Gap filling based on EOF analysis of temporal covariance of offset tracking displacement measurement time series

Alexandre Hippert-Ferrer[1], Yajing Yan[1], Philippe Bolon[1]

[1] Laboratoire d'Informatique, Systèmes, Traitement de l'Information et la Connaissance

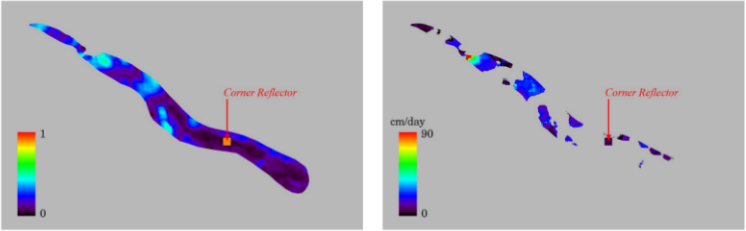
13/09/2018

Contents

Context

- Regular availability of satellite images (e.g. Sentinel data)
- Different displacement extraction techniques : (D-)InSAR, offset tracking...

- Presence of missing data in displacement fields in space and time
- Causes : image quality, missing image, possible limitations of displacement extraction techniques.



Argentiere glacier, offset tracking of TerraSAR-X in Summer 2010 (Fallourd et al. 2011)

Motivation

- Inverse distance weighting (IDW), simple kriging (SK), regression problems
- Spatial interpolation only
- It is necessary to make use of temporal information

How can we manage the presence of **spatio-temporal** missing data in time series ?

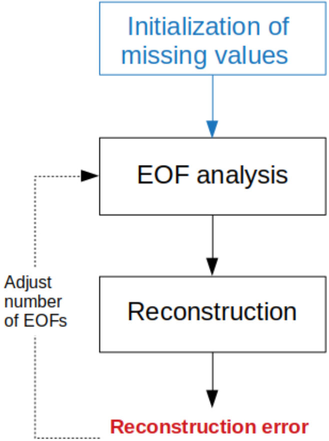
$\rightarrow$ We propose a statistical method based on spatio-temporal characteristics of the displacement time series.

The EM-EOF method

- Based on Empirical Orthogonal Functions (EOFs) which allow a signal representation in terms of orthogonal functions which describe a temporal or spatial mode of variation (comes from PCA), see (Hannachi et al. 2007) ;
- Based on analyses of the temporal or spatio-temporal covariance of a time series of displacement measurement ;
- Iterative processing with a first initialization of missing values by hand before EOF analysis.

→ Allows for a spatio-temporal interpolation

Workflow


Initialization of missing values
EOF analysis
Reconstruction
Reconstruction error
Adjust number of EOFs

Data representation

- Let $X(\mathbf{s}, t)$ be a spatio-temporal field containing the values of X at position $\mathbf{s}$ and time t :

$$X = (\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_n) = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & x_{p3} & \cdots & x_{pn} \end{pmatrix} \quad (1)$$

$(x_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ is the value at position $\mathbf{s}_i$ and time t_j and may be missing.

- We then initialize the missing values by a defined quantity.

Reconstruction using EOFs

- Estimation of the temporal covariance :

$$C = \frac{1}{p-1} X^t X \tag{2}$$

- Then we resolve :

$$CS = SD \tag{3}$$

$\rightarrow$ S contains the self-orthogonal eigenvectors $(\mathbf{s}_i)_{0 \leq i \leq p}$ called EOFs.

- We reconstruct X with M number of EOFs :

$$X = \sum_{i=1}^{r} a_i \mathbf{s}_i^t \rightarrow \hat{X} = \sum_{i=1}^{M \ll r} a_i \mathbf{s}_i^t \tag{4}$$

with $a_i = X\mathbf{s}_i$ and $r \leq min(p, n)$.

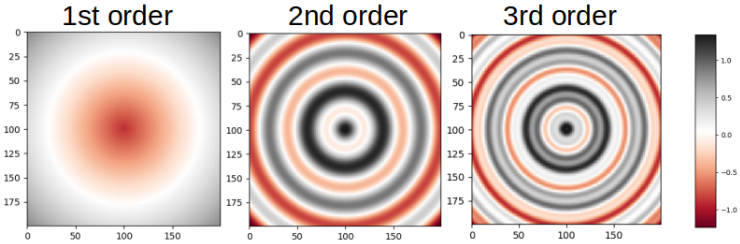
Key parameters

1. Missing data
   - % of missing values
   - Type : random, correlated
2. Noise
   - Signal-to-Noise Ratio (SNR)
   - Type : white Gaussian, correlated noise
3. Missing data initialization
   - Spatial mean
   - Spatial mean + noise

