Spatio-temporal missing data reconstruction in satellite displacement measurement time series

Alexandre Hippert-Ferrer\textsuperscript{1}†, Yajing Yan\textsuperscript{1}, Philippe Bolon\textsuperscript{1}, Romain Millan\textsuperscript{2}

\textsuperscript{1}Laboratoire d'Informatique, Systèmes, Traitement de l'Information et la Connaissance (LISTIC), Annecy, France
\textsuperscript{2}Institut des Géosciences de l'Environnement (IGE), Université Grenoble Alpes, CNRS, Grenoble, France

† Correspondence to: alexandre.hippert-ferrer@univ-smb.fr

Thursday, May 7
Introduction

- **Missing data** is a frequent issue in displacement time series in both space and time.
- It can prevent the full understanding of the **physical phenomena** under observation.
- **Causes**: rapid surface changes, missing image, technical limitations.

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

Surface velocity over Fox Glacier, offset tracking of Sentinel-2 images in February 2018 (Millan et al., 2019)

Slow slip event (interferometry), Mexico (Maubant et al., 2020)
Motivation of the study

Handling missing data in displacement time series

- Classical approach: spatial or temporal interpolation
- Not exploited (yet): spatio-temporal information

→ Manage spatio-temporal missing data in time series ←

Objective: propose a statistical gap-filling method addressing

1. Randomness and possible spatial, temporal and spatio-temporal correlation of
   - Noise
   - Missing data

2. Complex displacement behaviors (mixed frequencies)

3. Small time series
→ Extension of the EM-EOF method (Hippert et al., 2019, 2020) [3, 4]

**Key features of the extended EM-EOF method**:

- Low rank structure of the sample spatio-temporal covariance matrix.
- Displacement signal and noise decomposed in empirical orthogonal functions (EOFs).
- Reconstruction with an appropriate initialization of missing values.
- Expectation-Maximization (EM)-type algorithm for resolution.
Data representation

- Let $X_t = \{x_{ij}(t)\}_{1 \leq i \leq P_x, 1 \leq j \leq P_y}$ be a spatial grid observed at time $t = 1, \ldots, N$.
- Some elements of $X_t$ are missing.
- All $X_t$ are stacked into a spatio-temporal data matrix $Y = (X_1, X_2, \ldots, X_N)$.

Square window of size $M_x \times M_y = M$

- Each $X_t$ is augmented into a Hankel-block Hankel (HbH) matrix $D_t$ of size $K \times M = K_x K_y \times M_x M_y$, with $K_x = (P_x - M_x + 1)$, $K_y = (P_y - M_y + 1)$.
- All $D_t$ is stacked into a spatio-temporal matrix $D$ of size $(K \times NM)$, that is $D = (D_1, D_2, \ldots, D_N)$. 
Covariance estimation and decomposition

- Sample spatio-temporal covariance is estimated:
  \[ \hat{C} = \frac{1}{K} \mathcal{D}^T \mathcal{D} \]  
  (1)

- The eigenvalue decomposition (EVD) of matrix \( \hat{C} \) yields to:
  \[ \hat{C} \text{ EVD} = \sum_{i=1}^{NM} \lambda_i u_i u_i^T \]  
  (2)

Vectors \( u_i \) are the \( NM \) EOFs modes of matrix \( \mathcal{D} \). First modes capture the main spatio-temporal dynamical behavior of the signal, others represent perturbations.

- Reconstruction with an optimal number of EOF modes \( R \ll NM \) is obtained as
  \[ \hat{\mathcal{D}} = A_R U_R \]  
  (3)

\( A \) is the matrix of principal components, which are the projection of \( \mathcal{D} \) on each EEOF \( u \).

How do we find \( R \)?
Selection of the optimal number of EOF modes

1. Root-mean-square error (cross-RMSE) on cross-validated data \( \mathcal{Y} \in \mathcal{Y} \):

\[
\frac{1}{MN} \| \hat{\mathcal{Y}}_k - \mathcal{Y} \|_2
\]

- Requires no a priori information

2. Confidence index associated with each eigenvalue of \( \mathcal{D} \):

\[
C_k = \frac{\max(\Gamma_k) - \Gamma_k}{\max(\Gamma_k) - \min(\Gamma_k)} \quad k = 1, \ldots, NM
\]

with \( \Gamma_k = \log \left( \frac{\Delta \lambda_k}{\lambda_j - \lambda_k} \right) \).

- Investigation of eigenvalue degeneracy, which is linked to their uncertainties \( \frac{\Delta \lambda_k}{\lambda_j - \lambda_k} \).
- Over-estimation of EOF modes is addressed by building metric \( C_k \).
General principle of the method

Initialize missing values → First estimation of the optimal number of EOF $R$

Update missing values → Reconstruction with $r \leq R$ EOFs

Update $r$ → Compute $C_k$

EM iteration*

* For a fixed number of EOF modes, cross-RMSE is computed until it converges.
Trade-off between the amount of information extracted in the window (large $M$) and the number of repetitions of the window within each image (small $M$).

