

Reply to Editor

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

C1: I am glad to inform you that both the referees found that you satisfactorily addressed their comments and recommended the paper to publication. Before the paper will be published, I ask you to do one more round of revisions and correct the verb tense usage. The current version (abstract in particular) mixes the past, present, and future tenses. Please rewrite the paper using the same tense (e.g., "this paper develops", "the approach is evaluated", etc.)

R1: Thank you very much for your comments! our manuscript is corrected and the changes are listed as follows:

Line 9: 'will develop' has changed to '**develops**'.

Line 11: 'was' has changed to '**is**'.

Line 37: 'has been' has changed to '**was**'.

Line 48: 'will develop' has changed to '**develops**'.

Line 73: 'will be' has changed to '**is**'.

Line 125: 'will be' has changed to '**is**'.

Line 135: '100mg/L' has changed to '**100 mgL⁻¹**'.

Line 146: 'computed' has changed to '**compute**'.

Line 220: 'Maximum and minimum' has changed into '**Maximum, minimum**'.

Line 226: 'reproduced by using' has changed into '**reproduced using**'.

Line 228: 'were' has changed into '**are**'.

Line 248: 'Y. Shen proposed the improved singular spectrum analysis and F. Peng carried out the FORTRAN program and performed the simulations. Y. Shen, F. Peng and B. Li prepared the manuscript.' has changed into '**Y. Shen proposes the improved singular spectrum analysis and F. Peng carries out the FORTRAN program and performs the simulations. Y. Shen, F. Peng and B. Li prepare the manuscript.**'

Line 253: 'was' has changed into '**is**'.

Improved Singular Spectrum Analysis for Time Series with Missing Data

Y. Shen¹ F. Peng^{1,2} B. Li¹

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

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 **develops** an improved SSA (ISSA) for processing the incomplete time series and the modified SSA (SSAM) of Schoellhamer (2001) is its special case. The approach **is** 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 error and mean root mean squared error of the reconstructed time series by ISSA are also 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 standard deviation (SD) derived by ISSA is 12.27mg L⁻¹, smaller than 13.48 mg L⁻¹ derived by SSAM.

Keywords: Time series analysis, Singular spectrum Analysis, Missing Data

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 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 **was** successfully applied to analyze the time series of suspended-sediment concentration (SSC) in San Francisco Bay (Schoellhamer, 2002). This SSAM approach doesn't 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

85 principal component analysis approach for extracting common mode errors from the
 86 time series with missing data of a regional station network. The other kind of SSA
 87 approaches process the time series with missing data by filling the data gaps
 88 recursively or iteratively, such as the ‘‘Catterpillar’’-SSA method (Golyandina and
 89 Osipov, 2007), the imputation method (Rodrigues and Carvalho, 2013) or the iterative
 90 method (Kondrashov and Ghil, 2006).

91 This paper is motivated by Schoellhamer (2001) and Shen et al. (2014) and **develops**
 92 an improved SSA (ISSA) approach. In our ISSA, the lagged correlation matrix is
 93 computed with the same way as Schoellhamer (2001), the PCs are directly computed
 94 with both the eigenvalues and eigenvectors of the lagged correlation matrix. However,
 95 the PCs in Schoellhamer (2001) were calculated with the eigenvectors and a scale
 96 factor to compensate the missing value. Moreover, we do not need to fill the missing
 97 data recursively and iteratively as in Golyandina and Osipov (2007). The rest of this
 98 paper is organized as follows: the improvement of SSA for time series with missing
 99 data will be followed in Sect. 2, synthetic and real numerical examples are presented
 100 in Sects. 3 and 4 respectively, and then conclusions are given in last Sect. 5.

