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

DRP seminar

Andrés F. López-Lopera

September 17, 2020

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...

F. Bachoc
IMT, France

D. Idier
BRGM, France

J. Rohmer
BRGM, France

MULTI-OUTPUT GAUSSIAN PROCESSES WITH FUNCTIONAL DATA: A STUDY ON COASTAL FLOOD HAZARD ASSESSMENT

Andrés F. López-Lopera[†,‡]
andres-felipe.lopez@math.univ-toulouse.fr

Déborah Idier[‡]
d.idier@brgm.fr

Jérémy Rohmer[‡]
j.rohmer@brgm.fr

François Bachoc[†]
francois.bachoc@math.univ-toulouse.fr

[†] Institut de Mathématiques de Toulouse, Université Paul Sabatier, F-31062 Toulouse, France
[‡] BRGM, DRP/R3C, 3 avenue Claude Guillemin, F-45060 Orléans cédex 2, France

ABSTRACT. Most of the existing coastal flood Forecast and Early-Warning Systems do not model the flood, but instead, rely on the prediction of hydrodynamic conditions at the coast and on expert judgment. Recent scientific contributions are now capable to precisely model flood events, even in situations where wave overtopping plays a significant role. Such models are nevertheless costly-to-evaluate and surrogate ones need to be exploited for substantial computational savings. For the latter models, the hydro-meteorological forcing conditions (inputs) or flood events (outputs) are conveniently parametrised into scalar representations. However, they neglect the fact that inputs are actually functions (more precisely, time series), and that floods spatially propagate inland. Here, we introduce a multi-output Gaussian process model accounting for both criteria. On various examples, we test its versatility for both learning spatial maps and inferring unobserved ones. We demonstrate that efficient implementations are obtained by considering tensor-structured data and/or sparse-variational approximations. Finally, the proposed framework is applied on a coastal application aiming at predicting flood events. We conclude that accurate predictions are obtained in the order of minutes rather than the couples of days required by dedicated hydrodynamic simulators.

