



HAL
open science

Gap-filling based on iterative EOF analysis of temporal covariance: application to InSAR displacement time series

Alexandre Hippert-Ferrer, Y. Yan, P Bolon

► **To cite this version:**

Alexandre Hippert-Ferrer, Y. Yan, P Bolon. Gap-filling based on iterative EOF analysis of temporal covariance: application to InSAR displacement time series. IGARSS, Jul 2019, Yokohama, Japan. hal-02178695v1

HAL Id: hal-02178695

<https://hal.science/hal-02178695v1>

Submitted on 10 Jul 2019 (v1), last revised 23 Mar 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gap-filling based on iterative EOF analysis of temporal covariance : application to InSAR displacement time series.

A. Hippert-Ferrer *, *Student Member, IEEE*, Y. Yan *, *Member, IEEE*, P. Bolon *, *Member, IEEE*

*LISTIC, Université Savoie Mont-Blanc, Annecy, France

alexandre.hippert-ferrer@univ-smb.fr

Abstract—An iterative method, namely EM-EOF (Expectation Maximization-Empirical Orthogonal Functions) is proposed for the first time to retrieve missing values in InSAR displacement time series. The method decomposes the temporal covariance into different EOF modes by solving the eigenvalue problem, and then selects an optimal number of EOF modes to reconstruct the time series. After an appropriate initialization of missing values, the proposed method performs (i) a cross-validation error minimization to find an estimate of the optimal number of EOF modes used in the reconstruction and (ii) an iterative update of missing values which gives the best estimate of missing data points according to the cross-validation error. Results on a time series of Sentinel-1 A/B unwrapped interferograms over the Gorner glacier from November 2016 to March 2017 demonstrate the high efficiency of the proposed method at retrieving missing values and denoising the time series, even in case of time series with limited size and spatio-temporally correlated gaps.

Index Terms—Gap filling, EOFs, InSAR, displacement, time series

I. INTRODUCTION

TIME series analysis of remotely sensed images is an essential step to better understand some physical phenomena observed on Earth, especially since the recent development of online platforms allowing an easier access to satellite-derived products.

Differential interferometry (DInSAR) and offset tracking are well-developed techniques that have experienced continuous improvements over time to compute ground displacements maps from Synthetic Aperture Radar (SAR) images [1], [2]. Regardless of their capability to produce precise ground displacements estimations with enhanced spatial coverage, time series of displacement measurement derived from SAR images can suffer from missing data in space and time dimensions. Data gaps can occur for various reasons including raw data quality and/or technical limitations of the ground displacement computation method itself.

Of course, the missing data problem is not new [3]. In remote sensing, it is a well-documented problem, especially in optical and infrared satellite images where data quality is strongly dependent on cloud coverage [4–6]. An important framework of gap-filling methods already exists in ocean-atmosphere science and hydrology [7–10].

Despite the large variety of domains concerned with missing data, there has not been, to our knowledge, any study which fully focused on developing methods which can handle

missing data in SAR-derived products such as interferograms and offset time series. When analyzing displacement maps computed from SAR images, one has to consider various noises, besides the displacement signal itself. An efficient method aiming at filling in gaps in SAR-derived displacement maps should thus be able to handle noise and displacement signal complexity, e.g. noise correlated at different scales in both time and space dimensions, and multiple behaviors of displacement (linear, oscillatory, etc.).

From the existing methods (see [11] for a review), Empirical Orthogonal Functions (EOFs) have been used to infer missing data and to predict spatio-temporal signals [7], [12], thanks to their advantages such as ease of implementation, high efficiency and needless a priori information. The EOF analysis is based on the eigenvalue decomposition of the covariance of a time series into orthogonal functions, and allows a representation of the signal into different variability modes sometimes called trends, oscillatory patterns and noise [13], [14]. By initializing the missing points to a relevant value and by selecting an appropriate number of EOF modes used in the reconstruction, one can extract dominant features of a time series and retrieve the values of the initially missing points. In InSAR displacement measurement, the EOF analysis has recently been used for the first time to denoise and extract displacement signal from a time series of Sentinel-1 A/B interferograms over the Gorner glacier [15]. Spectacular results have been obtained, which confirms the efficiency of EOF-based methods in the analysis of InSAR displacement time series. Therefore, it also seems promising to use the EOF analysis to fill in data gaps in InSAR displacement time series.

