Fast Prediction of Coastal Flood Hazard Assessment: A Multi-Output Gaussian Process Metamodelling Approach

RISCOPE - Final Workshop

Andrés F. López-Lopera
Associate Professor at Polytechnic University of Hauts-de-France, France.

February 2, 2022

Project: Risk-based System for Coastal Flooding Early Warning (RISCOPE)
Institut de Mathématiques de Toulouse (IMT), France
Bureau de Recherches Géologiques et Minières (BRGM), France

Joint work...

D. Idier
BRGM, France

J. Rohmer
BRGM, France

F. Bachoc
IMT, France

Reliability Engineering and System Safety 218 (2022) 108139

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Multioutput Gaussian processes with functional data: A study on coastal flood hazard assessment

Andrés F. López-Lopera [a,*], Déborah Idier [b], Jérémy Rohmer [b], François Bachoc [c]

[a] *Univ. Polytechnique Hauts-de-France, Céramaths, Valenciennes, F-59313, France*
[b] *BRGM, DRP/R3C, 3 avenue Claude Guillemin, Orléans, F-45060, France*
[c] *Institut de Mathématiques de Toulouse, Univ. Paul Sabatier, Toulouse, F-31062, France*

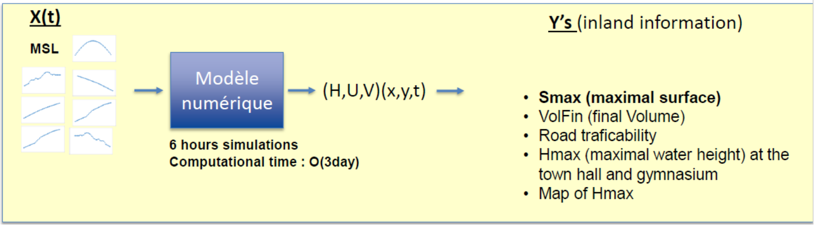
ARTICLE INFO

Keywords:
Surrogate model
Spatial flood data
Functional analysis
Dimensionality reduction
Separable kernel
Sparse-variational approximation

ABSTRACT

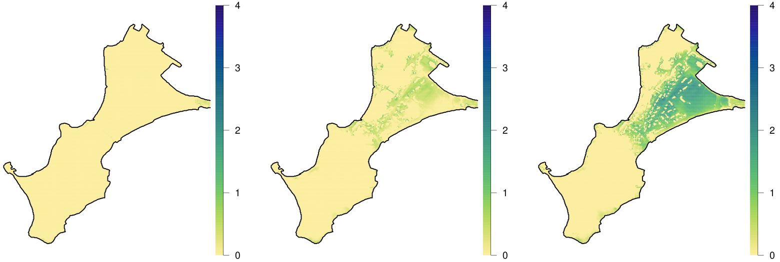
Surrogate models are often used to replace costly-to-evaluate complex coastal codes to achieve substantial computational savings. In many of those models, the hydrometeorological forcing conditions (inputs) or flood events (outputs) are conveniently parameterized by scalar representations, neglecting that the inputs are actually time series and that floods propagate spatially inland. Both facts are crucial in flood prediction for complex coastal systems. Our aim is to establish a surrogate model that accounts for time-varying inputs and provides information on spatially varying inland flooding. We introduce a multioutput Gaussian process model based on a separable kernel that correlates both functional inputs and spatial locations. Efficient implementations consider tensor-structured computations or sparse-variational approximations. In several experiments, we demonstrate the versatility of the model for both learning maps and inferring unobserved maps, numerically showing the convergence of predictions as the number of learning maps increases. We assess our framework in a coastal flood prediction application. Predictions are obtained with small error values within computation time highly compatible with short-term forecast requirements (on the order of minutes compared to the days required by hydrodynamic simulators). We conclude that our framework is a promising approach for forecast and early-warning systems.