Source: https://arxiv.org/abs/2007.14052

Géosciences pour une Terre durable

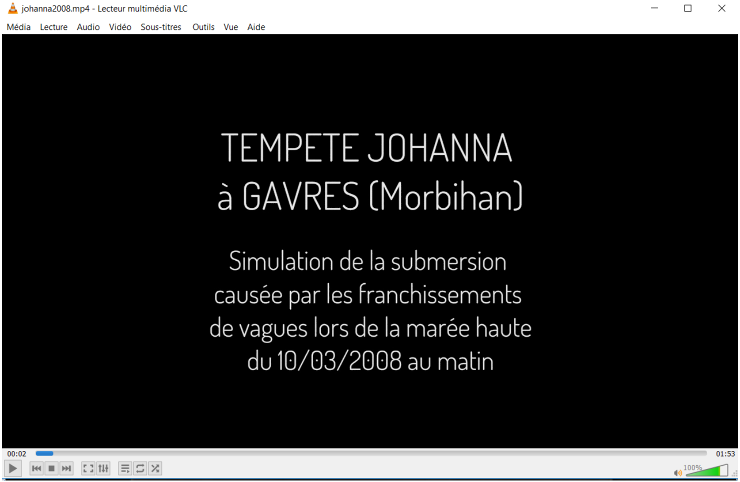
Table of contents

Institut de Mathématiques de Toulouse

brgm Géosciences pour une Terre durable

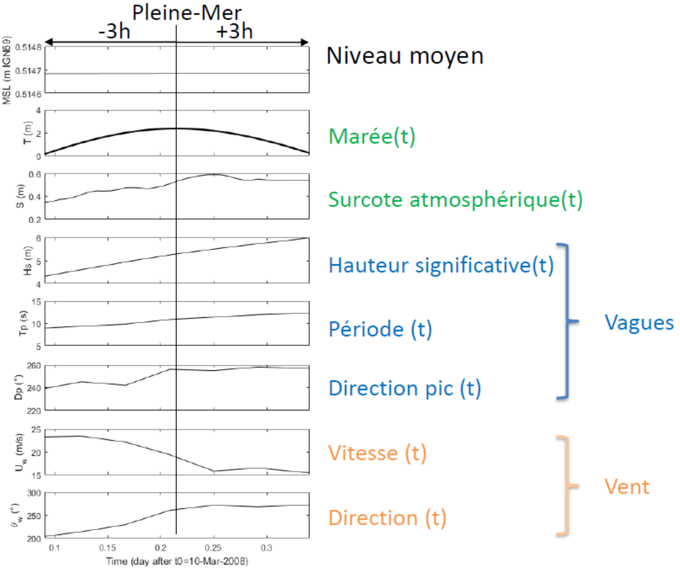
Motivation

Study site and dataset

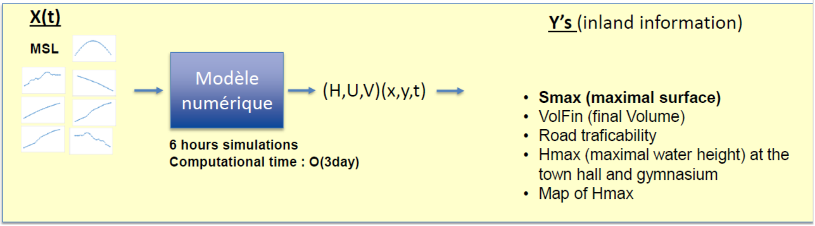


Géosciences pour une Terre durable

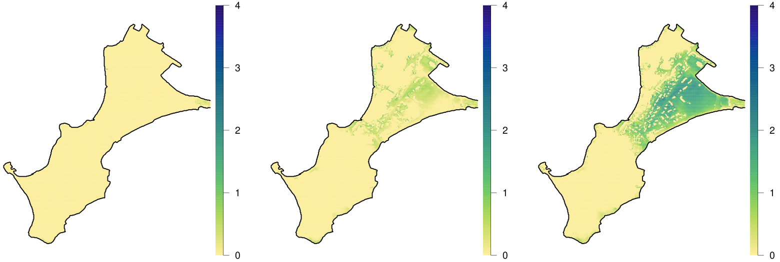
Study site and dataset



Study site and dataset



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

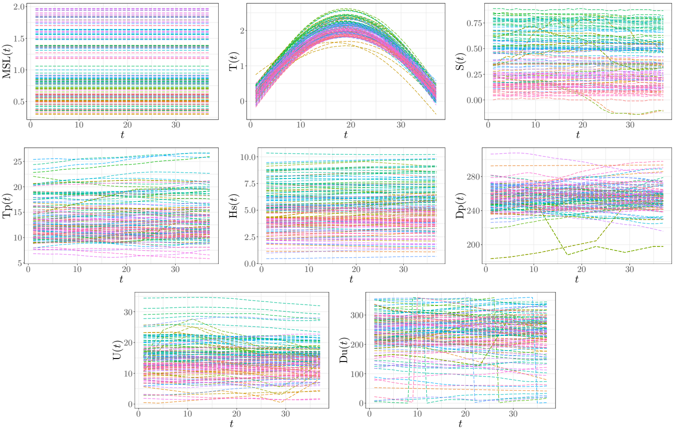


Challenge 1:

· Each flood event takes ∼ **3 days of simulation**.

Challenges

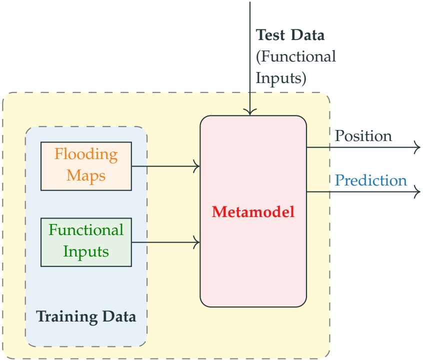
Drivers: hydro-meteorological conditions (tide, surge, wind speed, etc.)



Challenge 2:

- To consider inputs as functions (time-series) rather than scalars.

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



· This will lead to faster (approximate) predictions.

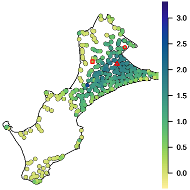
Application: forecasting and early warning systems.

Research subject

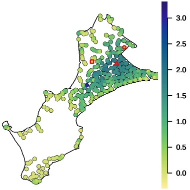
- To do so...
  - We further investigate a **metamodel** based on **Gaussian process**.
  - This results in a **multi-output process**:
    - correlated spatial flood events are driven by (functional) hydro-meteorological inputs.
  - Efficient implementations are based on **Python and R codes**.

