

ANALYZING GRACE/GRACE-FO MASS ANOMALY TIME SERIES USING EXTENDED SINGULAR SPECTRUM ANALYSIS

Kunpu Ji¹ and Yunzhong Shen*² ¹PhD Student, College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, People's Republic of China Email: <u>kunpuji@tongji.edu.cn</u>

²Professor, College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, People's Republic of China Email: <u>yzshen@tongji.edu.cn</u>

KEY WORDS: Singular Spectrum Analysis; GRACE/GRACE-FO; Mass Anomaly Time Series

ABSTRACT: The mass anomaly time series, derived from the monthly time-variable gravity fields of the Gravity Recovery and Climate Experiment (GRACE) and its follow-on mission (GRACE-FO), contain various deterministic signals such as long-term trends, annual and semi-annual signals. Singular spectrum analysis (SSA) is a data-driven method that effectively decomposes the time series into components ordered by variance intensity. However, the dozens of missing epochs in the monthly gravity product of GRACE and GRACE-FO impede the time series analysis. Besides, ordinary SSA does not consider the heterogeneity of mass anomaly time series (Level-3 products) inherited from the heterogeneity of the spherical harmonic coefficients (Level-2 products). To address these issues, we propose using extended singular spectrum analysis (ESSA) to extract the deterministic signals presented in the mass anomaly time series. Two well-studied areas (Qinghai-Tibet Plateau and Yangtze River Basin) are selected for analysis, utilizing the CSR RL06 mascon products and personal-computed mascon solutions from April 2002 to December 2022. Experimental results demonstrate that the ESSA successfully extracts dominant desired signals from the incomplete and heterogeneous mass anomaly time series. We compare the results from ESSA with those from improved SSA (ISSA) that handles incomplete time series by minimizing the quadratic norm of principal components using available observations. The comparison reveals that the ESSA approach outperforms the ISSA approach, exhibiting smaller fitting errors and higher root-mean-square ratios of extracted signals to residuals. Repeated simulations indicate that the signals extracted by ESSA are closer to the simulated true signals. Furthermore, the consideration of formal errors of time series indeed improves the filtering process. Notably, the ESSA approach also allows for retrospectively filling in the data gaps between the GRACE and GRACE-FO missions by taking the elements of extracted signals at the missing epochs.

1. INTRODUCTION

The Gravity Recovery and Climate Experiment (GRACE) and its subsequent GRACE-FO missions have been specifically designed to monitor the global gravity field with unprecedented spatiotemporal resolution (Tapley et al., 2004). The monthly gravity field models derived from these missions offer novel insights into various domains, including total water storage change (Syed et al., 2008; Landerer & Swenson, 2012), hydrological mass variations (Scanlon et al., 2016), sea-level change (Lombard et al., 2007; Yi et al., 2017), as well as polar and alpine ice melting (Baur & Sneeuw, 2011; Forsberg, 2017). The GRACE mission was operational from April 2002 until June 2017, whereas the GRACE-FO satellites were launched in early 2018 and initially provided data products in June of the same year. Consequently, there exists an 11-month gap between the GRACE and GRACE-FO missions, which poses a challenge to the continuous extraction of geophysical information.

The monthly mass anomaly time series derived from GRACE/GRACE-FO time-variable gravity fields contains abundant geophysical signals, such as long-term trends, annual oscillations, and semi-annual oscillations. These signals are highly correlated with various geophysical processes, such as precipitation, glacier melting, atmospheric pressure variations, and crustal activities (Chen et al., 2020; Velicogna et al., 2020). Singular spectrum analysis (SSA) is a data-driven method that can decompose the time series into several components ordered by variance (Vautard & Ghil, 1989). However, ordinary SSA assumes the time series is complete, which is invalid for GRACE/GRACE-FO observations. Although data gaps can be filled using interpolations beforehand (Wang et al., 2021; Yi & Sneeuw, 2021; Mo et al., 2022), interpolating large gaps may distort spectral peaks significantly (Kay & Marple, 1981). Another approach is to analyze incomplete time series using available data directly. Schoellhamer (2001) proposed the modified SSA (SSAM) to compute the principal components (PCs) with the observed data and the scalar factor related to the number of missing data; Shen et al. (2015) developed an improved SSA (ISSA) to compute the PCs by minimizing the quadratic form of PCs. It is proved that ISSA is a generalized version of SSAM. Recently, Ji et al. (2023) proposed the extended SSA (ESSA) approach, which can directly filter incomplete geodetic time series based on the optimal low-rank approximation. They demonstrated the superior performance of the ESSA approach over the ISSA approach with real Global Navigation