Motivation


X(t)
MSL
Modèle numérique
(H,U,V)(x,y,t)
6 hours simulations
Computational time : O(3day)
Y's (inland information)
• Smax (maximal surface)
• VolFin (final Volume)
• Road traficability
• Hmax (maximal water height) at the town hall and gymnasium
• Map of Hmax

Challenge

Spatial flood events: maximal inland water level (H_{max} [m])

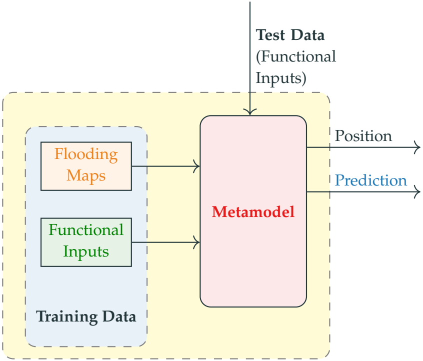


Challenge:

- Each flood event takes $\sim$ **3 days of simulation**.

Research subject

Goal: to build a **metamodel** accounting for both spatial and functional data



· This will lead to faster (approximate) predictions.

Application: forecasting and early warning systems.

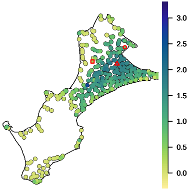
Research subject

- To do so...
  - We further investigate a metamodel based on **Gaussian process**.
  - This results in a **multi-output process**:
    - correlated spatial flood events are driven by (functional) hydro-meteorological inputs.
  - Efficient implementations are based on **Python and R codes**.
    - Codes used later in a early warning-oriented demonstrator (afternoon session!!).

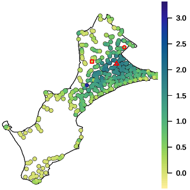
Géosciences pour une Terre durable

Research subject

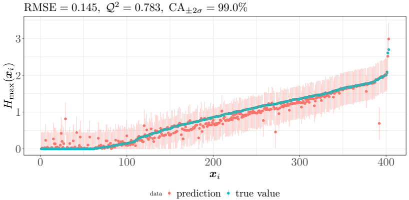
- To do so...
  - We further investigate a metamodel based on **Gaussian process**.
  - This results in a **multi-output process**:
    - correlated spatial flood events are driven by (functional) hydro-meteorological inputs.
  - Efficient implementations are based on **Python and R codes**.
    - Codes used later in a early warning-oriented demonstrator (afternoon session!!).
- Our developments provide fast (tens of seconds) and accurate predictions:



ground truth



prediction



ground truth vs prediction

Spatial Gaussian processes with functional inputs

· To deal with the functional inputs, Betancourt et al. [2020] explored a Gaussian process (GP)-based approach with a dedicated kernel function:

$$k_f(\mathcal{F}, \mathcal{F}') = \mathrm{cov}\left\{Y(\mathcal{F}), Y(\mathcal{F}')\right\}. \tag{1}$$

· Here, we have considered a new kernel function that accounts for both functional inputs and spatial information:

$$k((x, \mathcal{F}), (x', \mathcal{F}')) = \mathrm{cov}\left\{Y(x, \mathcal{F}), Y(x', \mathcal{F}')\right\} = k_s(x, x')k_f(\mathcal{F}, \mathcal{F}'), \tag{2}$$

with sub-kernels $k_s : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ and $k_f : \mathcal{F}(\mathcal{T}, \mathbb{R})^Q \times \mathcal{F}(\mathcal{T}, \mathbb{R})^Q \to \mathbb{R}$.

· This formulation allows us to enjoy efficient implementations based on well-founded mathematical operations, e.g., using Kronecker products.

Link with multi-output Gaussian processes

- The process Y can be written as a multi-output process Z:

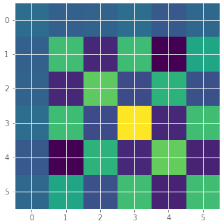
$$Z_i(x) := Y(\mathcal{F}_i, x), \quad \text{for } i = 1, \cdots, R.$$

- In that case, k can be rewritten as:

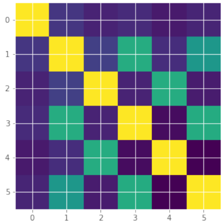
$$k_{i,j}(x, x') = b_{i,j}\, k_s(x, x'), \tag{3}$$

with $b_{i,j} := k_f(\mathcal{F}_i, \mathcal{F}_j)$, for $i, j = 1, \ldots, R$.

