Metamodelling for spatial outputs with functional PCA. Application to coastal flooding.

T.V.E. Perrin (elodie.perrin@emse.fr)

PhD supervision team: O. Roustant, J. Rohmer, D. Moncoulon, JP. Naulin, P. Tinard

MASCOT-NUM 2019, IFPEN, Rueil-Malmaison - 18-20 March 2019

Introduction

- **Aim :** To build a metamodel (Gaussian Process Regression (GPR)) for coastal flooding model, which is time consuming.
- In the literature, metamodel construction is based on principal component analysis (PCA) or wavelet basis [1] : but the spatial output of coastal flooding model is **high dimensional** (thousands of pixels) and has **strong irregularities.**
- We suggest a method using **functional PCA** [2]: here, PCA on wavelet basis.
- We propose a **criterion based on the energy** of an image to select the most important wavelet coefficients.

Mathematical model

$$\begin{array}{cccc} f & : & \mathcal{X} \subseteq \mathbb{R}^d & \to & \mathcal{L}_2([0,1]^2) \\ & & x & \mapsto & y_x(z) \end{array}$$

- We know n outputs of f for the data set : $x_1, \ldots, x_n$.
- We want to estimate $f(x^*)$.

Functional PCA (FPCA)

- **FPCA :** diagonalization of a covariance operator on $\mathcal{L}_2([0,1]^2)$.
- **Practical FPCA :** The functional space $\mathcal{L}_2([0,1]^2)$ is approximated by a finite dimensional one (splines, wavelets ...).
- Orthonormal basis: FPCA $\Leftrightarrow$ PCA on the coefficients of the orthonormal basis.
- Wavelet basis is an orthonormal basis.

Algorithm

1. **Wavelet decomposition :**

$$f(x_i)(z) = y_{x_i}(z) = \sum_{j=1}^{K} \beta_j(x_i)\phi_j(z)$$

with $\phi_j(z)$, the wavelet basis, and $\beta_j(x)$, the wavelet coefficients.

2. **To estimate** $(\beta_j(x^*))_{j=1,\ldots,K}$:
   (a) The most informative wavelet coefficients : $k << K$.
   - FPCA: to apply PCA on the coefficients.
   - To estimate the PCA coordinates of $\hat{\beta}_j(x^*))_{j=1,\ldots,K}$ by GPR.
   - To embed PCA coordinates on wavelet space.

   (b) The less informative wavelet coefficients : empirical mean.
3. **To predict the model output:** $\mathbf{\hat{f}(x^*)(z)} = \sum_{j=1}^{K} \hat{\beta}_j(x^*)\phi_j(z)$

Selection of the wavelet coefficients

- To select the wavelet coefficients which explain at most a proportion $p \in [0,1]$ of the image energy.
- Energy decomposition:

$$||y_x||_2^2 = \int (\sum_{j=1}^{K} \beta_j(x)\phi_j(z))^2 dz = \sum_{j=1}^{K} \beta_j(x)^2$$

- Proportion of energy of $\beta_j(.)$. Two possibilities :
  1. $\lambda_j = \mathbb{E}_X\left[\frac{\beta_j(X)^2}{||y_X||_2^2}\right] = \mathbb{E}_X\left[\frac{\beta_j(X)^2}{\sum_{j=1}^{K}\beta_j(X)^2}\right]$
  2. $\lambda_j = \frac{\mathbb{E}_X[\beta_j(X)^2]}{\mathbb{E}_X[||y_X||_2^2]} = \frac{\mathbb{E}_X[\beta_j(X)^2]}{\sum_{j=1}^{K}\mathbb{E}_X[\beta_j(X)^2]}$

