Nonlin. Processes Geophys. Discuss., 1, 1947–1966, 2014
www.nonlin-processes-geophys-discuss.net/1/1947/2014/
doi:10.5194/npgd-1-1947-2014

Nonlinear Processes
in Geophysics
Discussions
Open Access

Improved singular spectrum analysis for time series with missing data

Y. Shen[1], **F. Peng**[1,2], **and B. Li**[1]

[1]College of Surveying and Geo-informatics, Tongji University, Shanghai, China
[2]Center for Spatial Information Science and Sustainable Development, Shanghai, China

Received: 12 October 2014 – Accepted: 2 December 2014 – Published: 21 December 2014

Correspondence to: Y. Shen (yzshen@tongji.edu.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union & the American Geophysical Union.

Abstract

Singular spectrum analysis (SSA) is a powerful technique for time series analysis. Based on the property that the original time series can be reproduced from its principal components, this contribution will develop an improved SSA (ISSA) for processing the incomplete time series and the modified SSA (SSAM) of Schoellhamer (2001) is its special case. The approach was evaluated with the synthetic and real incomplete time series data of suspended-sediment concentration from San Francisco Bay. The result from the synthetic time series with missing data shows that the relative errors of the principal components reconstructed by ISSA are much smaller than those reconstructed by SSAM. Moreover, when the percentage of the missing data over the whole time series reaches 60 %, the improvements of relative errors are up to 19.64, 41.34, 23.27 and 50.30 % for the first four principal components, respectively. Besides, both the mean absolute errors and mean root mean squared errors of the reconstructed time series by ISSA are also much smaller than those by SSAM. The respective improvements are 34.45 and 33.91 % when the missing data accounts for 60 %. The results from real incomplete time series also show that the SD derived by ISSA is 12.27 mg L^{-1}, smaller than 13.48 mg L^{-1} derived by SSAM.

1 Introduction

Singular spectrum analysis (SSA) introduced by Broomhead and King (1986) for studying dynamical systems is a powerful toolkit for extracting short, noisy and chaotic signals (Vautard et al., 1992). SSA first transfers a time series into trajectory matrix, and carries out the principal component analysis to pick out the dominant components of the trajectory matrix. Based on these dominant components, the time series is reconstructed. Therefore the reconstructed time series improves the signal to noise ratio and reveals the characteristics of the original time series. SSA has been widely used in geosciences to analyze a wide variety of time series, such as the stream flow and sea-surface temperature (Robertson and Mechoso, 1998; Kondrashov and Ghil, 2006), the seismic tomography (Oropeza and Sacchi, 2011) and the monthly gravity field (Zotova and Shum, 2010). Schoellhamer (2001) developed a modified SSA for time series with missing data (SSAM), which has been successfully applied to analyze the time series of suspended-sediment concentration (SSC) in San Francisco Bay (Schoellhamer, 2002). This SSAM approach does not need to fill missing data. Instead, it computes the each principal component (PC) with observed data and a scale factor related to the number of missing data. Shen et al. (2014) developed a new principal component analysis approach for extracting common mode errors from the GNSS time series of a regional station network with missing data. The other kind of SSA approaches process the time series with missing data by filling the data gaps recursively or iteratively, such as the "Catterpillar"-SSA method (Golyandina and Osipov, 2007), the imputation method (Rodrigues and Carvalho, 2013) or the iterative method (Kondrashov and Ghil, 2006).

This paper is motivated by Schoellhamer (2001) and Shen et al. (2014), and will develop an improved SSA (ISSA) approach which involves that of Schoellhamer (2001) as a special case. The rest of this paper is organized as follows: the improvement of SSA for time series with missing data will be followed in Sect. 2, synthetic and real numerical examples are presented in Sects. 3 and 4 respectively, and then conclusions are given in last Sect. 5.