- (3) follows the structure of the *linear models of coregionalisation* (LMC):



B

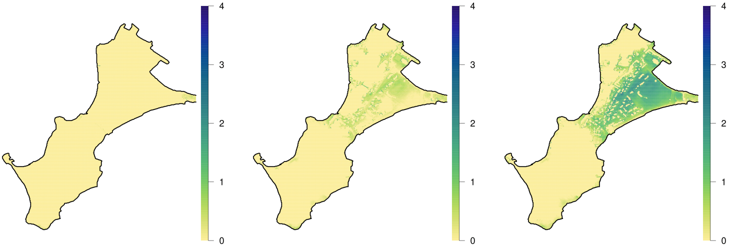


$k_f(\mathcal{F}, \mathcal{F}')$

Coregionalisation matrix

Coastal flooding application

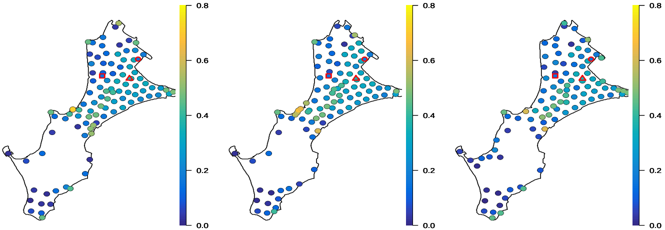
Spatial flood events: maximal inland water level (H_{max} [m])



- **174 flood/no flood events** containing 65k inland observations.
  - *21 historical* meteo-oceanic conditions (9 flood) [Idier et al., 2020]
  - *16 reinforced* historical conditions of the 9 historical flood events
  - *94 extreme* meteo-oceanic conditions (flood, non flood)
  - *43 modifications* of some of the 94 extreme conditions (flood/no flood)

Design of Experiments (DoE)

· Due to computational limitations, we focus on a tractable amount of points.



Replicate 1 Replicate 2 Replicate 3

Examples of DoE considering 103 spatial locations

Leave-one-out (LOO) test

LOO test: each scenario is predicted using data of the other ones.

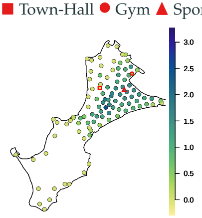
· Therefore, **174 GP metamodels** are trained and tested.

· We fix **103 spatial design points per map**.

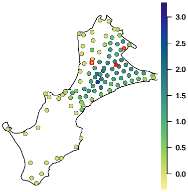
  – including the town-hall (■), gym (●) and sports field (▲) at Gâvres.

· We use **Matérn 5/2 kernels** as covariance matrices.

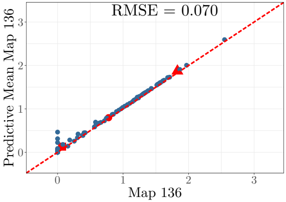
Flooding scenario 136: significant flood event