Satellite System (GNSS) position time series and synthetic time series. Moreover, the ESSA approach can also be used to process heterogeneous time series by considering the formal errors of time series. Actually, the mass anomaly time series, derived from the heterogeneous spherical harmonic coefficients (SHCs) (Shen et al., 2021), is typically heterogeneous. However, this heterogeneity has not been considered in previous signal analysis. In this paper, we employ the ESSA approach to analyze the incomplete and heterogeneous mass anomaly time series derived by time-variable gravity fields from GRACE/GRACE-FO measurements. The rest of the paper is organized as follows. Section 2 briefly introduces the ESSA approach; Section 3 presents the analytical results of the ESSA approach on two well-studied areas and compares them with ISSA. Some conclusions are summarized in Section 4.

2. METHODOLOGY

In this section, we provide a brief introduction to the extended SSA. For detailed mathematical aspects, please refer to Ji et al. (2023). Given a time series $x = \{x_i \mid 1 \le i \le N\}$, where x_i is the *i*th element and N is the length of x, we can form the trajectory matrix using a specified window size L,

$$\boldsymbol{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{i+1} & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{i+2} & \cdots & x_{K+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_L & x_{L+1} & \cdots & x_{i+L} & \cdots & x_M \end{bmatrix}$$
(1)

with K = N - L + 1. If the time series x contains missing data, the lag-covariance matrix C can only be constructed using the available data,

$$\boldsymbol{C} = \begin{bmatrix} c(0) & c(1) & \cdots & c(L-1) \\ c(1) & c(0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & c(1) \\ c(L-1) & \cdots & \cdots & c(0) \end{bmatrix}$$
(2)

with the element c(j) computed by,

$$c(j) = \frac{1}{N_j} \sum_{i \le N-j} x_i x_{i+j} \quad j = 0, 1, 2, \cdots, L-1$$
(3)

where both x_i and x_{i+j} must be observed rather than missed, and N_j is the number of the products of x_i and x_{i+j} within the sample index $i \le N - j$. By performing the eigenvalue decomposition (SVD) on C, we have,

$$\boldsymbol{C} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T \tag{4}$$

where V is an $L \times L$ orthogonal eigenvector matrix, Λ is the diagonal matrix comprised with the eigenvalues λ_i ($1 \le i \le L$) sorting in descending order. With the elements of V and X, one can reconstruct the time series at different scales. Ji et al. (2023) formulated the projection from the original time series to the reconstructed components (RCs) in matrix form as,

$$\boldsymbol{R}\boldsymbol{C}^{k} = [\boldsymbol{R}\boldsymbol{C}_{1}^{k} \quad \boldsymbol{R}\boldsymbol{C}_{2}^{k} \quad \cdots \quad \boldsymbol{R}\boldsymbol{C}_{N}^{k}]^{T} = \boldsymbol{B}^{k}\boldsymbol{x}$$

$$\tag{5}$$

where RC^k represents the *k*th RC, B^k is an $N \times N$ projection matrix whose elements solely dependent on the elements of V. One may refer to Ji et al. (2023) for detailed instructions on constructing B^k . The RCs corresponding to the larger eigenvalues represent the signals, while those associated with the smaller eigenvalues represent the noise. Suppose that the first leading *d* RCs represent the signals as *s* and the remaining RCs are residuals as *e*, we have,

$$s = Bx = B_1 x_1 + B_2 x_2 \tag{6}$$

$$\boldsymbol{e} = \boldsymbol{D}\boldsymbol{x} = \boldsymbol{D}_1 \boldsymbol{x}_1 + \boldsymbol{D}_2 \boldsymbol{x}_2 \tag{7}$$