101 2. Improved Singular Spectrum Analysis for Time Series with Missing Data

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

$$104 \quad \mathbf{C} = \begin{bmatrix} c(0) & c(1) & \cdots & c(L-1) \\ c(1) & c(0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & c(1) \\ c(L-1) & \cdots & \cdots & c(0) \end{bmatrix} \quad (1)$$

105 Each element $c(j)$ is computed by

$$106 \quad c(j) = \frac{1}{N-j} \sum_{i=1}^{N-j} x_i x_{i+j} \quad j = 0, 1, 2, \dots, L-1 \quad (2)$$

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

$$110 \quad a_{k,i} = \sum_{j=1}^L x_{i+j-1} v_{j,k} \quad 1 \leq i \leq N-L+1 \quad (3)$$

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

$$113 \quad x_i^k = \begin{cases} \frac{1}{i} \sum_{j=1}^i a_{k,i-j+1} v_{j,k} & 1 \leq i \leq L-1 \\ \frac{1}{L} \sum_{j=1}^L a_{k,i-j+1} v_{j,k} & L \leq i \leq N-L+1 \\ \frac{1}{N-i+1} \sum_{j=i-N+L}^L a_{k,i-j+1} v_{j,k} & N-L+2 \leq i \leq N \end{cases} \quad (4)$$

114 Since λ_k , the variance of the k th RC, is sorted in descending order, the first several
 115 RCs contain most of the signals of the time series, while the remaining RCs contain
 116 mainly the noises of time series. Thus the original time series **is** reconstructed with
 117 first several RCs.

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

$$120 \quad c(j) = \frac{1}{N_j} \sum_{i \leq N-j} x_i x_{i+j} \quad j = 0, 1, 2, \dots, L-1 \quad (5)$$

121 where, both x_i and x_{i+j} must be observed rather than missed, N_j is the number of the
 122 products of x_i and x_{i+j} within the sample index $i \leq N-j$. Then we compute the
 123 eigenvalues and eigenvectors from the lagged correlation matrix C . The PCs are also
 124 calculated with observed data,

$$125 \quad a_{k,i} = \frac{L}{L_i} \sum_{1 \leq j \leq L} x_{i+j-1} v_{j,k} \quad 1 \leq i \leq N-L+1 \quad (6)$$

126 where L_i is the number of observed data within the sample index from i to $i+L-1$. The
 127 reconstruction procedure of time series from PCs is the same as SSA. The scale factor
 128 L/L_i is used to compensate the missing value.

129 In order to derive the expression of computing PCs for the time series with missing
 130 data, the Eq. (3) is reformulated as,

$$131 \quad a_{k,i} = \sum_{i+j-1 \in S_i} x_{i+j-1} v_{j,k} + \sum_{i+j-1 \in \bar{S}_i} x_{i+j-1} v_{j,k} \quad (7)$$

132 where, $1 \leq i \leq N-L+1$, S_i and \bar{S}_i are the index sets of sampling data and missing
 133 data respectively within the integer interval $[i, i+L-1]$, i.e. $S_i \cap \bar{S}_i = 0$ and
 134 $S_i \cup \bar{S}_i = [i, i+L-1]$. If PCs are available, we can reproduce the missing values. Therefore,
 135 the missing values in Eq. (7) can be substituted with PCs as,

$$136 \quad x_{i+j-1} = \sum_{m=1}^L a_{m,i} v_{j,m} \quad (8)$$

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

$$138 \quad \left(1 - \sum_{i+j-1 \in \bar{S}_i} v_{j,k}^2 \right) a_{k,i} - \sum_{i+j-1 \in \bar{S}_i} \sum_{m=1, m \neq k}^L v_{j,m} v_{j,k} a_{m,i} = \sum_{i+j-1 \in S_i} x_{i+j-1} v_{j,k} \quad (9)$$