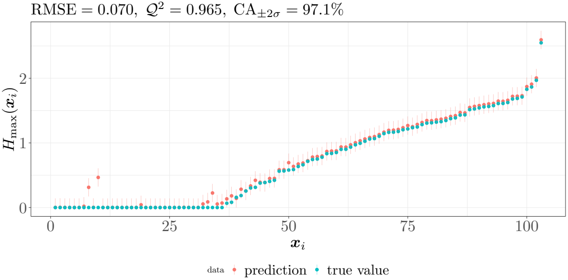
ground truth



prediction

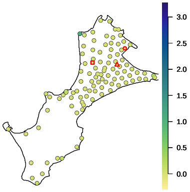


ground truth vs prediction

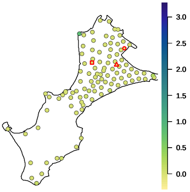


Flooding scenario 1: minor flood event

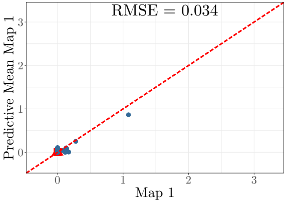
■ Town-Hall ● Gym ▲ Sports Field



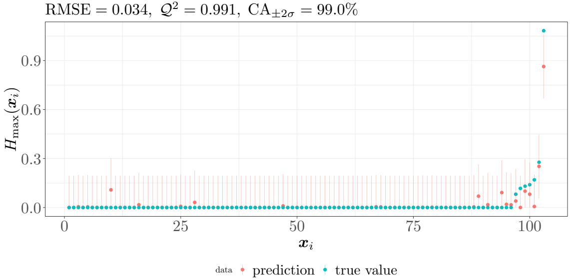
ground truth



prediction

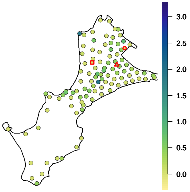


ground truth vs prediction

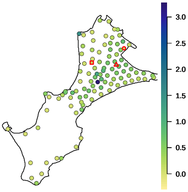


Flooding scenario 43: moderate flood event

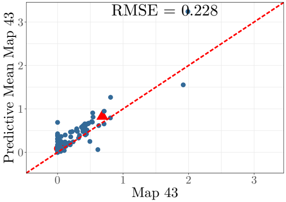
■ Town-Hall ● Gym ▲ Sports Field



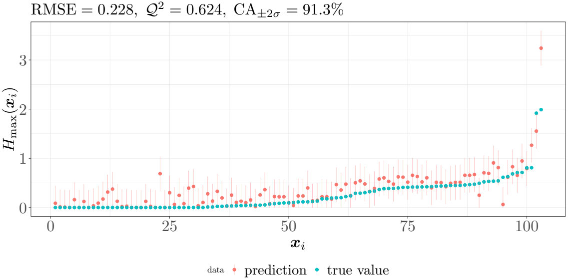
ground truth



prediction

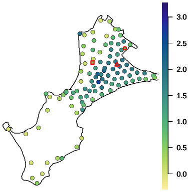


ground truth vs prediction

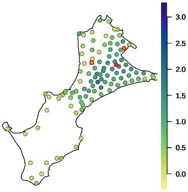


Flooding scenario 80: misprediction

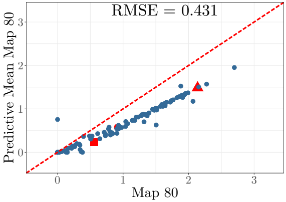
■ Town-Hall ● Gym ▲ Sports Field



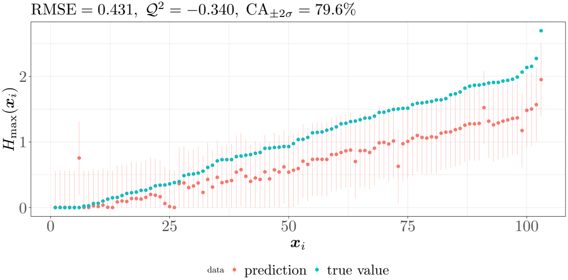
ground truth



prediction

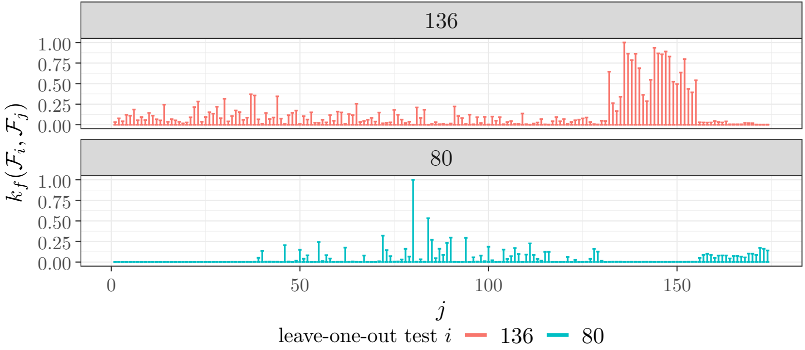


ground truth vs prediction



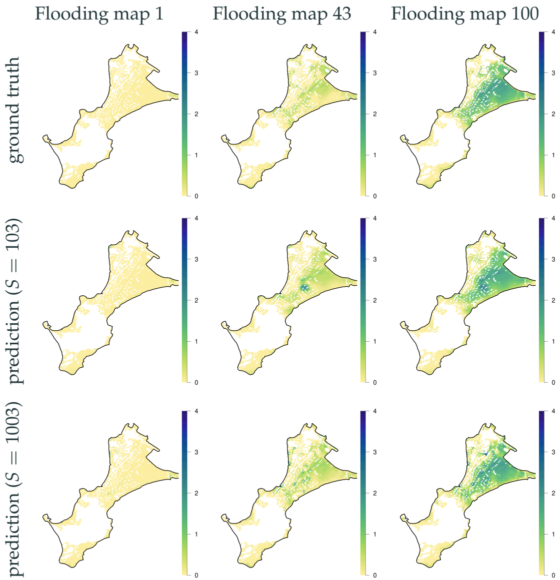
Leave-one-out (LOO) test

Flooding scenario 80: misprediction (discussion)