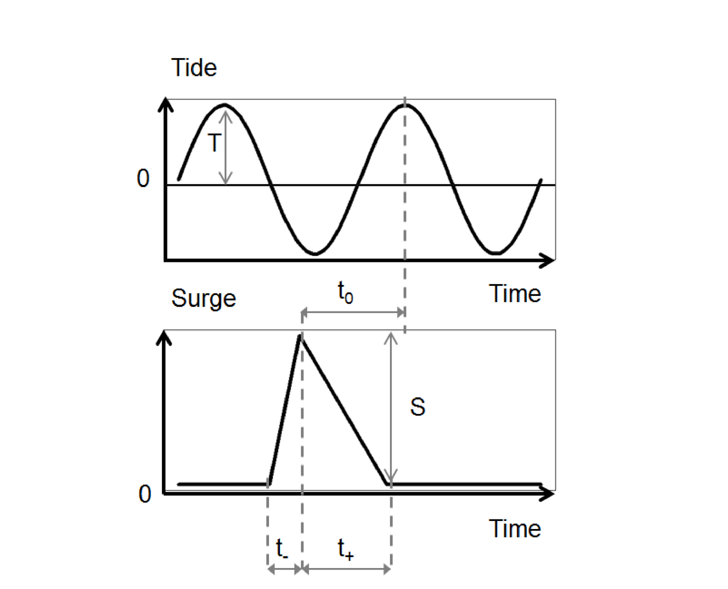
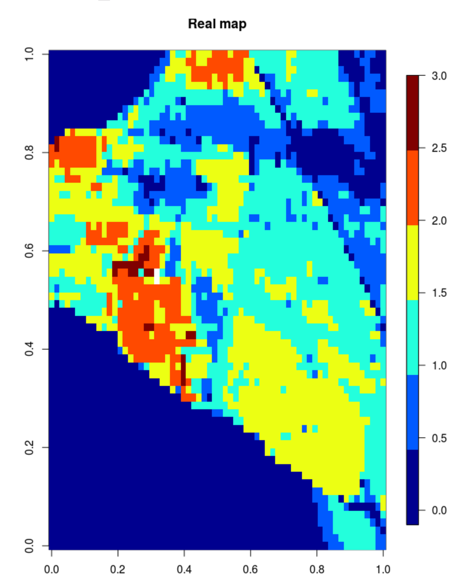
References

[1] A. Marrel, B. Iooss, M. Jullien, B. Laurent, and E. Volkova. Global sensitivity analysis for models with spatially dependent outputs. *Environmetrics*, 22(3):383–397, 2010.

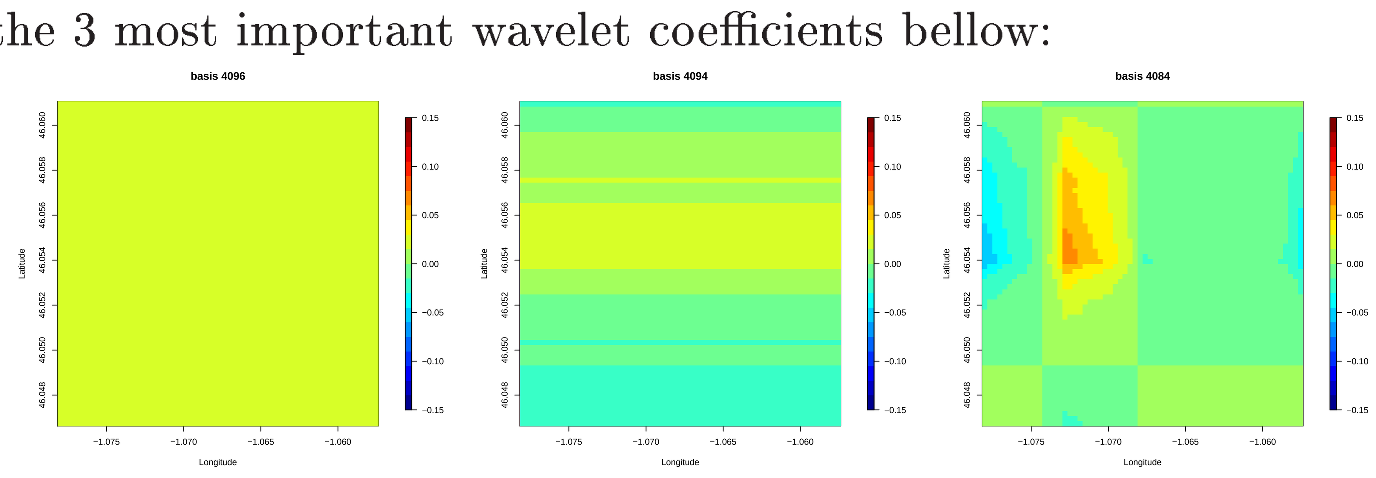
[2] J.O. Ramsay. *Functional data analysis.* Wiley Online Library, 2006.