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

$$140 \quad \mathbf{G}_i \boldsymbol{\xi}_i = \mathbf{y}_i \quad (10)$$

141 where,

$$142 \quad \mathbf{G}_i = \begin{bmatrix} 1 - \sum_{i+j-1 \in \bar{S}_i} v_{j,1}^2 & - \sum_{i+j-1 \in \bar{S}_i} v_{j,1} v_{j,2} & \cdots & - \sum_{i+j-1 \in \bar{S}_i} v_{j,1} v_{j,L} \\ - \sum_{i+j-1 \in \bar{S}_i} v_{j,2} v_{j,1} & 1 - \sum_{i+j-1 \in \bar{S}_i} v_{j,2}^2 & \cdots & - \sum_{i+j-1 \in \bar{S}_i} v_{j,2} v_{j,L} \\ \vdots & \vdots & \ddots & \vdots \\ - \sum_{i+j-1 \in \bar{S}_i} v_{j,L} v_{j,1} & - \sum_{i+j-1 \in \bar{S}_i} v_{j,L} v_{j,2} & \cdots & 1 - \sum_{i+j-1 \in \bar{S}_i} v_{j,L}^2 \end{bmatrix}, \quad (11)$$

$$143 \quad \boldsymbol{\xi}_i = \begin{bmatrix} a_{1,i} \\ a_{2,i} \\ \vdots \\ a_{L,i} \end{bmatrix}, \quad \mathbf{y}_i = \begin{bmatrix} \sum_{i+j-1 \in S_i} x_{i+j-1} v_{j,1} \\ \sum_{i+j-1 \in S_i} x_{i+j-1} v_{j,2} \\ \vdots \\ \sum_{i+j-1 \in S_i} x_{i+j-1} v_{j,L} \end{bmatrix} \quad (12)$$

144 Since \mathbf{G}_i is a symmetric and rank-deficient matrix with the number of rank-deficiency
 145 equaling to the number of missing data within the interval $[x_i, x_{i+L-1}]$, the PCs $a_{k,i}$
 146 ($k=1, 2, \dots, L$) are solved with Eq. (10) based on the following criterion (Shen et al.
 147 2014),

$$148 \quad \min : \boldsymbol{\xi}_i^T \mathbf{A}^{-1} \boldsymbol{\xi}_i \quad (13)$$

149 where, \mathbf{A} is diagonal matrix of eigenvalues λ_k , which is the covariance matrix of PCs.
 150 The solution of Eq. (10) is as follows,

$$151 \quad \boldsymbol{\xi}_i = \mathbf{A} \mathbf{G}_i^T (\mathbf{G}_i^T \mathbf{A} \mathbf{G}_i)^{-} \mathbf{y}_i \quad (14)$$

152 The symbol ‘-’ denotes the pseudo-inverse of a matrix.

153 If the non-diagonal elements of \mathbf{G}_i are all set to zero, the Eq. (14) can be further
 154 simplified as,

$$155 \quad a_{k,i} = \frac{1}{1 - \sum_{i+j-1 \in \bar{S}_i} v_{k,j}^2} \sum_{i+j-1 \in S_i} x_{i+j-1} v_{j,k} \quad 1 \leq k \leq L, 1 \leq i \leq N-L+1 \quad (15)$$

156 Supposing $v_{1,k} = v_{2,k} = \dots = v_{L,k} = 1/\sqrt{L}$ at the missing data points, the solution of Eq.
 157 (15) will be reduced to Eq. (6). Therefore, the SSAM approach is a special case of our

158 ISSA approach. By the way, the first several PCs contain most variance; the element
159 x_{i+j-1} can be approximately reproduced with the first several PCs in Eq. (8).

160 The main difference of our ISSA approach from the SSAM approach of Schoellhamer
161 (2001) is in calculating the PCs. We produce the PCs from observed data with Eq. (14)
162 according to the power spectrum (eigenvalues) and eigenvectors of the PCs. While
163 Schoellhamer (2001) calculates the PCs from observed data with Eq. (6) only
164 according to the eigenvectors and uses the scale factor L/L_i to compensate the missing
165 value. We have pointed out that this scale factor can be derived from Eq. (15), which
166 is the simplified version of our ISSA approach, by supposing the missing data points
167 with the same eigenvector elements. Therefore the performance of our ISSA approach
168 is better than SSAM of Schoellhamer (2001). The only disadvantage of our method is
169 that it will cost more computational effort.