where $B = \sum_{k=1}^{d} B^k$ and $D = I_N - B$; $x_1 = \{x_i \mid i \in S\}$ and $x_2 = \{x_i \mid i \in \overline{S}\}$ are the available data and missing data in x, S and \overline{S} are the index sets of available data and missing data, respectively; B_1 and B_2 are the sub-matrices of B with the column vectors related to the available epochs and missing epochs, D_1 and D_2 are the sub-matrices of D composed of the column vectors related to the available epochs and missing epochs, respectively. The missing data x_2 in time series can be estimated by minimizing the quadratic norm of e, i.e.,

$$\widehat{\boldsymbol{x}}_2 = \underset{\boldsymbol{x}_2}{\operatorname{argmin:}} \boldsymbol{e}^T \boldsymbol{e} = -(\boldsymbol{D}_2^T \boldsymbol{D}_2)^{-1} \boldsymbol{D}_2^T \boldsymbol{D}_1 \boldsymbol{x}_1$$
(8)

If the formal errors of time series are provided, they should be taken into account. Suppose the formal error for x_i is σ_i , then we can derive the weighted estimates of \hat{x}_2 with,

$$\widehat{\mathbf{x}}_2 = \underset{\mathbf{x}_2}{\operatorname{argmin:}} \mathbf{e}^T \mathbf{P} \mathbf{e} = -(\mathbf{D}_2^T \mathbf{P} \mathbf{D}_2)^{-1} \mathbf{D}_2^T \mathbf{P} \mathbf{D}_1 \mathbf{x}_1$$
(9)

where P is the diagonal weight matrix with the diagonal elements given by,

$$p_i = \begin{cases} \sigma_0^2 / \sigma_i^2 & i \in S \\ 1 & i \in \overline{S} \end{cases}$$
(10)

By substituting Eq. (8) or Eq. (9) into Eq. (6), we can obtain the signals. The missing data can be interpolated with the values of s in the corresponding epochs.



3. EXPERIMENTS AND RESULTS ANALYSIS

The RL06 mascon products $(0.25^{\circ} \times 0.25^{\circ})$, developed by the Center for Space Research (CSR), are used in this study. The research area includes Qinghai-Tibet Plateau (TP) and Yangtze River Basin (YRB). The period covered is from April 2002 to December 2022. Figure 1 presents the mass anomaly time series of TP and YRB areas, in equivalent water storage (EWH) form (top row), associated with their power spectrums (bottom row) computed with the Lomb-Scargle algorithm. It shows that the time series contains a noticeable long-term trend, annual oscillations, and other smaller periodical oscillations. Before applying the ESSA approach to extract these signals, it is necessary to determine the approximate window size. There is no universally accepted criteria for determining the window size. Golyandina (2010) recommended that the window size be proportional to the period, but not exceed half the length of the series. In this study, we have opted for a window size of 24 months, which is twice the annual period. Figure 2 presents the w-correlations between the first leading 24 RCs and the corresponding variance contribution percentage of the associated PCs. The findings indicate that when i, j > 5, the correlations between the neighbouring RC_i and RC_j are relatively small, and the first 5 leading PCs capture 93.4% and 88.2% variance. Therefore, the reconstruction order is chosen as 5. Figure 3 presents the first leading 5 RCs, in which RC1 represents the long-term trend, RC2 and RC3 represent the annual signal, RC4 and RC5 represent the semi-annual signal, respectively. We then reconstruct the time series by summarizing $RC1 \sim RC5$ and interpolate the missing data using the elements of reconstructed signals in missing epochs. The interpolated results are depicted in Fig. 4, from which we observe that the interpolated data aligns well with the available observations.