Application

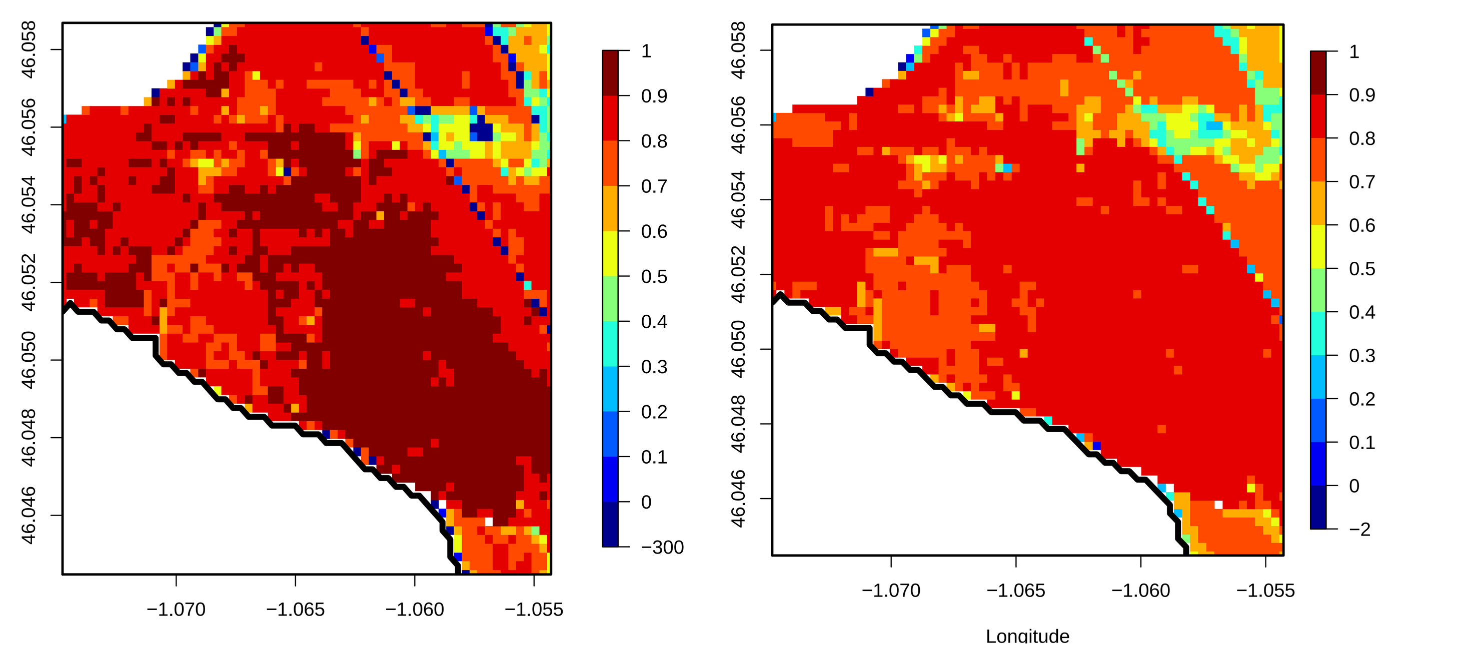
- **Case study:** Coastal flooding by overflow process
  - Location: Bouchôleurs (French Atlantic coast, near La Rochelle), which was touched by the Xynthia storm in 2010.
  - Observations of $n = 500$ runs

Inputs (Waves settings)	Outputs (Maximum water level)
$x = (T, S, t_0, t_+, t_-) \in I \subseteq \mathbb{R}^5$	a map of size 64×64
	

- **Comparison PCA and PCA based on wavelets:**
  1. Cross-validation 10-fold (CV-10-fold).
  2. Daubechies wavelets (6 levels of decomposition), see the basis of the 3 most important wavelet coefficients bellow:



  3. The number of principal components $\mathbf{n_{PC}}$ (for both methods) and the proportion of energy $\mathbf{p}$ are estimated by CV-10-fold.
  4. In the figure below, we have chosen the **first criterion** which gives the best result.



Map of mean Q^2 (dark line: coastline): on the left, for FPCA (first criterion), on the right, for PCA. Here: $\mathbf{n_{PC} = 1}$ (for both methods 94% of the explained variance) and $\mathbf{p = 0.99}$ (≈ 790 wavelet coefficients).

Conclusion and future works

- Competitive Q^2 value for wavelet/FPCA approach.
- Improved computational time: FPCA is much faster than PCA.
- Future works: sensitivity analysis with PCA on the wavelet basis.