In this paper, we propose a data-adaptive, iterative EOF-based, gap-filling method adapted for InSAR displacement time series with complex displacement signal and noise behaviors. We call the method EM-EOF for Expectation Maximization-EOF. Starting from an appropriate initialization, missing values are first updated in an iterative way given the observed values (Expectation step), and then the error between initial and reconstructed fields is minimized with the current estimate of missing values (Maximization step). The method is organized in two steps, with each of them using the described EM principle. In order to deal with the case where no ground truth is available (a frequent issue in displacement measurement), we propose a cross-validation technique to compute errors. The EM-EOF method is applied to a time

series of Sentinel-1 A/B interferograms over the Gorner glacier between November 2016 and March 2017 where data gaps are observed.

II. METHODOLOGY

The general principle of EM-EOF is given in the diagram in Fig. 1. Missing values are initialized by the appropriate value (section B). Then, the temporal covariance of the time series is formed and decomposed into EOF modes. The appropriate number of EOF modes to reconstruct the time series is estimated by minimizing the error between initial and reconstructed field (section C), and a refinement of the estimation of missing values based on an iterative update is performed (section D). Finally, the field is reconstructed with the optimal number of EOF modes and the updated missing values (section E).

A. Theoretical background

Let us suppose that we have a data grid composed by a spatio-temporal field X which contains missing data. The field can be written in matrix form:

$$X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix} \quad (1)$$

where each column \mathbf{x}_t is an observation at a given time t , ($t = 1, \dots, n$) and each line is a time series of a given grid point s (with $s = 1, \dots, p$). The spatial mean of the field at each time (i.e. mean of each column) is subtracted to form the *spatial anomaly* X' :

$$X' = X - \mathbf{1}_n \bar{X} \quad (2)$$

where $\mathbf{1}_n$ is the line vector of length n containing only ones and $\bar{X} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ is a line vector with each \bar{x}_t being the spatial mean of the observation at time t .

The *sample temporal covariance matrix* \hat{C} of X is then given by:

$$\hat{C} = \frac{1}{p-1} X'^T X' \quad (3)$$

The eigenvectors of the matrix \hat{C} can simply be found by resolving the eigenvalue equation:

$$\hat{C}U = U\Lambda \quad (4)$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ contains the eigenvalues λ of matrix \hat{C} in decreasing order, with each indicating the fraction of total variance explained by the corresponding EOF mode. Each column \mathbf{u}_i of U is a eigenvector of \hat{C} corresponding to the eigenvalue λ_i . Each eigenvector is orthogonal to each other, hence the name EOF. Each EOF mode describes the spatio-temporal variability of the spatial anomaly of the field [14].

X' can be totally reconstructed by summing the Principal Components (PCs) a_i multiplied by the EOFs:

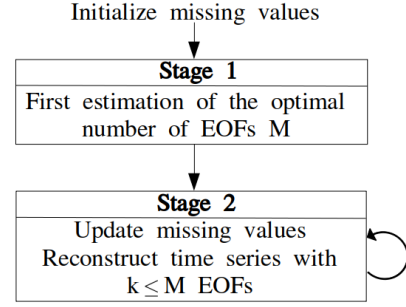


Fig. 1. Diagram of the 2-steps procedure of the EM-EOF method. Data gaps are first initialized. Step 1 estimates the optimal number of EOF modes M for reconstruction which minimizes the error function. Step 2 is a refinement step which iteratively uses updates of missing values to reconstruct the time series.

$$\hat{X}' = \sum_{i=1}^n a_i \mathbf{u}_i^t \quad (5)$$

where $a_i = X' \mathbf{u}_i$ is a spatial pattern map associated with each EOF mode \mathbf{u}_i . The PCs refer to the spatial modes of variability of the time series and the EOFs show how these modes oscillate in time.

B. Initialization of missing values

Initialization of missing values is a key parameter of the EM-EOF method because it impacts the estimation of the temporal covariance matrix \hat{C} and thus the computation of the EOFs. The value of initialization should be set in accordance to the distribution of the observed values, because it can lead to a biased estimate of the optimal number of EOF modes in step 1 and can increase the time of convergence [16] of the iterative update in step 2. However, synthetic tests have shown that no significant difference is observed on the reconstruction performance between a initialization of X by the spatial mean or by the spatial mean plus a Gaussian noise. We therefore choose to initialize missing values by the spatial mean as it avoid any bias in the computation of the anomaly matrix.