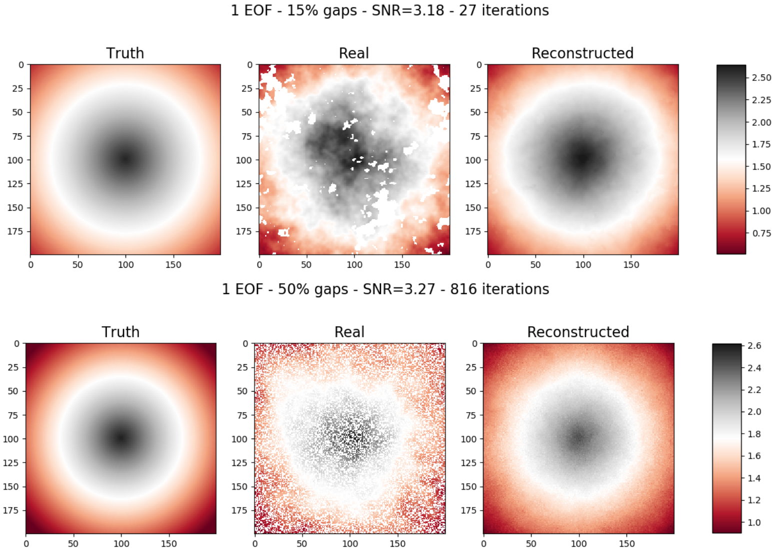
Synthetic data

Different spatio-temporal fields are generated for the experience :

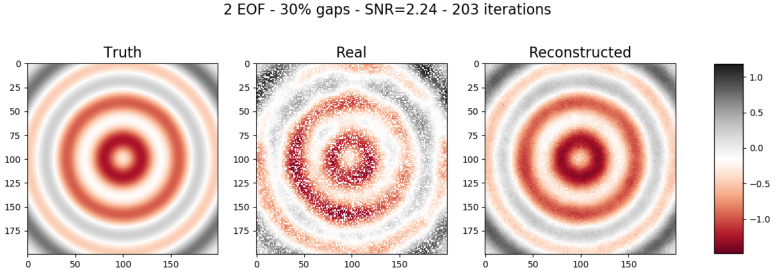
$f(r,t)$	Order
$(1-2r)t$	1
$\sin(\frac{\pi}{2}t)\cos(\frac{\pi}{2}r) + 0.5\cos(\frac{3\pi}{2}t)\cos(5\pi r)$	2
$\sin(\frac{\pi}{2}t)\cos(\frac{\pi}{2}r) + 0.5\cos(\frac{3\pi}{2}t)\cos(5\pi r) + 0.3\sin(\frac{7\pi}{2}t)\sin(15\pi r)$	3