2 Improved singular spectrum analysis for time series with missing data

For a stationary time series x_i $(1 \le i \le N)$, we can construct an $L \times (N-L+1)$ trajectory matrix with a window size L, its Toeplitz lagged correlation matrix $\mathbf{C}$ is formulated by

$$\mathbf{C} = \begin{bmatrix} c(0) & c(1) & \cdots & c(L-1) \\ c(1) & c(0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & c(1) \\ c(L-1) & \cdots & \cdots & c(0) \end{bmatrix}. \tag{1}$$

Each element $c(j)$ is computed by

$$c(j) = \frac{1}{N-j} \sum_{i=1}^{N-j} x_i x_{i+j} \quad j = 0, 1, 2, \cdots, L-1. \tag{2}$$

For matrix $\mathbf{C}$, we can compute its eigenvalues λ_k and the corresponding eigenvectors $\boldsymbol{v}_k$ in descending order of λ_k $(1 \le k \le L)$. Then the kth principal components (PCs) is $\boldsymbol{a}_k$ and its ith element $a_{k,i}$ is computed by

$$a_{k,i} = \sum_{j=1}^{L} x_{i+j-1} v_{k,j} \quad 1 \le i \le N-L+1 \tag{3}$$

where $v_{k,j}$ is the jth element of $\boldsymbol{v}_k$. We compute the kth reconstructed components (RCs) of the time series with the kth PCs as (Vautard et al., 1992)

$$x_i^k = \begin{cases} \frac{1}{i} \sum_{j=1}^{i} a_{k,i-j+1} v_{k,j} & 1 \le i \le L-1 \\ \frac{1}{L} \sum_{j=1}^{L} a_{k,i-j+1} v_{k,j} & L \le i \le N-L+1 \\ \frac{1}{N-i+1} \sum_{j=i-N+L}^{L} a_{k,i-j+1} v_{k,j} & N-L+2 \le i \le N \end{cases}. \tag{4}$$

The variance contained in the kth RC is λ_k. Since λ_k is sorted in descending order, the first several RCs contain most of the signals of the time series, while the remaining RCs contain mainly the noises of time series. Thus the original time series will be reconstructed with first several RCs.

The SSAM approach developed by Schoellhamer (2001) computes the elements $c(j)$ of the lagged correlation matrix by,

$$c(j) = \frac{1}{N_j} \sum_{i \le N-j} x_i x_{i+j} \quad j = 0, 1, 2, \cdots, L-1 \tag{5}$$

where, both x_i and x_{i+j} must be observed rather than missed, N_j is the number of the products of x_i and x_{i+j} within the sample index $i \le N-j$. The PCs are also calculated with observed data,

$$a_{k,i} = \frac{L}{L_i} \sum_{1 \le j \le L} x_{i+j-1} v_{k,j} \quad 1 \le i \le N-L+1 \tag{6}$$

where L_i is the number of observed data within the sample index from i to $i+L-1$. The reconstruction procedure of time series from PCs is the same as SSA. The scale factor, L/L_i, is used to compensate the missing value. In order to derive the expression of computing PCs for the time series with missing data, the expression Eq. (3) is reformulated as,

$$a_{k,i} = \sum_{i+j-1 \in S_i} x_{i+j-1} v_{k,j} + \sum_{i+j-1 \in \overline{S}_i} x_{i+j-1} v_{k,j} \tag{7}$$

where, $1 \le i \le N-L+1$, S_i and $\overline{S}_i$ are the index sets of sampling data and missing data respectively within the integer interval $[i, i+L-1]$, i.e. $S_i \cap \overline{S}_i = 0$ and $S_i \cup \overline{S}_i = [i, i+L-1]$. If PCs are available, we can reproduce the missing values. Therefore, the missing values in Eq. (7) can be substituted with PCs as,

$$x_{i+j-1} = \sum_{m=1}^{L} a_{m,i} v_{m,j}. \tag{8}$$

Substituting Eq. (8) into the second term of the right hand of Eq. (7) yields,

$$\left(1 - \sum_{i+j-1 \in \overline{S}_i} v_{k,j}^2\right) a_{k,i} - \sum_{i+j-1 \in \overline{S}_i} \sum_{m=1, m \neq k}^{L} v_{m,j} v_{k,j} a_{m,i} = \sum_{i+j-1 \in S_i} x_{i+j-1} v_{k,j}. \tag{9}$$

