \{ y_{i1}, y_{i+1}, ..., y_{i+L} \} \) indicates the vector of length \( L \) and \( K = (N/d) - L + 1 \) and \( Y_{N/d} (t) = \{ y_{1}, y_{2}, ..., y_{N/d} \} \) is the windowed series of length \( N/d \).

Step (iii): the authors apply SVD to decompose the trajectory matrix \( T \) into eigentriplet \( (\sqrt{\varepsilon_i}, U_i, V_i) \) in order to examine features of data:

\[
T = \sum_{i=1}^{d} \sqrt{\varepsilon_i} U_i V_i^T
\]

(2)

Here \( \sqrt{\varepsilon_i}, U_i, V_i \) are eigenvalues, left and right eigenvectors of \( i^{th} \) eigentriplet.

Step (iv): in this step, the \( d \) triplets correspond to non-zero eigenvalues are subdivided into groups based on the periodicity of eigenvectors and eigenvalues. The trajectory matrix reconstructed using the selected triplet groups. As the data are processed in short windows, the first few triplets (those corresponds to log term trend and cycles) must be included in reconstruction processes along with the other higher order triplets to retain the low frequency content of the data:

\[
T_r = \sum_{i=a,b,c} \sqrt{\varepsilon_i} U_i V_i^T
\]

(3)

\[
T_r = \begin{bmatrix}
Y(1,1) & \ldots & Y(1,K) \\
\vdots & \ddots & \vdots \\
Y(L,1) & \ldots & Y(L,K)
\end{bmatrix}
\]

(4)

where, \( a, b, c \) are the triplets that are in the selected groups.

Step (v): the reconstructed trajectory matrix \( T_r \) obtained in the above process will be diagonally averaged to get the reconstructed series of length \( N/d \).

Let us denote the reconstructed series by \( Y_r (t) = \{ g_1, g_2, ..., g_k, ..., g_N \} \). The elements of \( Y_r \) computed from the diagonal averaging of reconstructed trajectory matrix as follows.

The authors set \( L^* = \min (L, K^*) = \max (L, K) \) and let \( y^{*}_{ij} = y_{ij} \) if \( L < K \) and \( y^{*}_{ij} = y_{ij} \) otherwise:

\[
g_k = \begin{cases} 
\frac{1}{k+1} \sum_{m=1}^{k+1} y^{*}_{m, k-m+2} & \text{for } 1 \leq k < L^* \\
\frac{1}{L^*} \sum_{m=k+1}^{L^*} y^{*}_{m, k-m+2} & \text{for } L^* \leq k < K^* \\
\frac{1}{N-k} \sum_{m=k-K^*+1}^{N-k-1} y^{*}_{m, k-m+2} & \text{for } K^* \leq k < N 
\end{cases}
\]

(5)
and $g_1 = y_{(1, 1)}$.

Step (vi): the WSSA processed series of original length ($N$) obtained by merging all the $d$ reconstructed outputs in the final step.

The WSSA reduces the floating point operation in SVD computation of $n \times m$ trajectory matrix from $O(nm^2)$ to $d \times O(n_1m_1^2)$, where $d$ is the number of windows and $n_1 \times m_1$ is the size of the trajectory matrix in each window.

3. Applications

The authors present here two examples from diverse data sources to demonstrate the robustness of the method.

Example 1:

In the first example, the method was applied on three-million-year long (3 million Yr BP-present) mutually consistent records of surface air temperature [21]. According to Ref. [21], the surface air temperature data was extracted from marine benthic oxygen isotopes sampled at 100 year interval using comprehensive ice-sheet model and simple ocean-temperature models. For the present analysis, the total data were divided into seven windows of each length 4,096 samples. The SSA method was then applied on each window to produce reconstructed series. The minimum window length for the trajectory formulation was obtained greater than or equal to 200 samples by plotting the autocorrelation (Fig. 1) of the data as explained in the methodology. The data have been reconstructed by retaining only the first six significant eigentriplets that contain trends and low frequency components which are also of our interest to demonstrate that the underlying method is efficient in preserving the low frequency content. After processing the data using short windows, the authors have reproduced the reconstructed data series of original length by simply combining the windowed data outputs obtained. Fig. 2 shows the original and reconstructed data sets. Fig. 3 shows the power spectral densities of original and reconstructed data series of length 2.8672 million years. It can be observed from the power spectrum of original and reconstructed data sets (Fig. 3), that the low frequency content is preserved in the data even after the windowed processing. As the authors want to suppress the high frequency components, the authors have neglected the higher order triplets of low variance. The efficiency of the time domain filtering using SSA can be observed from the power spectrum (Fig. 3) which shows flattened high frequency component in the processed data.

Example 2:

In the next example, the method has been applied to a single seismic trace of reflection data containing 8,000 samples with 0.25 ms sampling interval. Total data has been divided in to four equal parts with 2,000

![Sample autocorrelation function](image)

Fig. 1 Autocorrelation of surface air temperature data used to calculate the optimal window length ($L$) for SSA.
samples and each part is processed through WSSA using a window length 500 samples. The data reconstructed using the first 10 eigentriplets by viewing the eigenvectors periodicities to filter high frequency content with frequency above 85 Hz. The combined output of all the four parts is shown in Fig. 4 along with original trace. One can see that the structure of the trace is almost the same after the processing. The power spectral densities of original trace and its windowed SSA output are shown in Fig. 5. Clearly, one can observe from the power spectrums of seismic trace the components with frequency 85 Hz are completely suppressed in the WSSA output. As the main object of the paper is of the low frequency preserving characteristics of WSSA, the authors have not filtered the low frequency content of the trace. In order to perform the high frequency filtering, the authors have omitted the triplets corresponding to that
low periods in the reconstruction step. However, the major low frequency spectral components observed in the original data are preserved in the processed data. The results suggest that the WSSA is a robust and fast technique and it does not alter the low frequency content of the data series.