170 3. Performance of ISSA with synthetic time series

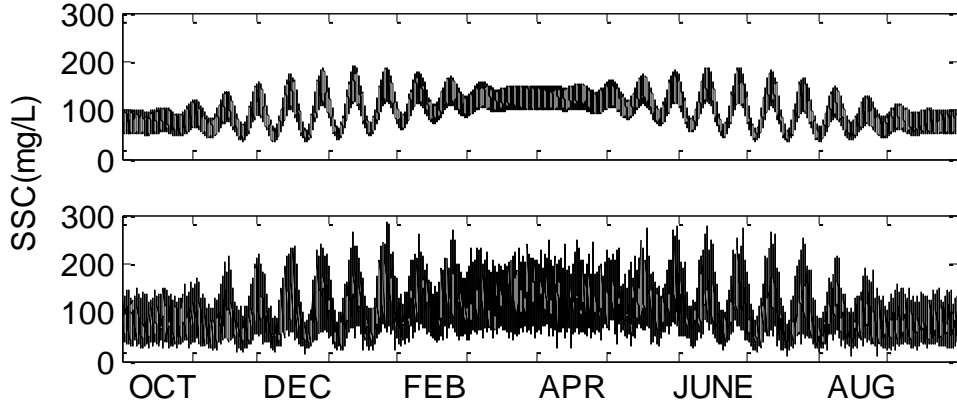
171 The same synthetic time series as Schoellhamer (2001) are used to analyze the
172 performance of ISSA compared to SSAM. The synthetic SSC time series is expressed
173 as,

$$174 \quad c(t) = 0.2R(t)c_s(t) + c_s(t) \quad (16)$$

175 where, $R(t)$ is a time series of Gaussian white noise with zero mean and unit standard
176 deviation; $c_s(t)$ is the periodic signal expressed as,

$$177 \quad c_s(t) = 100 - 25 \cos \omega_s t + 25(1 - \cos 2\omega_s t) \sin \omega_{sn} t \\ + 25(1 + 0.25(1 - \cos 2\omega_s t) \sin \omega_{sn} t) \sin \omega_a t \quad (17)$$

178 The periodic signal oscillates about the mean value 100mg L^{-1} including the signals
179 with seasonal frequency $\omega_s = 2\pi / 365 \text{ day}^{-1}$, spring/neap angular frequency
180 $\omega_{sn} = 2\pi / 14 \text{ day}^{-1}$ and advection angular frequency $\omega_a = 2\pi / (12.5 / 24) \text{ day}^{-1}$. The one
181 year of synthetic SSC time series $c(t)$, starting at October 1 with 15-minute time step,
182 is presented on the bottom of Fig. 1, the corresponding periodic signal $c_s(t)$ is
183 shown on the top of Fig. 1.



184

185

Figure 1. periodic signal $c_s(t)$ (top) and Synthetic time series (bottom)

186

Although the selection of window length is an important issue for SSA (Hassani 2012, 2013), this paper chooses the same window length ($L=120$) as that in Schoellhamer (2001) in order to compare the performance of the proposed method with that of Schoellhamer (2001). Using the synthetic time series we compute the lagged correlation matrix and the variances of each mode. The first 4 modes contain the periodic components, which account for 72.3% of the total variance; particularly, the first mode contains 50.2% of the total variance. In order to evaluate the accuracies of reconstructed PCs from the time series with different percentages of missing data, following the way of Shen et al. (2014), we compute the relative errors of the first four modes derived by ISSA and SSAM with the following expression,

