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

T.V.E. Perrin
Mines Saint-tienne (EMSE), France

Supervisor(s): O.Roustant (EMSE), J.Rohmer (BRGM), D.Moncoulon (CCR), J.Naulin (CCR), P.Tinard (CCR)

Ph.D. expected duration: Oct. 2017 - Sep. 2020

Adress: 158 Cours Fauriel, 42023 Saint-tienne

Email: elodie.perrin@emse.fr

Abstract:

This abstract is for a poster submission. Gaussian process (GP) is one of the most attractive metamodels for emulating time-consuming computer codes. Here, we focus on problems when outputs are spatial maps. Without loss of generality, we consider a spatial domain $D_z = [0,1]^2$. The computer code is viewed as a function:

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

where x is a vector of scalar inputs and $y_x(z)$ is the ouput at the location $z \in D_z$.

A common technique [1] is to vectorize and to reduce dimensionality of the output map by principal component analysis (PCA). Then, PCA coordinates are emulated independantly by different GP models, which seems reasonable given the orhonormality of the axis. However, output dimensionality can be too high and makes intractable the covariance matrix diagonalization. For instance, marine flooding maps may have more than ten thousand pixels. Furthermore, PCA does not take into account the spatial nature of the data and their properties, such as smoothness.

In our work, $\forall x \in \mathcal{X}$, $y_x(z)$ is decomposed onto a finite basis of functions, using functional principal components analysis (FPCA) [3]. This dimension reduction method allows to keep at most the functional/spatial nature of the outputs. In the litterature, wavelets are often chosen as the basis of functions for spatial maps [2], due to their ability in revealing information at different levels and areas in the maps. Thus we have:

$$y_x(z) = \sum_{j=1}^{K} \beta_j(x)\phi_j(z), \ \forall x \in \mathcal{X}, \ \forall z \in D_z \tag{2}$$

where $j \in \{1, \dots, K\}$, $\beta_j(x)$ are the wavelet coefficients, and $(\phi_j(z))$ the wavelet basis. This basis is orthonormal, so FPCA is equivalent to PCA on the wavelet coefficients. Obviously, in order to reduce the dimension, it is necessary to limit PCA on the most informative wavelet coefficients. We propose to order them in two different ways according to the decomposition of the energy: $||y_x||_2^2 = \int_0^1 y_x(z)^2 dz = \sum_{j=1}^{K} \beta_j(x)^2$ (equation (3)). The unselected coefficients are estimated by the mean.

$$\lambda_j = \mathbb{E}_X \left[\frac{\beta_j(X)^2}{\sum_{j=1}^{K} \beta_j(X)^2} \right] \quad \textbf{or} \quad \lambda_j = \frac{\mathbb{E}_X[\beta_j(X)^2]}{\sum_{j=1}^{K} \mathbb{E}_X[\beta_j(X)^2]} \tag{3}$$

The efficiencies of metamodels obtained with FPCA and PCA are compared on a dataset of 500 computer code simulations of marine flooding at Bouchôleur site, in France. Figure 1 shows their performances using 50 different learning set (chosen at random uniformly from the whole dataset) of size 50, and emulating two principal components for FPCA and PCA. The methods are assessed using the Q^2 criterion, computed at each pixel of the maps.

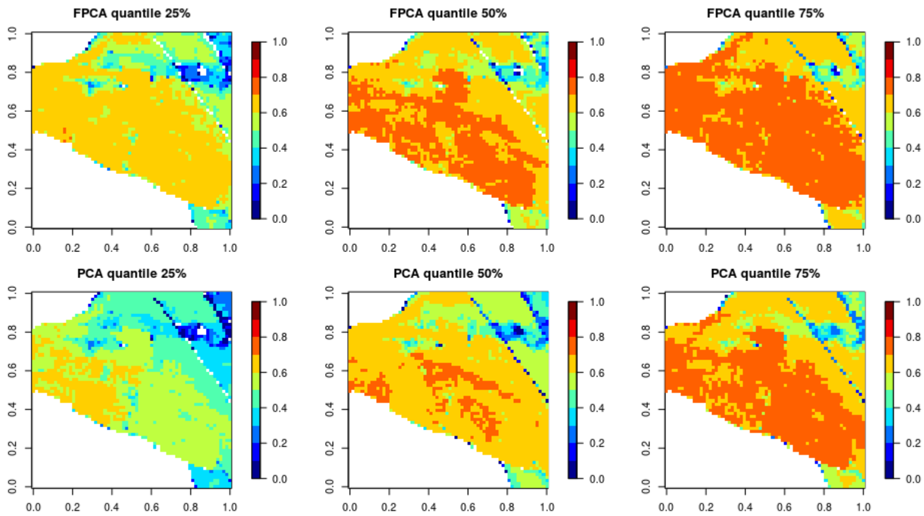

Figure 1: Q^2 maps of 25%, 50% and 75% quantiles (from left to right): metamodelling on FPCA coordinates (first line), metamodelling on PCA coordinates (second line).

It is observed that compared to the standard PCA technique, the proposed methodology leads to a better prediction accuracy (cf. Figure 1) and time computation efficiency. In our case study, applying FPCA is five times faster than PCA. On the other hand, FPCA is sensitive to the number of wavelet coefficients preselectioned, which justifies improving this aspect potentially by relying on cross-validation techniques.

References

[1] T. Chen, K. Hadinoto, W. Yan, and Y. Ma. Efficient meta-modelling of complex process simulations with time–space-dependent outputs. *Computers & chemical engineering*, 35(3):502–509, 2011.

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

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

Short biography – T.V.E. Perrin graduated the MapI^3 master in applied mathematics from Paul Sabatier University, at Toulouse, France. Now, she is doing a PhD in applied mathematics at EMSE, France, on *Propagation of uncertainties and calibration of numerical simulations to estimate the cost linked to marine flooding*. The thesis is supervised by EMSE, BRGM, and CCR.