4. Conclusion

The analysis of low frequency content of a time series using the conventional spectral method depends
solely on the length of the time series. Consequently, for the large data series, the computational cost and time considerably increases in SSA processing. Here, the authors have proposed a windowed SSA to lessen the computational cost and time without missing any low frequency components of the data. The authors have applied the method on two different data sets representing the long historical climate time series and seismic reflection trace to verify efficiency of the proposed method for the preserving the low frequency contents. The authors have demonstrated that the proposed WSSA scheme effectively reduces the floating point operation in SVD computation compared to classical SSA. Also unlike the conventional spectral methods, the size of the window does not affect the long period/low frequency components of the data series.

Acknowledgements

The authors thank director, CSIR-NGRI for his permission to publish this paper. We also thank Prof. Mrinal K. Sen for his encouragement. The first author is grateful to CSIR for SRF funding, AcSIR and Faculty members of AcSIR.

References

[1] Vautard, R., and Ghil, M. 1989. “Singular Spectrum Analysis in Nonlinear Dynamics, with Applications to Paleoclimatic Time Series.” *Physica D: Nonlinear Phenomena* 35 (3): 395–424.

[2] Vautard, R., Mo, K. C., and Ghil, M. 1990. “Statistical Significance Test for Transition Matrices of Atmospheric Markov Chains.” *J. Atmos. Sci.* 47: 1926–31.

[3] Vautard, R., Yiou, P., and Ghil, M. 1992. “Singular-Spectrum Analysis: A Toolkit for Short, Noisy Chaotic Signals.” *Physica D: Nonlinear Phenomena* 58 (1–4): 95–126.

[4] Jiang, N., Ghil, M., and Neelin, D. 1996. “Forecasts of Niño-3 SST Anomalies and SOI Based on Singular Spectrum Analysis Combined with the Maximum Entropy Method.” *Experimental Long-Lead Forecast Bulletin* 5: 2–4.

[5] Ghil, M., and Yiou, P. 1996. “Spectral Methods: What They Can and Cannot Do for Climatic Time Series.” *Springer Berlin Heidelberg* 44: 446–82.

[6] Alexandrov, T. 2009. “A Method of Trend Extraction Using Singular Spectrum Analysis.” *REVSTAT—Statistical Journal* 7 (1): 1–22.

[7] Serita, A., Hattori, K., Yoshino, C., Hayakawa, M., and Isezaki, N. 2005. “Principal Component Analysis and Singular Spectrum Analysis of ULF Geomagnetic Data Associated with Earthquakes.” *Natural Hazards and Earth System Sciences* 5 (January): 685–9.

[8] Tiwari, R. K., and Rajesh, R. 2014. “Imprint of Long-Term Solar Signal in Groundwater Recharge Fluctuation Rates from Northwest China.” *Geophys. Res. Lett.* 41 (9): 3103–9.

[9] Kondrashov, D., Shprits, Y., and Ghil, M. 2010. “Gap Filling of Solar Wind Data by Singular Spectrum Analysis.” *Geophysical Research Letters* 37 (15). doi: 10.1029/2010GL044138.

[10] Kondrashov, D., and Ghil, M. 2006. “Spatio-Temporal Filling of Missing Points in Geophysical Data Sets.” *Nonlin. Processes Geophys* 13: 151–9.

[11] Schoellhamer, D. H. 2001. “Singular Spectrum Analysis for Time Series with Missing Data.” *Geophysical Research Letters* 28 (16): 3187–90. doi: 10.1029/2000GL012698.

[12] Varadi, F., Pap, J. M., Ulrich, R. K., Bertello, L., and Henney, C. J. 1999. “Searching for Signal in Noise by Random Lag Singular Spectrum Analysis.” *Astrophys. J.* 526 (2): 1052–61.

[13] Palus, M., and Dvorak, I. 1992. “Singular Value Decomposition in Attractor Reconstruction: Pitfalls and Precautions.” *Physica D: Nonlinear Phenomena* 55 (February): 221–34.

[14] Golyandina, N., and Stepanov, D. 2005. “SSA-Based Approaches to Analysis and Forecast of Multidimensional Time Series.” In *Proceedings of the 5th Workshop on Simulation*, 293–8.

[15] Canales, L. L. 1984. “Random Noise Reduction.” In *Proceedings of the 54th Ann. Internat. Mtg., Soc. Expl. Geophys.*, 525–7.

[16] Sacchi, M. D. 2009. “FX Singular Spectrum Analysis.” In *Proceedings of the CSPG CSEG CWLS Convention*, 392–5.

[17] Vicente, O., and Mauricio, S. 2011. “Simultaneous Seismic Data De-noising and Reconstruction via Multichannel Singular Spectral Analysis.” *Geophysics* 7 (3): V25–32.

[18] Zotov, L. V. 2012. “Application of Multichannel Singular Spectrum Analysis to Geophysical Fields and Astronomical Images.” *Advances in Astronomy and Space Physics* 2: 82–4.

[19] Rokhlin, V., Szlam, A., and Tygert, M. 2009. “A Randomized Algorithm for Principle Component Analysis.” *SIAM Journal on Matrix Analysis and
References

[20] Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P. 2002. “Advanced Spectral Methods for Climatic Time Series.” *Rev. Geophys.* 40 (1): 1–41. doi: 10.1029/2001RG000092.

[21] Bintanja, R., and van de Wal, R. S. W. 2008. “North American Ice-Sheet Dynamics and the Onset of 100,000-Year Glacial Cycles.” *Nature* 454 (August): 869–72. doi: 10.1038/nature07158.