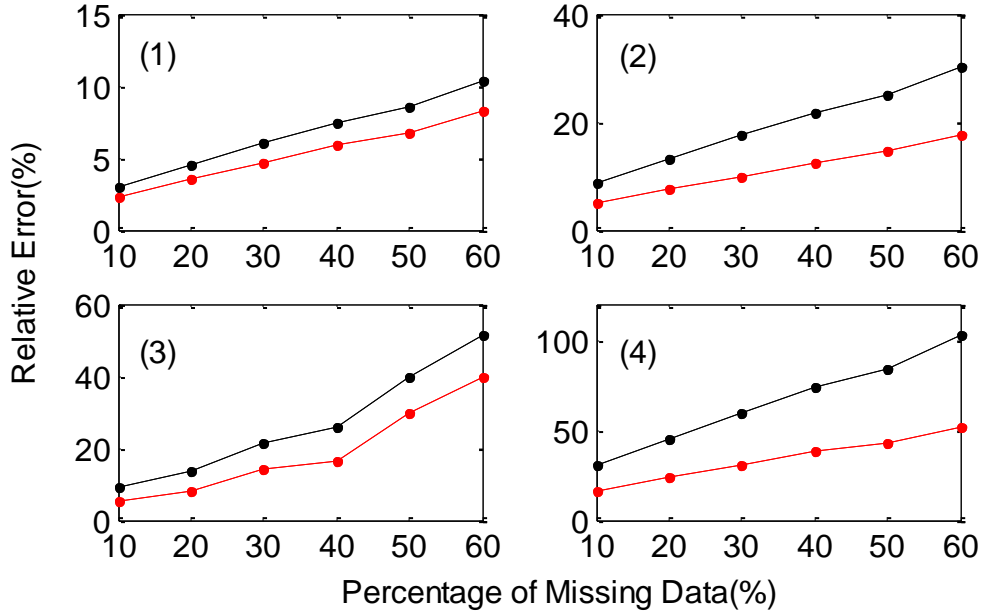
196

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

197

where, The symbol ' T ' denotes the transpose of a matrix; p denotes relative error; N is the number of repeated experiments; \mathbf{a}_i is the reconstructed PCs of i th experiment from data missing time series, \mathbf{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 clearly 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.

209



210

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

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

220
$$\text{RMS} = \sqrt{\frac{\sum_{j=1}^M \Delta\hat{c}^2(t_j)}{M}} \quad (19)$$