We compare the ESSA approach with the ISSA approach (Shen et al., 2015), another SSA-based method to directly handle incomplete time series. To evaluate the effectiveness of these two approaches, we compute the fitting errors $\hat{\sigma}$ and RMS ratios \hat{r} of the extracted signals over residuals with Eq.(11) and Eq.(12),

$$\hat{\sigma} = \sqrt{\frac{1}{N_S} \sum_{i \in S} (x_i - s_i)^2} \tag{11}$$

$$\hat{r} = \sqrt{\sum_{i \in S} s_i^2} / \sqrt{\sum_{i \in S} (x_i - s_i)^2}$$
(12)

where N_s is the length of available data, x_i and s_i are the *i*th element of x and s. A smaller fitting error or a larger RMS ratio indicate a stronger extraction of signals. For TP area, the fitting errors and RMS ratio are 0.7109 cm and 4.8729 for ESSA approach, and 0.7213 cm and 4.7738 for ISSA approach. For YSB area, the fitting errors and RMS ratio are 1.4753 cm and 2.9481 for ESSA approach, and 1.4931 cm and 2.8977 for ISSA approach. To intuitively compare the two approaches, we analyze the time series of each point in regional grid. Due to space limitations, we only present the computed results for TP area. Figure 5 presents the fitting errors and RMS ratios for the two approaches, where (a) and (b) represent for $\hat{\sigma}_{ESSA}$ and \hat{r}_{ESSA} , respectively, while (c) and (d) represent for the corresponding improvements of ESSA over ISSA in reducing fitting error $(\hat{\sigma}_{ISSA} - \hat{\sigma}_{ESSA})/\hat{\sigma}_{ISSA} \times 100\%$ and enhancing RMS ratio $(\hat{r}_{ESSA} - \hat{r}_{ISSA})/\hat{r}_{ISSA} \times 100\%$ 100%, respectively. Figure 6 presents the linear trend, annual and semi-annual amplitudes derived by least squares fitting of each time series of grid point. The results indicate that while the ESSA method is slightly inferior to the ISSA approach in a few areas with less significant signals, it significantly outperforms ISSA in regions with stronger signals. The mean and maximum improvements in terms of fitting errors are 2.87% and 22.03%, respectively, and in terms of RMS ratios are 5.10% and 34.72%. To further compare the two approaches, we conduct 500 simulations. In each simulation, we regard the signals extracted by the ESSA approach from real data as true signals \overline{s} , and then generate random noise ewith zero mean and unit weight variance derived from the real data. As a result, we obtain the full noisy time series x = $\overline{s} + e$. We delete the time series according to the locations of missing epochs in real data. Since the true signals are available here, we can use Eq. (13) to calculate the root mean square errors (RMSEs) of the extracted signals,

$$RMSE = \sqrt{\frac{1}{N_S} \sum_{i \in S} (s_i - \overline{s}_i)^2}$$
(13)

The results are presented in Fig.7. It is evident from the results that the signals obtained through the ESSA approach are closer to those obtained through the ISSA approach. The average RMSE is reduced from 0.3433 cm for ESSA to 0.3299 cm for ISSA, representing a reduction of 3.90%.

Since the CSR mascon solutions do not provide the formal errors, we construct the mascon equation based on the pointmass methodology (Forsberg, 2007; Baur & Sneeuw, 2011). The ITSG-Grace2018 and ITSG-Grace_op spherical harmonic coefficients (SHCs) products are used since they have achieved state-of-the-art performance among existing solutions. They also provide the full covariance matrices of the SHCs, which are essential in point-mass inversion modelling (Chen et al., 2016). The study area is TP and a rectangle grid with $1^{\circ} \times 1^{\circ}$ is constructed. The period is from April 2002 to December 2022. The point-mass modelling employs a linear equation to relate gravitational disturbances in space derived from GRACE/GRACE-FO SHCs to the variations of individual point-mass on the Earth's surface (Baur & Sneeuw, 2011). Since the equation is typically ill-posed, the Tikhonov regularization approach is commonly utilized