Collecting all equations of Eq. (9) for $k = 1, 2, \ldots, L$, we have,

$$\mathbf{G}_i \boldsymbol{\xi}_i = \boldsymbol{y}_i \tag{10}$$

where,

$$\mathbf{G}_i = \begin{bmatrix} 1 - \sum\limits_{i+j-1 \in \overline{S}_i} v_{1,j}^2 & - \sum\limits_{i+j-1 \in \overline{S}_i} v_{1,j} v_{2,j} & \cdots & - \sum\limits_{i+j-1 \in \overline{S}_i} v_{1,j} v_{L,j} \\ - \sum\limits_{i+j-1 \in \overline{S}_i} v_{2,j} v_{1,j} & 1 - \sum\limits_{i+j-1 \in \overline{S}_i} v_{2,j}^2 & \cdots & - \sum\limits_{i+j-1 \in \overline{S}_i} v_{2,j} v_{L,j} \\ \vdots & \vdots & \ddots & \vdots \\ - \sum\limits_{i+j-1 \in \overline{S}_i} v_{L,j} v_{1,j} & - \sum\limits_{i+j-1 \in \overline{S}_i} v_{L,j} v_{2,j} & \cdots & 1 - \sum\limits_{i+j-1 \in \overline{S}_i} v_{L,j}^2 \end{bmatrix}, \tag{11}$$

$$\boldsymbol{\xi}_i = \begin{bmatrix} a_{1,i} \\ a_{2,i} \\ \vdots \\ a_{L,i} \end{bmatrix}, \quad \boldsymbol{y}_i = \begin{bmatrix} \sum\limits_{i+j-1 \in S_i} x_{i+j-1} v_{1,j} \\ \sum\limits_{i+j-1 \in S_i} x_{i+j-1} v_{2,j} \\ \vdots \\ \sum\limits_{i+j-1 \in S_i} x_{i+j-1} v_{L,j} \end{bmatrix}. \tag{12}$$

Since $\mathbf{G}_i$ is a symmetric and rank-deficient matrix with the number of rank-deficiency equaling to the number of missing data within the interval $[x_i, x_{i+L-1}]$, the PCs $a_{k,i}$ $(k = 1, 2, \ldots, L)$ are solved with Eq. (10) based on the following criterion (Shen et al., 2014),

$$\min : \boldsymbol{\xi}_i^T \boldsymbol{\Lambda} \boldsymbol{\xi}_i \tag{13}$$

where, $\boldsymbol{\Lambda}$ is diagonal matrix of eigenvalues λ_k, which is the covariance matrix of PCs. The solution of Eq. (10) is as follows,

$$\boldsymbol{\xi}_i = \boldsymbol{\Lambda} \mathbf{G}_i^T \left(\mathbf{G}_i^T \boldsymbol{\Lambda} \mathbf{G}_i \right)^{-} \boldsymbol{y}_i . \tag{14}$$

The symbol "–" denotes the pseudo-inverse of a matrix. If the non-diagonal elements of $\mathbf{G}_i$ are all set to zero, the expression Eq. (14) can be further simplified as,

$$a_{k,i} = \frac{1}{1 - \sum_{i+j-1 \in \bar{S}_i} v_{k,j}^2} \sum_{1 \le j \le L} x_{i+j-1} v_{k,j} \quad 1 \le k \le L, \;\; 1 \le i \le N - L + 1. \tag{15}$$

Supposing $v_{k,1} = v_{k,2} = \cdots = v_{k,L} = 1/\sqrt{L}$ at the missing data points, the solution of Eq. (15) can be reduced to Eq. (6). Therefore, the SSAM approach is a special case of our ISSA approach. By the way, the first several PCs contain most variance; the element x_{i+j-1} can be approximately reproduced with the first several PCs in Eq. (8).