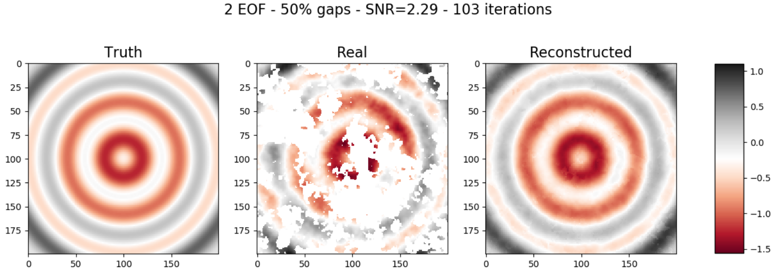


Some results on synthetic data



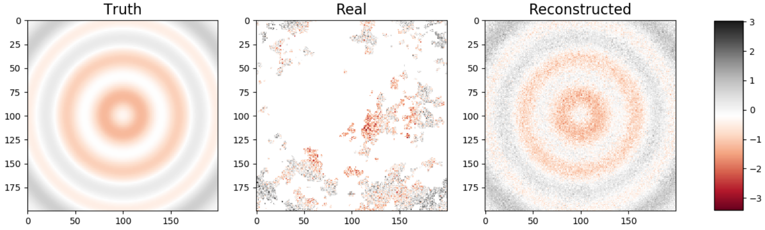
Some results on synthetic data



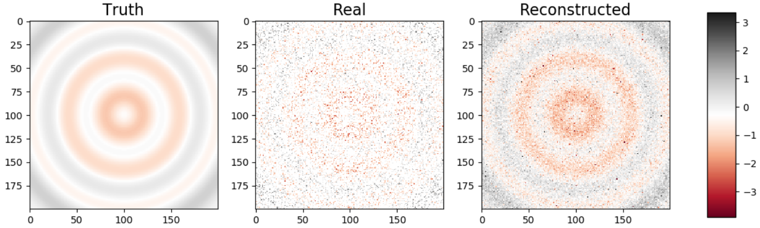


Some results on synthetic data : worst cases

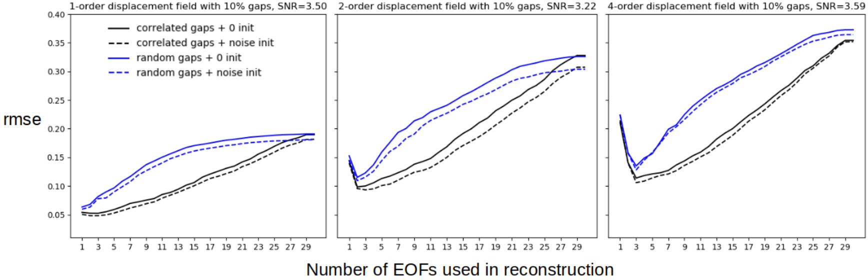
1 EOF - 70% gaps - SNR=0.52 - 72 iterations



1 EOF - 70% gaps - SNR=0.52 - 326 iterations



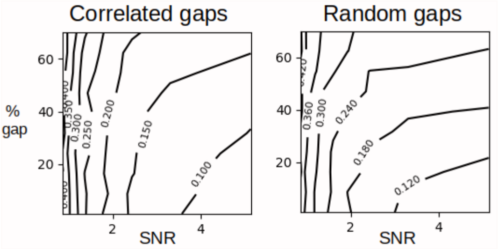
Reconstruction error : RMSE vs. number of EOFs



- *min*(RMSE) is reached at $M \approx$ signal order
- M greatly depends on missing data type and less on missing data initialization.

Reconstruction error

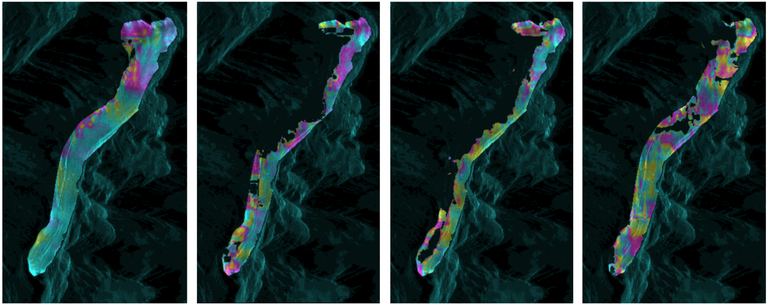
min(RMSE) for % of gaps vs. SNR with noise initialization :



- Best case : high SNR (>3) and few missing values (<10%)
- The impact of SNR is more important than the % of missing values.

Time series of displacement fields

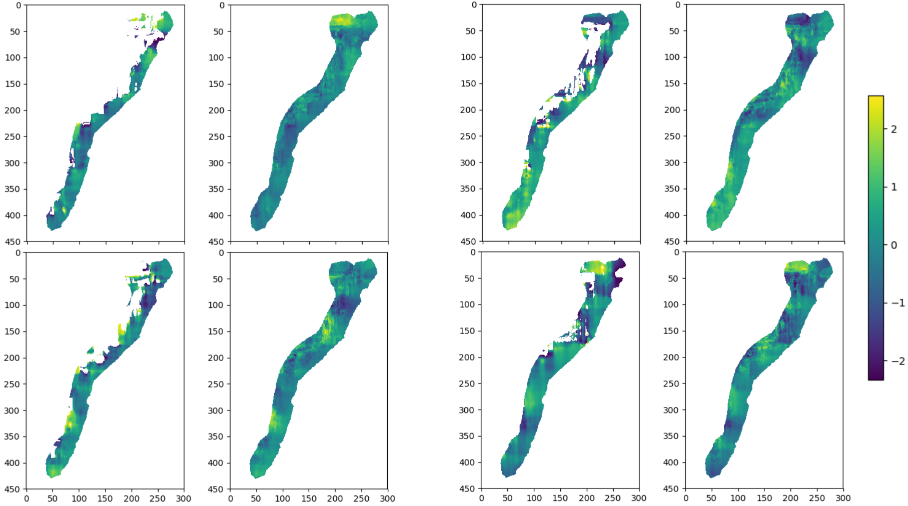
- 65 displacement fields of Mont-Blanc glaciers computed by *offset tracking* of Sentinel 1 SAR images
- Time period : september 2016 to december 2017
- Presence of missing data in space and time



09/2016 04/2017 08/2017 12/2017

Preliminary results

12 days-displacement [m] in radar LOS (5-30% gaps, 8 EOF, RMSE$\leq$10%)



Conclusion

1. EM-EOF method seems promising according to synthetic simulations :
   - The proposed method is robust against % of gaps (reconstruction still working for $\geq$50% gaps) ;
   - The reconstruction performs better with correlated gaps initialized with noise ;
   - The reconstruction is more demanding in case of correlated noise ;
   - The impact of the SNR is more significant than the quantity of gaps.
2. Satisfactory preliminary results over Argentiere glacier obtained ;
3. Further investigation and validation by InSAR data in future work.