Upper limit based on covariance estimation theory: $M < P/6$

Lower limit:

We use the spatial decorrelation decay $\tau$ defined as:

$$\tau = -\frac{\Delta P}{\log r}$$

(6)

$r$ : lag-one auto-correlation

$\Delta P$ : spatial sampling rate, here 1 pixel.

$M$ can be approximated by $M \approx P/\tau$ (Ghil et al., 2002) [1] which gives $M > P/20$ with $r < 0.95$. 
Surface velocities on Fox Glacier, New Zealand

<table>
<thead>
<tr>
<th>Period</th>
<th>Platform</th>
<th>Data type</th>
<th>Time series size</th>
<th>[min, max]% missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>02/2018-09/2018</td>
<td>Sentinel-2</td>
<td>Offset tracking</td>
<td>12</td>
<td>[10, 60]%</td>
</tr>
</tbody>
</table>

Time series description.

- Surface velocities computed from the study of (Millan et al., 2019) [5].

Surface velocities (m/year) on Fox Glacier. P1 and P2 locations are selected for temporal evolution analysis.
13 EOFs modes; $M=225$; cross-validation data: 1% of observed values.

- Seasonal variation is retrieved, consistent values with the literature (4.5 m/day below the main ice fall in winter).

- Improved accuracy of $\approx 15$ m/year compared to the EM-EOF method.
Optimal number of EOF modes (13) corresponds to a peak in $C_k$ which coincides with a break in the eigenvalue spectrum.

Eigenvalues multiplets are kept in the reconstructed data.

$$C_k = \frac{\max(\Gamma_k) - \Gamma_k}{\max(\Gamma_k) - \min(\Gamma_k)}$$

$$\Gamma_k = \log\left(\frac{\Delta \lambda_k}{\lambda_j - \lambda_k}\right)$$
Conclusion

- **Extension** of the EM-EOF method to impute spatio-temporal missing values.
  - Can handle small time series with high incompleteness
  - Extraction of the displacement signal from heterogeneous perturbations (noise)

- **Robust selection of the optimal number of EOF modes** based on:
  - Iterative computation of the cross-validation error
  - Confidence metric based on eigenvalue uncertainties to address potential over-estimation due to eigenvalue degeneracy

- A range of spatial lag $M$ is provided

- Limitations: potential edge effect due to spatial square window.

**Perspective**: Use a shaped window (adaptive spatial lag) instead of a square window.
Thank you for your attention.


This work has been supported by the Programme National de Télédétection Spatiale (PNTS, http://www.insu.cnrs.fr/pnts), grant PNTS-2019-11, and by the SIRGA project.
Diagonal averaging, called *hankelization*, [2] is applied successively to each matrix $H_{i,t}$ and to each matrix $D_t$, so that we have the following averaging:

$$x_{ik}(t) = \frac{1}{\#A_k} \sum_{(l,l') \in A_k} x_{ll'}(t)$$  \hspace{1cm} (7)$$

$$H_{k,t} = \frac{1}{\#B_k} \sum_{(l,l') \in B_k} H_{ll',t}$$  \hspace{1cm} (8)$$

with $A_k = \{(l,l') : 1 \leq l \leq K_y, 1 \leq l' \leq M_y, l + l' = k + 1\}$ and $B_k = \{(l,l') : 1 \leq l \leq K_x, 1 \leq l' \leq M_x, l + l' = k + 1\}$. 

Reconstruction averaging
Confidence index and effective sample size

North's et al. "rule of thumb" (North, 1982) to approximate the eigenvalue uncertainty:

\[ \Delta \lambda_k \approx \sqrt{\frac{2}{L^* \lambda_k}} \quad \Delta u_k \approx \frac{\Delta \lambda_k}{\lambda_j - \lambda_k} u_j \]  

(9)

with \( L^* = N^* M^* \).

- \( N^* = N \left[ 1 + 2 \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \rho(k) \right]^{-1} \) is the temporal ESS (Thiébaux, 1984)
- \( M^* \) is the spatial ESS within each spatial window of size \( M \). We estimate it by:

\[ M^* = M \left( 1 + 2 \sum_{k=1}^{M} \left(1 - \frac{k}{M}\right) \nu(k) \right)^{-1} \]  

(10)

Then \( \Gamma_k = \log \left( \frac{\Delta \lambda_k}{\lambda_j - \lambda_k} \right) \) and \( C_k \) is computed as:

\[ C_k = \frac{\max(\Gamma_k) - \Gamma_k}{\max(\Gamma_k) - \min(\Gamma_k)} \quad k = 1, \ldots, NM \]  

(11)