The main difference of our ISSA approach from the SSAM approach of Schoellhamer (2001) is in calculating the PCs. We produce the PCs from observed data with Eq. (14) according to the power spectrum (eigenvalues) and eigenvectors of the PCs. While Schoellhamer (2001) calculates the PCs from observed data with Eq. (6) only according to the eigenvectors and uses the scale factor L/L_i to compensate the missing value. We have pointed out that this scale factor can be derived from Eq. (15), which is the simplified version of our ISSA approach, by supposing the missing data points with the same eigenvector elements. Therefore the performance of our ISSA approach will be better than SSAM of Schoellhamer (2001).

3 Performance of ISSA with synthetic time series

The same synthetic time series as Schoellhamer (2001) are used to analyze the performance of ISSA compared to SSAM. The real time series was collected in San Francisco Bay every 15 min from 1991 to 2001, which contains many missing data. The synthetic SSC time series is expressed as,

$$c(t) = 0.2R(t)c_\mathrm{s}(t) + c_\mathrm{s}(t) \tag{16}$$

where, $R(t)$ is a time series of Gaussian white noise with zero mean and unit SD; $c_\mathrm{s}(t)$ is the periodic signal expressed as,

$$\begin{aligned} c_\mathrm{s}(t) = {} & 100 - 25\cos\omega_\mathrm{s}t + 25(1 - \cos 2\omega_\mathrm{s}t)\sin\omega_\mathrm{sn}t \\ & + 25(1 + 0.25(1 - \cos 2\omega_\mathrm{s}t)\sin\omega_\mathrm{sn}t)\sin\omega_\mathrm{a}t. \end{aligned} \tag{17}$$

The periodic signal oscillates about the mean value 100 mg L^{-1} including the signals with seasonal frequency $\omega_\mathrm{s} = 2\pi/365\,\mathrm{day}^{-1}$, spring/neap angular frequency $\omega_\mathrm{sn} = 2\pi/14\,\mathrm{day}^{-1}$ and advection angular frequency $\omega_\mathrm{a} = 2\pi/(12.5/24)\,\mathrm{day}^{-1}$. The one year of synthetic SSC time series $c(t)$, starting at 1 October with 15 min time step, is presented on the bottom of Fig. 1, the corresponding periodic signal $c_\mathrm{s}(t)$ is shown on the top of Fig. 1.

We use the 30 h window size ($L = 120$), same as Schoellhamer (1996, 2001), to form the covariance matrix and then compute the variances of each modes. The first 4 modes contain the periodic components, which account for 72.3 % of the total variance; particularly, the first mode contains even 50.2 % of the total variance. In order to evaluate the accuracies of recovered PCs from the time series with different percentages of missing data, following the idea in Shen et al. (2014), we compute the relative errors of the first four modes derived by ISSA and SSAM with the following expression,

$$p = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{(\boldsymbol{a}_i - \boldsymbol{a}_0)^T(\boldsymbol{a}_i - \boldsymbol{a}_0)}{\boldsymbol{a}_0^T\boldsymbol{a}_0}} \times 100\,\% \tag{18}$$

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

where, p denotes relative error; N is the number of repeated experiments; $\boldsymbol{a}_i$ is the reconstructed PCs of ith experiment from data missing time series, $\boldsymbol{a}_0$ denotes the PCs reconstructed from the time series without missing data. We design the experiment of missing data by randomly deleting the data from the synthetic time series. The percentage of deleted data is from 10 to 60 % with an increase of 10 % each time. Then, we reconstruct the first four PCs from the data deleted synthetic time series using both SSAM and ISSA, and repeat the experiments for 50 times. The relative errors of the first four PCs are presented in Fig. 2, from which we clear see that the accuracies of reconstructed PCs by our ISSA are obviously higher than those by SSAM, especially for the second and fourth PCs. In the case of 60 % missing data, the accuracy improvements are up to 19.64, 41.34, 23.27 and 50.30 % for the first four PCs, respectively.