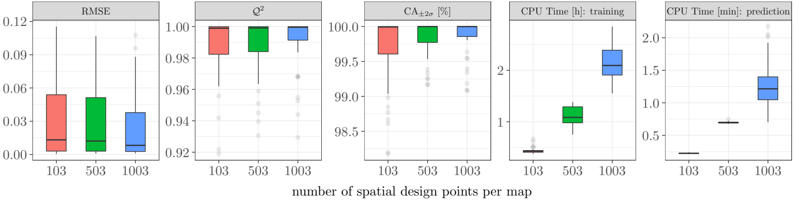


Correlations between functional inputs

Influence of the number of design points



Influence of the number of design points



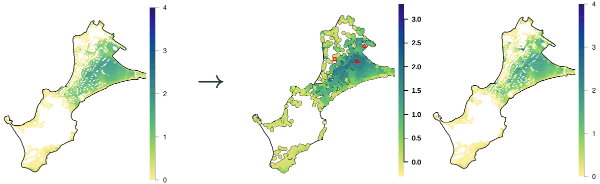
Performance indicators computed on a dataset based on **131 flood events**

Advantages and remaining challenges

Advantages

- Approximate predictions under **smoothness assumptions**.

Numerical simulator GP metamodel



$\sim$ 3 days ($\sim$ 65k points) | $\sim$ 2 seconds ($\sim$ 1k points) | $\sim$ 1 min ($\sim$ 34k points)

Challenges:

- **Modelling non-stationary spatial data:**
  - e.g. considering non-stationary kernels [e.g., Remes et al., 2017].
- **Dimension reduction of the output space:**
  - e.g. via Wavelets decomposition [e.g., Perrin et al., 2020].

Conclusions

In summary...

- We further investigated a **Multi-output GP-based framework**:
  - It accounts for both **spatial outputs** and **functional inputs**.
  - Regularity assumptions are encoded into **kernel functions**.
- We applied our approach as a **metamodel of coastal computer codes**:
  - **Fast and (locally) accurate predictions** are obtained.
  - The **availability and diversity of learning events** are crucial.
- We provided **Python** and **R codes** (github.com/anfelopera/spatfGPs):
  - **Efficient implementations** are developed based on *sparse-variational inference* or *Kronecker-based operations*.

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

brgm Géosciences pour une Terre durable

References

M A Alvarez, L Rosasco, and N Lawrence. Kernels for vector-valued functions: A review. *Foundations and Trends in Machine Learning*, 4(3):195, 2012.

J Betancourt, F Bachoc, T Klein, D Idier, R Pedreros, and J Rohmer. GP metamodeling of functional-input code for coastal flood hazard assessment. *Reliability Engineering & System Safety*, 198:106870, 2020.

D Idier, J Rohmer, R Pedreros, S Le Roy, J Lambert, J Louisor, G Le Cozannet, and E Le Cornec. Coastal flood: A composite method for past events characterisation providing insights in past, present and future hazard-joining historical, statistical and modelling approaches. *Natural Hazards*, 101:465–501, 2020.

T V E Perrin, O Roustant, J Rohmer, O Alata, J P Naulin, D Idier, R Pedreros, D Moncoulon, and P Tinard. Functional principal component analysis for global sensitivity analysis of model with spatial output. *arXiv e-prints*, May 2020.

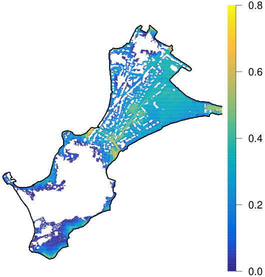
C E Rasmussen and C K I Williams. *GPs for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, Cambridge, MA, 2005.