C. Step 1 : find the optimal number of EOF modes

In this step, we perform the computation of equations (3), (4) and reconstruct the time series as in equation (5) by successively adding one EOF mode at a time. At each step, we compute a cross-validation root-mean-square error (cross-RMSE) [17] which estimates the reconstruction error :

$$E(k) = \left[\frac{1}{N} \sum_{i=1}^N |\hat{\mathcal{X}}_k - \mathcal{X}|^2 \right]^{1/2} \quad (6)$$

where $\mathcal{X} = \{x_i\}_{1 \leq i \leq N}$ is the cross validation data subset of X which contains N points randomly chosen in space and time among existing data. These points are set as artificial missing data with their values set aside. After each reconstruction with k EOF modes, the cross validation data set \mathcal{X} is compared to the new estimated set $\hat{\mathcal{X}}_k$. The number of points N must be neither too small nor too large : a small N cannot provide

a good statistical representation of the time series whereas a large N can affect the reconstruction error since the quantity of missing data is increased. The advantage of cross-RMSE lies in its independence of the true field if it exists, which makes it a robust indicator of the reconstruction error when dealing with real data with limited access to ground truth. Each error $E(k)$, $k = 1, \dots, n$ between the time series reconstructed using k EOF modes is computed, and we then find the optimal number M which minimizes the error function $E(k)$:

$$M = \arg \min_{M \in [1, n]} E(k) \quad (7)$$

Synthetic tests have shown that strong correlated (in both space and time dimensions) noise present in the data can lead to an over-estimation of M , mainly because correlated noise will add significant variance to secondary modes, resulting in a selection of more EOF modes in the reconstruction.

D. Step 2 : update missing values

Step 2 is a refinement step : equations (3), (4) and (5) are computed in an iterative loop. At iteration i , the reconstruction is performed using the missing values obtained at iteration $i - 1$. The cross-RMSE is computed at each iteration. Once the convergence is obtained, i.e. the difference $\Delta E = E_i(k) - E_{i-1}(k)$ is smaller than a predefined convergence value β , one more EOF mode is added to the reconstruction and the iteration process starts again with the added EOF mode. If the error starts to increase, the procedure stops, otherwise it keeps going until the number of EOF modes, M , is reached.

E. Reconstruction of the field

The final step consists in reconstructing the time series using the estimated optimal number of EOF modes M at step 1 and the updated missing values found at step 2 :

$$\hat{X}' = \sum_{i=1}^M a_i \mathbf{u}_i^t \quad (8)$$

which is simply a truncation of equation (5) by its first M terms. To get the reconstructed field \hat{X} , we finally add the spatial mean back to the anomaly :

$$\hat{X} = \hat{X}' + \mathbf{1}_n \bar{X} \quad (9)$$

III. APPLICATION TO A TIME SERIES OF INCOMPLETE UNWRAPPED INTERFEROGRAMS

17 Sentinel 1 A/B SAR images acquired from November 2016 to March 2017 were available for displacement measurement over the Gornier glacier located in the canton of Valais, Switzerland, and from which 16 interferograms were computed from consecutive acquisitions with an interval of 6 days. All interferograms contain missing data because of coherence loss due to rapid glacier flow and/or snow falls. The quantity of data gaps per interferogram ranges from 11.8 to 27.4 %, and the time series also contains a missing interferogram because of total coherence loss. The signal to noise ratio (SNR) varies from one interferogram to another, with sometimes significant noise. The relatively small

size of the time series (16 displacement maps), the spatio-temporal distribution of data gaps and the presence of possible significant noise, and a total missing interferogram in the time series altogether make the reconstruction challenging.

The EM-EOF method is then applied to this time series of interferograms. Examples of the reconstruction are presented in figure 2 : case 1 (first line) contains 14.6% of data gaps, case 2 (second line) is the reconstruction of the missing interferogram (100% of data gaps) and case 3 (third line) contains 27.4% of data gaps. Missing values are initialized by the spatial mean. We choose to set the number of cross-validation points to 1% of the total observed points per interferogram, thus making it proportionally equal at all times. The optimal number of EOF modes to reconstruct the time series was found to be 3.

In case 1, the displacement signal shows consistent patterns in missing data areas with the overall displacement pattern, but with a smoother texture. In observed areas, the reconstruction does not degrade the displacement pattern. The center part of the glacier shows small zero-centered residuals, which demonstrates the capacity of the method to retrieve missing values and to reduce noise represented by higher order EOF modes. Large residuals are observed close to the glacier edges because localized phase unwrapping errors exist in the original interferogram in transition zones between static rock and ice in motion.