We reconstruct the time series $\hat{c}(t)$ using the first four PC modes and then evaluate the quality of reconstructed series by examining the error $\Delta\hat{c}(t) = \hat{c}(t) - c_s(t)$. For the cases whose missing data are between 10 to 50 % over the whole time series, the reconstructed component of the time series is calculated only when the percentage of missing data in the window size is less than 50 %; while for the cases whose overall missing data already reach 60 %, it is allowed 60 % missing data in the window size. In Fig. 3, we demonstrate the root mean squared errors (RMS) of each experiment of different percentages of missing data. The RMS is computed with $\Delta\hat{c}(t)$ as

$$\mathrm{RMS} = \sqrt{\sum_{j=1}^{M} \Delta\hat{c}^2\left(t_j\right)/M} \tag{19}$$

where M is the number of data points involved in the experiment.

As we can see from the Fig. 3, the RMSs of ISSA are much smaller than those of SSAM for all same experiment scenarios. In Table 1, we present the mean absolute error (MAE) and mean root mean squared errors (MRMS) of 50 experiments with different percentages of missing data.

Obviously, if there is no missing data, the ISSA coincides with SSAM. If the percentage of missing data increases, both MAE and MRMS will become larger. In Table 1, all

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀| |▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

the MAE and MRMS of ISSA are smaller than those of SSAM. When the percentage of missing data reaches 50 %, the MAE and MRMS are 3.17 and 4.14 $\mathrm{mg\,L^{-1}}$ for ISSA, and 4.57 and 5.89 $\mathrm{mg\,L^{-1}}$ for SSAM, respectively. The improved percentage (IMP) of ISSA with respect to SSAM is also listed in Table 1. As the missing data increases, the IMPs of both MAE and MRMS increase as well. Moreover, when the synthetic time series with the missing data is same as the real SSC time series of Fig. 4, the IMPs of MAE and MRMS are 8.87 and 15.19 %, respectively.

4 Performance of ISSA with real time series

The mid-depth SSC time series at San Mateo Bridge is presented in Fig. 4, which contains about 61 % missing data. This time series was reported by Buchanan and Schoellhamer (1999) and Buchanan and Ruhl (2000), and analyzed by Schoellhamer (2001) using SSAM. We analyze this time series using our ISSA with the window size of 30 h ($L = 120$) comparing with Schoellhamer's SSAM. The first 10 modes represent dominant periodic components as shown in Schoellhamer (2001) which contain 89.1 % of the total variance. Therefore, we reconstruct the time series with first 10 modes when the missing data in a window size is less than 50 %.

The residual time series, e.g. the differences of observed minus reconstructed data, are presented in Fig. 5. The maximum, minimum and mean absolute residuals as well as the SD are presented in Table 2. It is clear that both maximum and minimum residuals are significantly reduced by using ISSA approach. The SD of our ISSA is reduced by 8.6 %. The squared correlation coefficients between the observations and the reconstructed data from ISSA and SSAM are 0.9178 and 0.9046, respectively, which reflect that the reconstructed time series with our ISSA can indeed, to very large extent, specify the real time series.