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

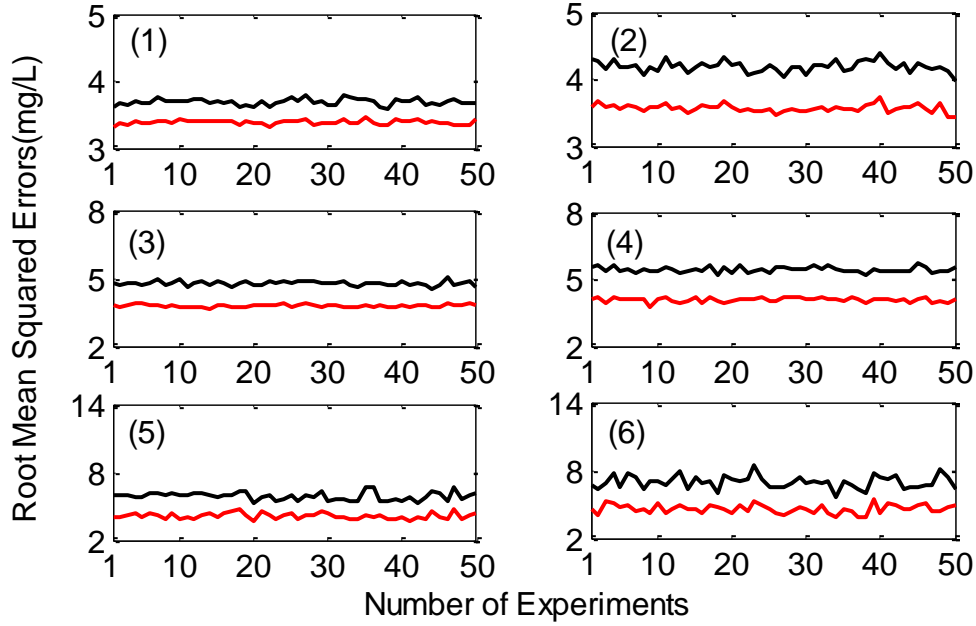


Figure 3. RMSE of 50 experiments, (1)~(6) represent percentage of missing data ranging from 10% to 60% with 10% increments.

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 reconstruction error (MARE) and mean root mean squared errors (MRMSE) of 50 experiments with different percentages of missing data.

Table 1: Mean absolute reconstruction error and mean root mean squared error of simulated time series with different percentage of missing data (mg L^{-1})

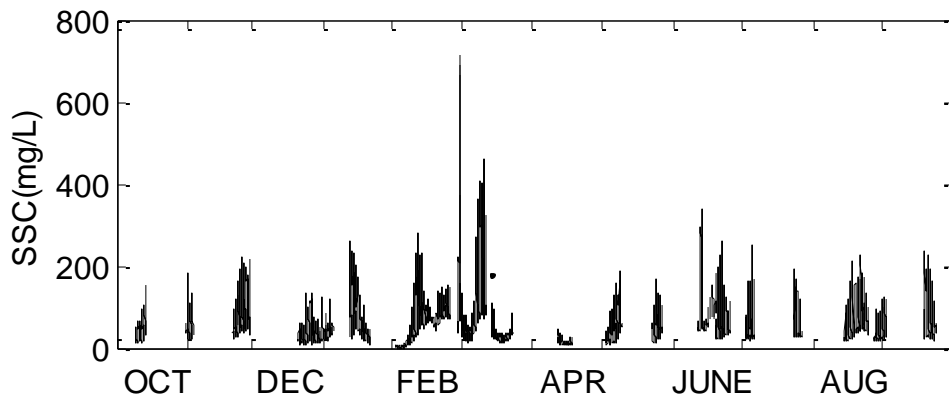
Percentage of Missing Data (%)	MARE			MRMSE		
	SSAM	ISSA	IMP (%)	SSAM	ISSA	IMP (%)
0	2.48	2.48	0	2.06	2.06	0%
10	2.87	2.60	9.41	3.68	3.38	2.21
20	3.26	2.73	16.26	4.19	3.56	15.04
30	3.71	2.90	21.83	4.76	3.78	20.59
40	4.22	3.11	26.30	5.42	4.07	24.91
50	4.57	3.17	30.63	5.89	4.14	29.71
60	5.37	3.52	34.45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

Obviously, if there is no missing data, the ISSA coincides with SSAM. If the percentage of missing data increases, both MARE and MRMSE will become larger. In Table 1, all the MARE and MRMSE of ISSA are smaller than those of SSAM. When the percentage of missing data reaches 50%, the MARE and MRMSE are 3.17 mg L^{-1} and 4.14 mg L^{-1} for ISSA, and 4.57 mg L^{-1} and 5.89 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 MARE and

238 MRMSE increase as well. Moreover, when the synthetic time series with the missing
239 data is same as the real SSC time series of Fig. 4, the IMPs of MARE and MRMSE
240 are 8.87% and 15.19%, respectively.