To reconstruct the missing interferogram (case 2), the spatial mean is replaced by the temporal mean given by $\bar{X} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)^T$. \bar{X} is added as in equation (9), which means that the displacement pattern is mainly determined from the temporal mean of the time series. As some grid points are only observed a few times in the time series, their temporal mean is statistically less representative than the theoretical temporal mean, which can potentially lead to a biased reconstructed value.

The reconstruction of case 3 also shows a consistent displacement pattern with the initial field. On the original interferogram, data gaps (due to low coherence) induced phase unwrapping errors in some localized area, which results in discontinuities of displacement values in the residuals, where large values are mainly observed. However, no displacement signal can be identified in the residual.

Despite the high efficiency of the EM-EOF method, note however that this method is not sensitive to displacement signals related to unique and local events. Therefore, this part of the signal cannot always be retrieved in the reconstruction, which constitutes one limitation of the method.

IV. CONCLUSION

EM-EOF is an iterative method to retrieve missing values in a time series of SAR-derived displacement measurement. This method does not need any a priori information, is data adaptive and has a high computational efficiency since it only needs a few eigenvalues to fill in the missing data. Satisfactory results are obtained in the case of the Gornier glacier for retrieving consistent displacement patterns in missing data areas with no degradation of the observed parts. In particular, the efficiency