5 Conclusions

We have developed the ISSA approach in this paper for processing the incomplete time series by using the principle that a time series can be reproduced by using its principal components. We proved that the SSAM developed by Schoellhamer (2001) is a special case of our ISSA. The performances of ISSA and SSAM were demonstrated with a synthetic time series, and the results show that the relative errors of the first four principal components by ISSA are significantly smaller than those by SSAM. With the missing data gets more, the improvements of the relative errors becomes more evident. When the percentage of missing data reaches 60 %, the improvements of the first four principal components are up to 19.64, 41.34, 23.27 and 50.30 %, respectively. Moreover, when the missing data accounts for 60 %, the MAE and MRMS derived by ISSA are 3.52 and 4.60 $\mathrm{mg\,L^{-1}}$, and by SSAM are 5.37 and 6.96 $\mathrm{mg\,L^{-1}}$. The corresponding improvements of ISSA with respect to SSAM are 34.45 and 33.91 %. When the missing data of synthetic time series is the same as the real SSC time series, the improvements of MAE and MRMS are 8.87 and 15.19 %, respectively. The SD derived from the real SSC time series at San Mateo Bridge by ISSA and SSAM are 12.27 and 13.48 $\mathrm{mg\,L^{-1}}$, and the squared correlation coefficients between the observations and the reconstructed data from ISSA and SSAM are 0.9178 and 0.9046, respectively. Therefore, ISSA can indeed, to a great extent, retrieve the informative signals from the original incomplete time series.

Author contributions. Y. Shen proposed the improved singular spectrum analysis and F. Peng carried out the matlab program and performed the simulations. Y. Shen, F. Peng and B. Li prepared the manuscript.

Acknowledgements. This work was sponsored by Natural Science Foundation of China (Projects: 41274035, 41474017) and partly supported by State Key Laboratory of Geodesy and Earth's Dynamics (SKLGED2013-3-2-Z).

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◂◂ ▸▸
◂ ▸
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

References

Broomhead, D. S. and King, G. P.: Extracting qualitative dynamics from experimental data, Physica D, 20, 217–236, 1986.

Buchanan, P. A. and Ruhl, C. A.: Summary of suspended-solids concentration data, San Francisco Bay, California, water year 1998, US Geological Survey Open File Report 2000-88, US Geological Survey, 41 pp., 2000.

Buchanan, P. A. and Schoellhamer, D. H.: Summary of Suspended-Solids Concentration Data, San Francisco Bay, California, Water Year 1997, US Geological Survey Open-File Report 99-189, US Geological Survey, p. 52, 1999.

Golyandina, N. and Osipov, E.: The “Catterpillar”-SSA method for analysis of time series with missing data, J. Stat. Plan. Inf., 137, 2642–2653, 2007.

Kondrashov, D. and Ghil, M.: Spatio-temporal filling of missing points in geophysical data sets, Nonlin. Processes Geophys., 13, 151–159, doi:10.5194/npg-13-151-2006, 2006.

Oropeza, V. and Sacchi, M.: Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis, Geophysics, 76, 25–32, 2011.

Robertson, A. W. and Mechoso, C. R.: Interannual and decadal cycles in river flows of southeastern South America, J. Climate, 11, 2570–2581, 1998.

Rodrigues, P. C. and de Carvalho, M.: Spectral modeling of time series with missing data, Appl. Math. Model., 37, 4676–4684, 2013.

Schoellhamer, D. H.: Factors affecting suspended-solids concentrations in South San Francisco Bay, California, J. Geophys. Res., 101, 12087–12095, 1996.

Schoellhamer, D. H.: Singular spectrum analysis for time series with missing data, Geophys. Res. Lett., 28, 3187–3190, 2001.

Schoellhamer, D. H.: Variability of suspended-sediment concentration at tidal to annual time scales in San Francisco Bay, USA, Cont. Shelf Res., 22, 1857–1866, 2002.

Shen, Y., Li, W., Xu, G., and Li, B.: Spatiotemporal filtering of regional GNSS network’s position time series with missing data using principal component analysis, J. Geodesy., 88, 1–12, doi:10.1007/s00190-013-0663-y, 2014.

Vautard, R. and Ghil, M.: Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series, Physica D, 35, 395–424, 1989.

Vautard, R., Yiou, P., and Ghil, M.: Singular-spectrum analysis: a toolkit for short, noisy, chaotic signals, Physica D, 58, 95–126, 1992.

Wang, X. L., Corte-Real, J., and Zhang, X.: Intraseasonal oscillations and associated spatial–temporal structures of precipitation over China, J. Geophys. Res., 101, 19035–19042, 1996.