241 4. Performance of ISSA with real time series

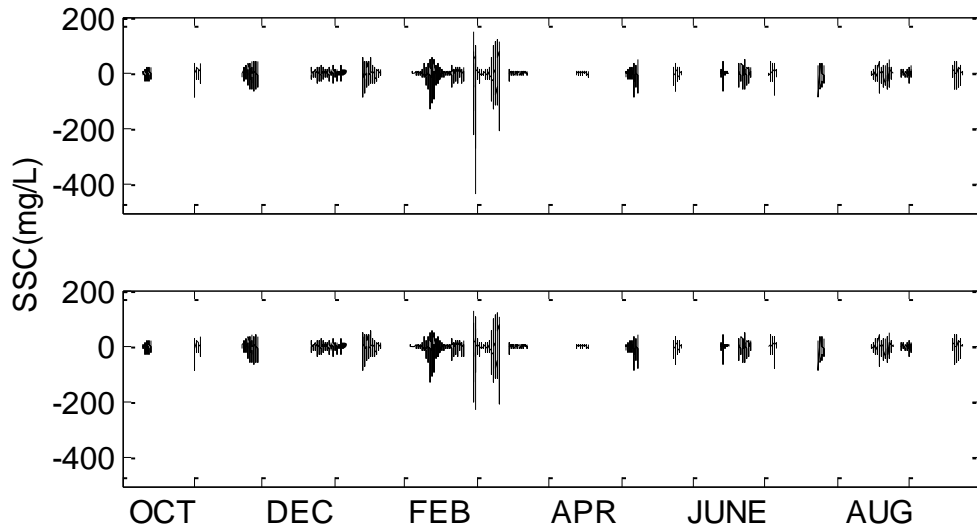
242 The mid-depth SSC time series at San Mateo Bridge is presented in Fig. 4, which
243 contains about 61% missing data. This time series was reported by Buchanan and
244 Schoellhamer (1999) and Buchanan and Ruhl (2000), and analyzed by Schoellhamer
245 (2001) using SSAM. We analyze this time series using our ISSA with the window
246 size of 30h ($L=120$) comparing with SSAM. The first 10 modes represent dominant
247 periodic components as shown in Schoellhamer (2001) which contain 89.1% of the
248 total variance. Therefore, we reconstruct the time series with first 10 modes when the
249 missing data in a window size is less than 50%.



250

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

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



260

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

263 Table 2: **Maximum, minimum** and mean absolute residuals of SSAM and ISSA

Residuals(mg L ⁻¹)	SSAM	ISSA
Maximum	145.05	126.61
Minimum	-432.20	-227.70
Mean absolute residuals	8.19	8.00
SD	13.48	12.27

264

265

266

267 5. Conclusions

268 We have developed the ISSA approach in this paper for processing the incomplete
 269 time series by using the principle that a time series can be **reproduced using** its
 270 principal components. We prove that the SSAM developed by Schoellhamer (2001) is
 271 a special case of our ISSA. The performances of ISSA and SSAM **are** demonstrated
 272 with a synthetic time series, and the results show that the relative errors of the first
 273 four principal components by ISSA are significantly smaller than those by SSAM. As
 274 the fraction of missing data increases, the improvement of the relative error becomes
 275 greater. When the percentage of missing data reaches 60%, the improvements of the
 276 first four principal components are up to 19.64, 41.34, 23.27 and 50.30%, respectively.
 277 Moreover, when the missing data accounts for 60%, the MARE and MRMSE derived
 278 by ISSA are 3.52 mg L⁻¹ and 4.60 mg L⁻¹, and by SSAM are 5.37 mg L⁻¹ and 6.96 mg
 279 L⁻¹. The corresponding improvements of ISSA with respect to SSAM are 34.45 and
 280 33.91%. When the missing data of synthetic time series is the same as the real SSC
 281 time series, the improvements of MARE and MRMSE are 8.87 and 15.19%,
 282 respectively. The SD derived from the real SSC time series at San Mateo Bridge by
 283 ISSA and SSAM are 12.27 mg L⁻¹ and 13.48 mg L⁻¹, and the squared correlation

284 coefficients between the observations and the reconstructed data from ISSA and
285 SSAM are 0.9178 and 0.9046, respectively. Therefore, ISSA can indeed, to a great
286 extent, retrieve the informative signals from the original incomplete time series.