to solve for the mass anomalies. The square roots of the diagonal elements of the mean square error (MSE) matrix are considered as the formal errors of the solutions. We then apply weighted ESSA to filter the mass anomaly time series, considering the formal errors, and compare the fitting errors and RMS ratios with those by unweighted ESSA. The statistical results are presented in Fig.8, where panels (a) and (b) denote the fitting errors for ISSA and ESSA, panels (c) and (d) denote the RMS ratios for ISSA and ESSA, respectively. The results clearly show that considering formal errors can enhance the filtering process. Compared to the unweighted ESSA, the fitting error for ESSA is reduced by an average of 21.81%, while the RMS ratio is improved by an average of 27.18%. Similar to the unweighted case, we carry out the simulation experiments. In each simulation, the random errors are generated under the normal distribution with zero mean and a diagonal variance matrix with the squares of the formal errors. The computed RMSEs with Eq. (13) are presented in Fig.9. The results show that the weighted ESSA approach performs better than the ESSA approach, with the RMSE of extracted signals reduced by an average of 1.24% and a maximum of 5.10%.



Figure 1. The mass anomaly time series and power spectrums of the TP area and YRB area. (a) and (b): mass anomaly time series. (c) and (d): power spectrums



Figure 2. The w-correlations of the first 24 leading RCs (top row) and the variance contribution percentage of the associated PCs (bottom row)



Figure 3. The first leading 5 reconstructed components





Figure 5. The fitting errors $\hat{\sigma}$ and RMS ratios \hat{r} . (a): $\hat{\sigma}_{ESSA}$, (b): \hat{r}_{ESSA} , (c): $(\hat{\sigma}_{ISSA} - \hat{\sigma}_{ESSA})/\hat{\sigma}_{ISSA} \times 100\%$, (d): $(\hat{r}_{ESSA} - \hat{r}_{ISSA})/\hat{r}_{ISSA} \times 100\%$



Figure 6. The linear trend, annual amplitude, and semi-annual amplitude. (a): linear trend, (b): annual amplitude, (c): semi-annual amplitude.



Figure 7. RMSEs of signals extracted by ISSA and ESSA approaches for 500 simulations



Figure 8. Fitting errors and RMS ratios. (a) and (b) are the fitting errors for ISSA and ESSA, respectively, (c) and (d) are the RMS ratios for ISSA and ESSA, respectively.



Figure 9. RMSEs of signals extracted by ESSA and weighted ESSA approaches for 500 simulations.

4. CONCLUSIONS

The monthly mass anomaly time series, derived from GRACE/GRACE-FO time-variable gravity fields, contains missing values and exhibits visible heterogeneity inherited from the Level-2 spherical harmonic coefficients. We introduce the ESSA approach to extract the deterministic signals, including long-term trends, annual and semi-annual oscillations, from incomplete and heterogeneous mass anomaly time series. Real data analysis in the TP and YSB areas demonstrates the ESSA approach's applicability for signal extraction and filling in data gaps in time series. We also compare the ESSA approach with the ISSA approach using real and synthetic time series. The comparison shows that the ESSA approach outperforms the ISSA approach. In addition, we find that considering the formal errors can further improve the filtering process.



5. ACKNOWLEDGMENTS

This study is sponsored by the National Natural Science Foundation of China (41974002, 42192532, and 42274005). The CSR RL06 mascon products are from <u>https://www2.csr.utexas.edu/grace/RL06 mascons.html</u>. The ITSG-Grace2018/ ITSG-Grace_op Level-2 products and the corresponding full variance-covariance matrices are from <u>http://ftp.tugraz.at/outgoing/ITSG/GRACE/</u>.

6. REFERENCES

Baur, O., & Sneeuw, N. (2011). Assessing Greenland ice mass loss by means of point-mass modeling: a viable methodology. *Journal of Geodesy*, 85(9), 607-615. https://doi.org/10.1007/s00190-011-0463-1

Chen, J. L., Tapley, B., Rodell, M., Seo, K. W., Wilson, C., Scanlon, B. R., & Pokhrel, Y. (2020). Basin-Scale River Runoff Estimation From GRACE Gravity Satellites, Climate Models, and In Situ Observations: A Case Study in the Amazon Basin. *Water Resources Research*, 56(10).