Yiou, P., Fuhrer, K., Meeker, L. D., Jouzel, J., Johnsen, S., and Masked, P. A.: Paleoclimatic variability inferred from the spectral analysis of Greenland and Antarctic ice-core data, J. Geophys. Res., 102, 26441–26454, 1997.

Zotova, L. V. and Shum, C. K.: Multichannel singular spectrum analysis of the gravity field from grace satellites, AIP Conf. Proc., 1206, 473–479, 2010.

Table 1. Mean absolute error and MRMS ($mg\,L^{-1}$).

Missing Data (%)	MAE			MRMS		
	SSAM	ISSA	IMP (%)	SSAM	ISSA	IMP (%)
0	2.48	2.48	0	2.06	2.06	0
10	2.87	2.60	9.41	3.68	3.38	2.21
20	3.26	2.73	16.26	4.19	3.56	15.04
30	3.71	2.90	21.83	4.76	3.78	20.59
40	4.22	3.11	26.30	5.42	4.07	24.91
50	4.57	3.17	30.63	5.89	4.14	29.71
60	5.37	3.52	34.45	6.96	4.60	33.91
As Bay Example	3.38	3.08	8.87	2.70	2.29	15.19

Table 2. Maximum and minimum and mean absolute residuals of SSAM and ISSA.

Residuals ($\mathrm{mg\,L^{-1}}$)	SSAM	ISSA
Maximum	145.05	126.61
Minimum	−432.20	−227.70
SD	13.48	12.27

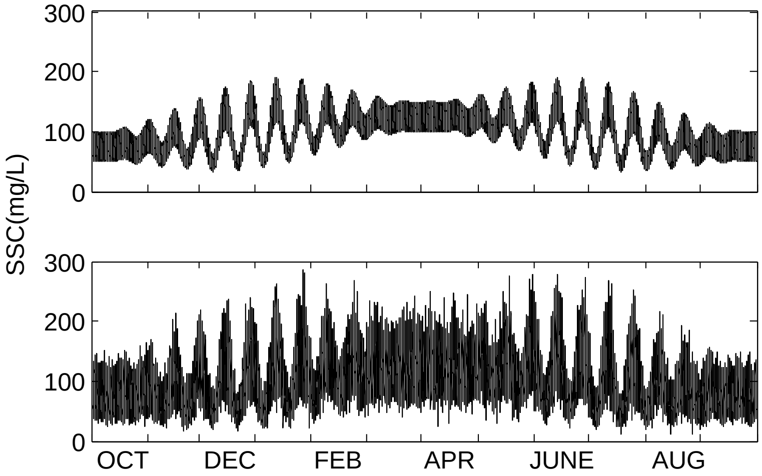

Figure 1. Periodic signal $c_s(t)$ (top panel) and synthetic time series (bottom panel).

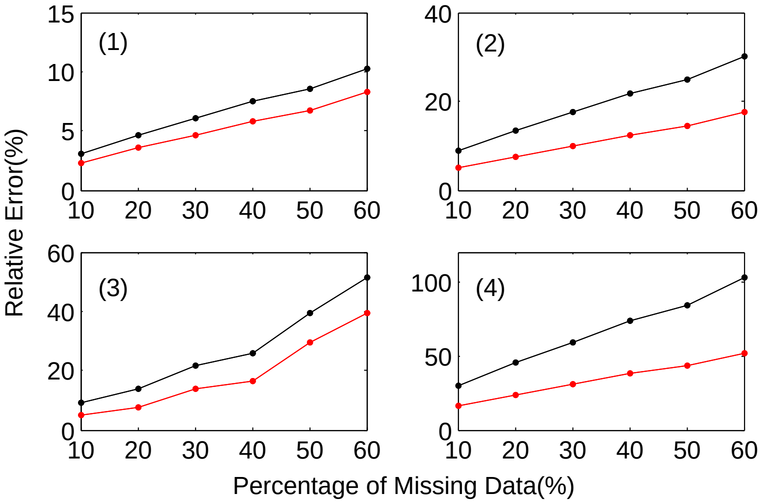

Figure 2. Relative errors of first four PCs (ISSA: red line; SSAM: black line).