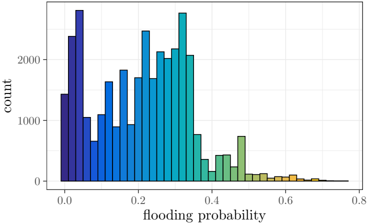
Sami Remes, Markus Heinonen, and Samuel Kaski. Non-stationary spectral kernels. In *Advances in Neural Information Processing Systems*, volume 30, pages 4642–4651. 2017.

M Van der Wilk, V Dutordoir, ST John, A Artemev, V Adam, and J Hensman. A framework for interdomain and multioutput GPs. *arXiv:2003.01115*, 2020.

Géosciences pour une Terre durable

Design of Experiments (DoE)

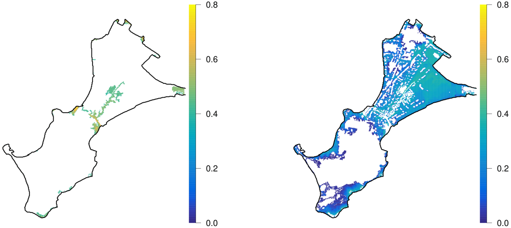




Non-Zero Empirical Flooding Probability (EFP)

- Since the dataset seems "unbalanced", we split it into two classes:
  - **Class 1:** spatial locations with EFP $\in [0.4, 0.8]$ ($\sim$ 2.5k points).
  - **Class 2:** spatial locations with EFP $\in (0, 0.4)$ ($\sim$ 31.5k points).

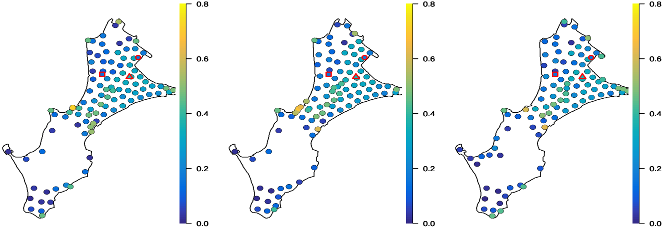
Design of Experiments (DoE)



Class 1: EFP $\in [0.4, 0.8]$ Class 2: EFP $\in (0, 0.4)$

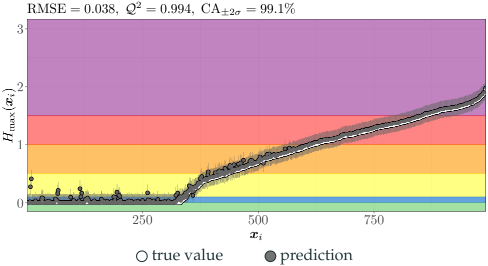
Non-Zero Empirical Flooding Probability (EFP)

- We then propose DoEs based on k-means clustering: inputs $= (x_1, x_2, \text{EFP})$



Replicate 1 Replicate 2 Replicate 3

Assessment using the flood categories from the French Risk Prevention Plan



- minor: $H_{\mathsf{max}} \leq 0.5$
- moderate: $0.5 < H_{\mathsf{max}} \leq 1$
- serious: $1 < H_{\mathsf{max}} \leq 1.5$
- severe: $H_{\mathsf{max}} > 1.5$

Flood Category	Proportions [%] per Category								
	Scenario 1			Scenario 43			Scenario 100		
	H_{max}	$\widehat{H}^{(103)}_{\mathsf{max}}$	$\widehat{H}^{(1003)}_{\mathsf{max}}$	H_{max}	$\widehat{H}^{(103)}_{\mathsf{max}}$	$\widehat{H}^{(1003)}_{\mathsf{max}}$	H_{max}	$\widehat{H}^{(103)}_{\mathsf{max}}$	$\widehat{H}^{(1003)}_{\mathsf{max}}$
minor	99.9	**99.7**	99.6	90.9	**86.6**	83.3	50.6	51.3	**50.8**
moderate	0.1	**0.3**	**0.3**	7.9	**12.0**	15.4	16.1	**17.3**	19.9
serious	0.0	**0.0**	0.1	0.9	0.8	**0.9**	20.0	**20.9**	18.0
severe	0.0	**0.0**	**0.0**	0.3	0.6	**0.4**	13.4	10.5	**11.2**

Proportions computed on a dataset based on **131 flood events**