287

288 **Author contribution**

289 Y. Shen proposes the improved singular spectrum analysis and F. Peng carries out the
290 FORTRAN program and performs the simulations. Y. Shen, F. Peng and B. Li prepare
291 the manuscript.

292

293 **Acknowledgements**

294 This work is sponsored by Natural Science Foundation of China (Projects: 41274035,
295 41474017) and partly supported by State Key Laboratory of Geodesy and Earth's
296 Dynamics (SKLGED2013-3-2-Z).

297

298 **References**

- 299 Broomhead, D.S., G.P. King, Extracting qualitative dynamics from experimental data.
300 *Physica D*, 20, 217-236, 1986.
- 301 Buchanan, P.A., and C.A Ruhl, Summary of suspended-solids concentration data, San
302 Francisco Bay, California, water year 1998, Open File Report 99-189, 41 pp.,
303 U.S. Geological Survey, 2000.
- 304 Buchanan, P.A., and D. H. Schoellhamer, Summary of suspended solids concentration
305 data, San Francisco Bay, California, water year 1997, Open File Report
306 00-88 URL <http://ca.water.usgs.gov/rep/ofr99189/>, 52 pp., U.S. Geological
307 Survey, 1999.
- 308 Golyandina, N., E. Osipov, The “Catterpillar”-SSA method for analysis of time series
309 with missing data, *J. Stat. Plan. Inf.*, 137, 2642-2653, 2007.
- 310 Hassani H., Mahmoudvand R., Zokaei M., et al. On the Separability between signal
311 and noise in singular spectrum analysis, *Fluct. Noise Lett.* 11(2), 1-11, 2012.
- 312 Hassani H., Mahmoudvand R. Multivariate singular spectrum analysis: a general view
313 and new vector forecasting approach, *Int. J. Energy Stat.*, 1(1), 55-83, 2013.
- 314 Kondrashov, D. M. Ghil, Spatio-temporal filling of missing points in geophysical data
315 sets, *Nonlin. Processes Geophys.*, 13, 151-159, 2006.
- 316 Oropeza, V., M. Sacchi, Simultaneous seismic data denoising and reconstruction via
317 multichannel singular spectrum analysis, *Geophysics*, 76(3), 25-32, 2011.
- 318 Robertson, A.W. and C. R. Mechoso, Interannual and decadal cycles in river flows of
319 southeastern South America, *Journal of Climate*, 11(10), 2570-2581, 1998.
- 320 Rodrigues, P.C., M. de Carvalho, Spectral modeling of time series with missing data,
321 2013
- 322 Schoellhamer, D.H., Factors affecting suspended-solids concentrations in South San
323 Francisco Bay, California, *J. Geophys. Res.*, 101(C5), 12087-12095, 1996.

324 Schoellhamer, D.H., Singular spectrum analysis for time series with missing data,
325 Geophys. Res. Lett. 28(16), 3187-3190, 2001.

326 Schoellhamer, D.H., Variability of suspended-sediment concentration at tidal to
327 annual time scales in San Francisco Bay, USA, Continental Shelf Research,
328 22, 1857-1866, 2002

329 Shen, Y., W. Li, G. Xu, B. Li. Spatiotemporal filtering of regional GNSS network's
330 position time series with missing data using principal component analysis,
331 Journal of Geodesy, DOI 10.1007/s00190-013-0663-y, Vol.88: 1-12, 2014

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

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

336 Wang, X.L., J. Corte-Real, and X. Zhang, Intraseasonal oscillations and associated
337 spatial-temporal structures of precipitation over China, J. Geophys. Res.,
338 101(D14), 19035-19042, 1996.

339 Yiou, P., K. Fuhrer, L.D. Meeker, J. Jouzel, S. Johnsen, and P.A. Masked,
340 Paleoclimatic variability inferred from the spectral analysis of Greenland
341 and Antarctic ice-core data, J. Geophys. Res., 102(C12), 26441-26454, 1997.

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

344