NPGD

1, 1947–1966, 2014

Improved singular spectrum analysis for time series with missing data

Y. Shen et al.

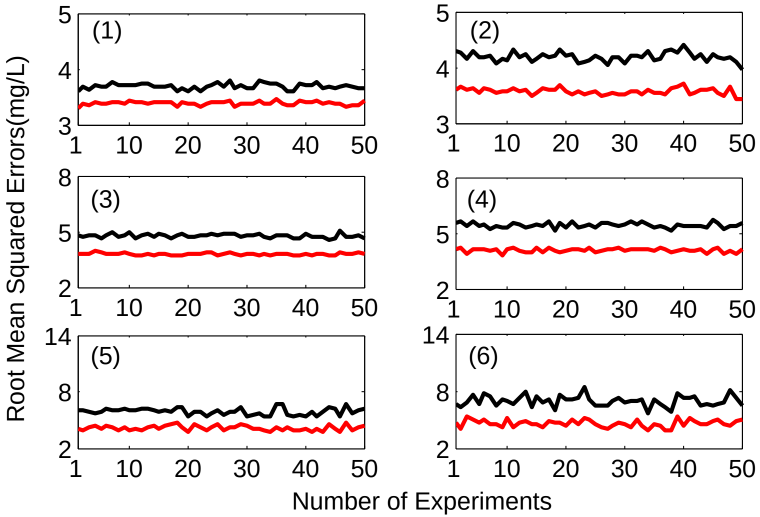

Figure 3. RMS of 50 experiments, (1)–(6) represent percentage of missing data ranging from 10 to 60 % with 10 % increments.

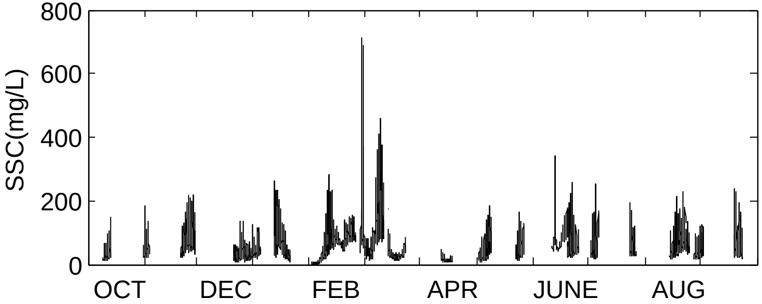

Figure 4. Mid-depth SSC time series at San Mateo Bridge during water year 1997.

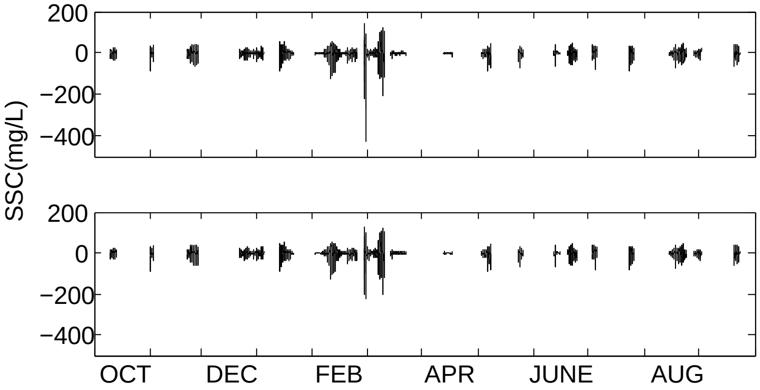

Figure 5. Residual series after removing reconstructed signals from first 10 modes (top panel: SSAM; bottom panel: ISSA).