Questions ?

References :

[1] R. Fallourd, O. Harant, E. Trouvé and P. Bolon. "Monitoring temperate glacier displacement by multi-temporal TerraSAR-X images and continuous GPS measurements." *J-STARS*, 4 : 372-386, 2011.

[2] J. M. Beckers and M. Rixen. "EOF Calculation and Data Filling from Incomplete Oceanographic Datasets". *J. Atmos. Oceanic Technol*, 20(12) : 1836-1856, 2003.

[3] A. Hannachi, I. Jollife and D. Stephenson. "Empirical orthogonal functions and related techniques in atmospheric science : A review". *Int. J. Climatol.*, 27 : 1119-1152, 2007.

Reconstruction error

Crossed validation reconstruction error :

$$rmse = \frac{1}{N}\sqrt{\sum_{i=1}^{N} |\mathbf{x}_r - \mathbf{x}|^2} \qquad (5)$$

where $\mathbf{x}_r$ and $\mathbf{x}$ are vectors containing N points randomly chosen in each image before reconstruction.

Singular Spectrum Analysis

Data representation :

$$\boldsymbol{x}_t = (\mathbf{x}_t, \mathbf{x}_{t+1}, \dots \mathbf{x}_{t+M-1}) \tag{6}$$

$$\mathcal{X} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_M \\ \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_{M+1} \\ \vdots & \vdots & & \vdots \\ \mathbf{x}_{n-M+1} & \mathbf{x}_{n-M+2} & \cdots & \mathbf{x}_n \end{pmatrix} \tag{7}$$

"Grand" covariance matrix :

$$\mathcal{C} = \frac{1}{n-M+1}\mathcal{X}^T\mathcal{X} = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1p} \\ C_{21} & C_{22} & \cdots & C_{2p} \\ \vdots & \vdots & & \vdots \\ C_{p1} & C_{p2} & \cdots & C_{pp} \end{pmatrix} \tag{8}$$

where $(C_{ij})_{1\leq i,j\leq p}$ is a delayed covariance matrix between points i and j

$$C_{ij} = \frac{1}{n-M+1}\sum_{t=1}^{n-M+1} \boldsymbol{x}_t^{i\,T}\boldsymbol{x}_t^i \tag{9}$$

Singular Spectrum Analysis

Calcul des EOFs étendus $\mathbf{v}_k$:

$$\mathcal{X} = U\Theta V^t \tag{10}$$

$$\boldsymbol{x}_t^T = \sum_{k=1}^{d} \theta_k u_k \mathbf{v}_k \tag{11}$$

Reconstruction du signal :

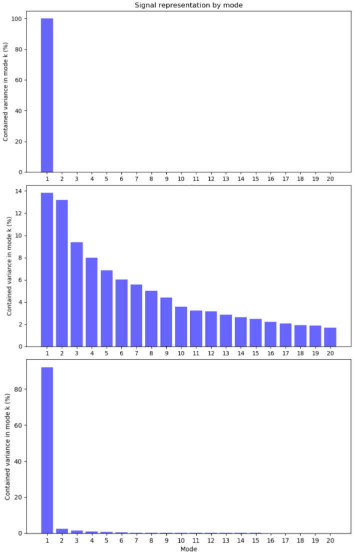
$$RCs(t) = \begin{cases} \frac{1}{t}\sum_{j=1}^{t}\sum_K \theta_k v_k(t-j+1)u_k, & 1 \leq t \leq M-1 \\ \frac{1}{M}\sum_{j=1}^{M}\sum_K \theta_k v_k(t-j+1)u_k, & M \leq t \leq n-M+1 \\ \frac{1}{n-t+1}\sum_{j=t-n+M}^{M}\sum_K \theta_k v_k(t-j+1)u_k, & n-M+2 \leq t \leq n \end{cases}$$

Space correlated noise

A partir d'une fonction d'auto-corrélation de type $c(r) = r^{-\beta}$ et d'une image de bruit blanc b :

1. Calcul de la densité spectrale de puissance de c : $\Gamma(c) = |\mathcal{F}\{c\}|$
2. Calcul de la TF de b : $\mathcal{F}\{b\}$
3. Filtrage fréquentiel : $\mathcal{F}\{b\}\Gamma(c)$
4. Calcul de $\mathcal{F}^{-1}\{\mathcal{F}\{b\}\Gamma(c)\}$

System energy



Variance expliquée par les M premiers modes :

$$\frac{100 \sum_{k=1}^{M} \lambda_k}{\sum_{k=1}^{r} \lambda_k} \% \tag{12}$$

min(rmse) vs. image number