Chen, T. Y., Shen, Y. Z., & Chen, Q. J. (2016). Mass Flux Solution in the Tibetan Plateau Using Mascon Modeling. *Remote Sensing*, 8(5).

Forsberg, R., & Reeh N. (2007). Mass change of the Greenland ice sheet from GRACE. In: Proceedings of the 1st international symposium of IGFS, Harita Dergisi, vol 18, pp 454–458. ftp://ftp.spacecenter. dk/pub/igfs2006-proceedings/IGFS 2006 Proceedings.pdf.

Forsberg, R., Sorensen, L., & Simonsen, S. (2017). Greenland and Antarctica Ice Sheet Mass Changes and Effects on Global Sea Level. *Surveys in Geophysics*, 38(1), 89-104.

Golyandina, N. (2010). On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods. *Statistics and Its Interface*, 3(3), 259-279.

Ji, K. P., Shen, Y. Z., Chen, Q. J., & Wang, F. W. (2023). Extended singular spectrum analysis for processing incomplete heterogeneous geodetic time series. *Journal of Geodesy*, 97(8).

Kay, S. M., & Marple, S. L. (1981). Spectrum Analysis - a Modern Perspective. Proceedings of the Ieee, 69(11), 1380-1419.

Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, 48.

Lombard, A., Garcia, D., Ramillien, G., Cazenave, A., Biancale, R., Lemome, J. M., Flechtner, F., Schmidt, R., & Ishii, M. (2007). Estimation of steric sea level variations from combined GRACE and Jason-1 data. *Earth and Planetary Science Letters*, 254(1-2), 194-202.

Mo, S. X., Zhong, Y. L., Forootan, E., Mehrnegar, N., Yin, X., Wu, J. C., Feng, W., & Shi, X. Q. (2022). Bayesian convolutional neural networks for predicting the terrestrial water storage anomalies during GRACE and GRACE-FO gap. *Journal of Hydrology*, 604.

Scanlon, B. R., Zhang, Z. Z., Save, H., Wiese, D. N., Landerer, F. W., Long, D., Longuevergne, L., & Chen, J. l. (2016). Global evaluation of new GRACE mascon products for hydrologic applications. *Water Resources Research*, 52(12), 9412-9429.

Schoellhamer, D. H. (2001). Singular spectrum analysis for time series with missing data. *Geophysical Research Letters*, 28(16), 3187-3190.

Shen, Y., Peng, F., & Li, B. (2015). Improved singular spectrum analysis for time series with missing data. *Nonlinear Processes in Geophysics*, 22(4), 371-376.

Shen, Y. Z., Wang, F. W., & Chen, Q. J. (2021). Weighted multichannel singular spectrum analysis for post-processing GRACE monthly gravity field models by considering the formal errors. *Geophysical Journal International*, 226(3), 1997-2010.

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., & Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research*, 44(2).



Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, 305(5683), 503-505.

Vautard, R., & Ghil, M. (1989). Singular Spectrum Analysis in Nonlinear Dynamics, with Applications to Paleoclimatic Time-Series. *Physica D*, 35(3), 395-424.

Velicogna, I., Mohajerani, Y., Geruo, A., Landerer, F., Mouginot, J., Noel, B., Rignot, E., Sutterley, T., van den Broeke, M., van Wessem, M., & Wiese, D. (2020). Continuity of Ice Sheet Mass Loss in Greenland and Antarctica From the GRACE and GRACE Follow-On Missions. *Geophysical Research Letters*, 47(8).

Wang, F. W., Shen, Y. Z., Chen, Q. J., & Wang, W. (2021). Bridging the gap between GRACE and GRACE follow-on monthly gravity field solutions using improved multichannel singular spectrum analysis. *Journal of Hydrology*, 594.

Yi, S., Heki, K., & Qian, A. (2017). Acceleration in the Global Mean Sea Level Rise: 2005-2015. *Geophysical Research Letters*, 44(23), 11905-11913.

Yi, S., & Sneeuw, N. (2021). Filling the Data Gaps Within GRACE Missions Using Singular Spectrum Analysis. *Journal of Geophysical Research-Solid Earth*, 126(5).