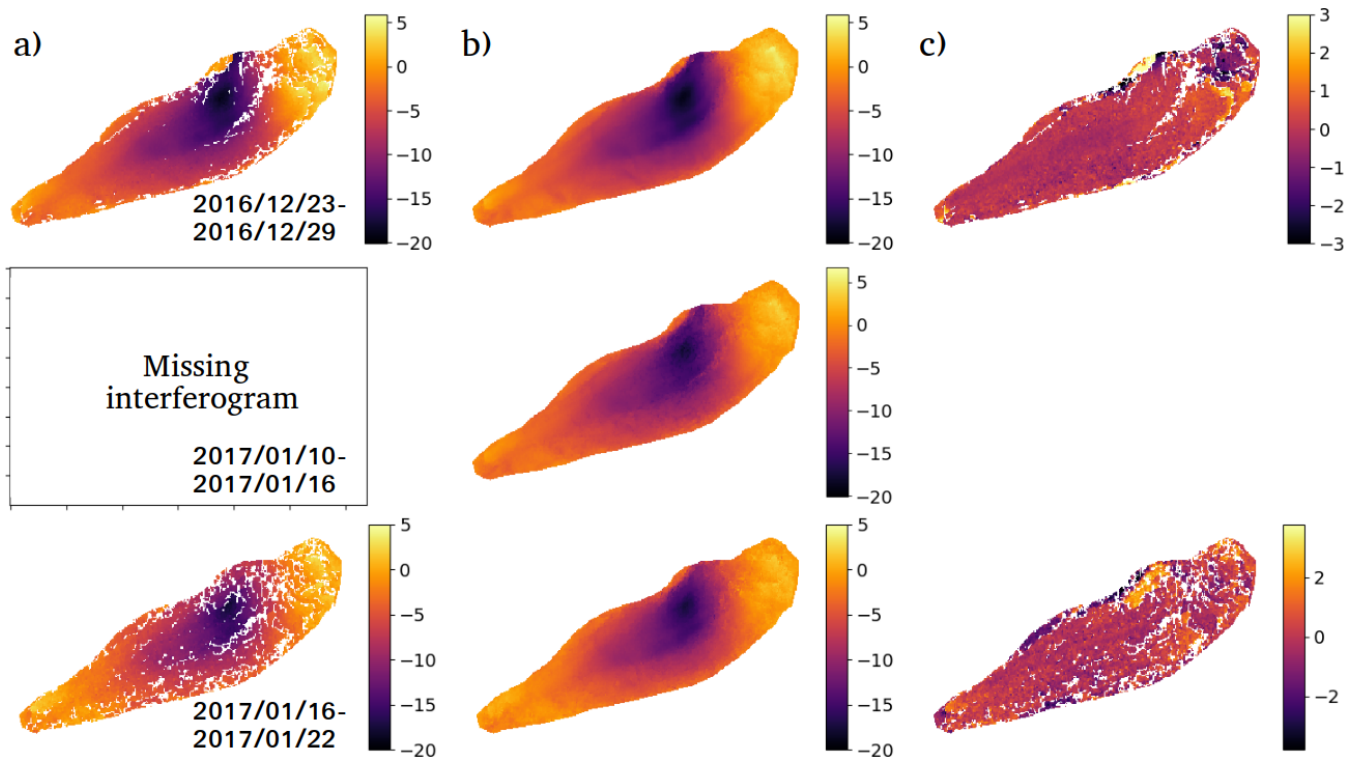


Fig. 2. a) Original b) reconstructed interferograms and c) residuals (original - reconstructed) in radar geometry at three time intervals (2016/12/23-2016/12/29, 2017/01/10-2017/01/16, 2017/01/16-2017/01/22). Displayed values are in centimeters in the radar line-of-sight (LOS).

of the method is highlighted for reconstructing a completely missing interferogram in the time series by working on the temporal mean instead of the spatial mean. The EM-EOF method can greatly help to increase the effective size of the time series, and thus to facilitate the understanding of the phenomenon under observation, particularly if low coherence in interferograms is a frequent issue. In the future work, instead of the temporal covariance, a spatio-temporal covariance will be taken into account in order to better characterize the displacement behaviour in the time series, e.g. reconstruct more propagating structures in space and time in the signal.

REFERENCES

- [1] T. Strozzi, A. Luckman, T. Murray, U. Wegmuller, and C. L. Werner, "Glacier motion estimation using SAR offset-tracking procedures," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2384–2391, Nov 2002.
- [2] A. Ferretti, C. Prati, and F. Rocca, "Permanent Scatterers in SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 8–20, Jan 2001.
- [3] R. J. A. Little and D. B. Rubin, *Statistical analysis with Missing Data*. New York: Wiley, 1987.
- [4] F. Melgani, "Contextual Reconstruction of Cloud-Contaminated Multitemporal Multispectral Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pp. 442–455, Feb. 2006.
- [5] C.-H. Lin, K.-H. Lai, Z.-B. Chen, and J.-Y. Chen, "Patch-based information reconstruction of cloud-contaminated multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 163–174, 2014.
- [6] W. Wu, L. Ge, J. Luo, R. Huan, and Y. Yan, "A spectral-temporal patch-based missing area reconstruction for time-series images," *Remote Sensing*, vol. 10, p. 1560, 2018.
- [7] J. M. Beckers and M. Rixen, "EOF calculations and data filling from incomplete oceanographic datasets," *J. Atmos. Oceanic Technol.*, vol. 20(12), pp. 1836–1856, 2003.
- [8] D. Kondrashov and M. Ghil, "Spatio-temporal filling of missing points in geophysical data sets," *Nonlinear Processes Geophys.*, vol. 13, pp. 151–159, 2006.
- [9] A. Alvera-Azcarate, A. Barth, J.-M. Beckers, and R. H. Weisberg, "Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields," *J. Geophys. Res.*, vol. 112, no. C03008, 2007.
- [10] F. Gerber, R. de Jong, M. E. Schaepman, G. Schaepman-Strub, and R. Furrer, "Predicting missing values in spatio-temporal remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2841–2853, 2018.
- [11] H. Shen, X. Li, Q. Chen, C. Zeng, G. Yang, H. Li, and L. Zhang, "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, pp. 61–85, 09 2015.
- [12] C. Xu, "Reconstruction of gappy GPS coordinate time series using empirical orthogonal functions," *J. Geophys. Res. Solid Earth*, vol. 121, pp. 9020–9033, 2016.
- [13] M. Ghil, M. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. Mann, A. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou, "Advanced spectral methods for climatic time series," *Review of Geophysics*, vol. 40, 1, pp. 1–41, 2002.
- [14] A. Hannachi, I. Jolliffe, and D. Stephenson, "Empirical orthogonal functions and related techniques in atmospheric science: A review," *Int. J. Climatol.*, vol. 27, pp. 1119–1152, 2007.
- [15] R. Prébet, Y. Yan, M. Jauvin, and E. Trouvé, "A data-adaptive EOF based method for displacement signal extraction from interferogram time series," in *IEEE Int. Geosci. and Remote Sens. Symposium*, 2018.
- [16] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, 2003.
- [17] J.-M. Brankart and P. Brasseur, "Optimal analysis of in situ data in the western Mediterranean using statistics and cross-validation," *J. Atmos. Oceanic Technol.*, vol. 13, pp. 477–491, 1995.