Research subject

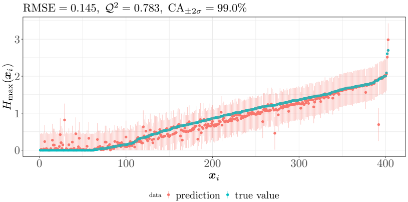
- To do so...
  - We further investigate a **metamodel** based on **Gaussian process**.
  - This results in a **multi-output process**:
    - correlated spatial flood events are driven by (functional) hydro-meteorological inputs.
  - Efficient implementations are based on **Python and R codes**.
- Our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



ground truth



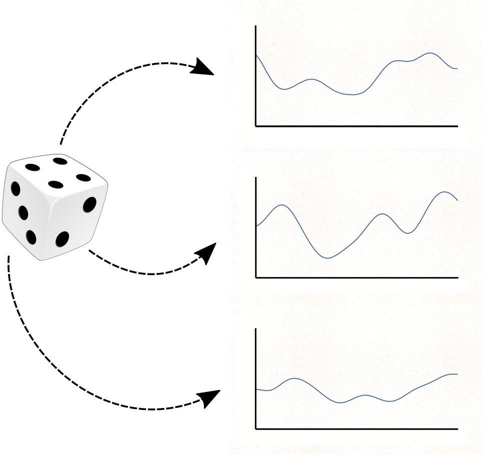
prediction



ground truth vs prediction

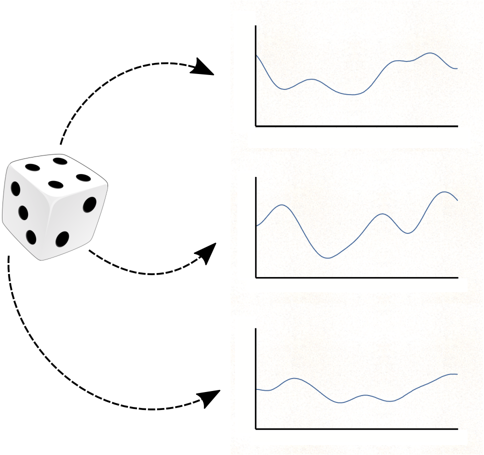
Spatial Gaussian processes

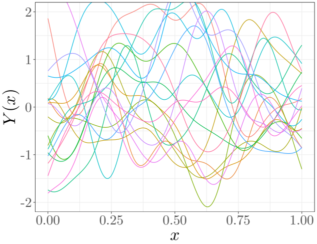
- A **random process** can be defined as a collection of *random variables (r.v.'s)*.



Random processes

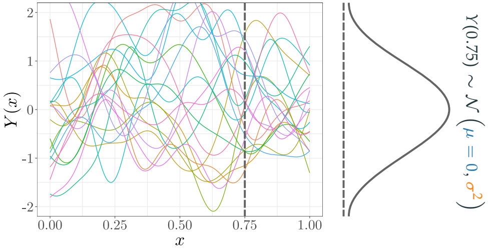
· A **random process** can be defined as a collection of *random variables (r.v.'s)*.





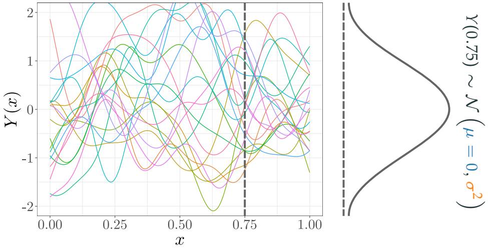
Gaussian processes (GPs)

- A GP $\{Y(x); x \in \mathbb{R}\}$ is then defined as a collection of *Gaussian r.v.'s*:



Gaussian processes (GPs)

· A GP $\{Y(x); x \in \mathbb{R}\}$ is then defined as a collection of *Gaussian r.v.'s*:



· For a given finite number of design points $(x_1, \cdots, x_n)$:

$$\mathbf{Y}_n = \left[Y(x_1), \cdots, Y(x_n)\right]^\top \sim \mathcal{N}\left(\boldsymbol{\mu}(= \mathbf{0}), \mathbf{K}\right), \tag{1}$$

with mean vector μ and covariance matrix $\mathbf{K} = (k(x_i, x_j))_{1 \leq i,j \leq n}$.

· More generally, for the infinite-dimensional case, we have:

$$Y \sim \mathcal{GP}(\mu(= 0), k), \tag{2}$$

with mean function μ and covariance function (kernel) k.

Gaussian processes (GPs)

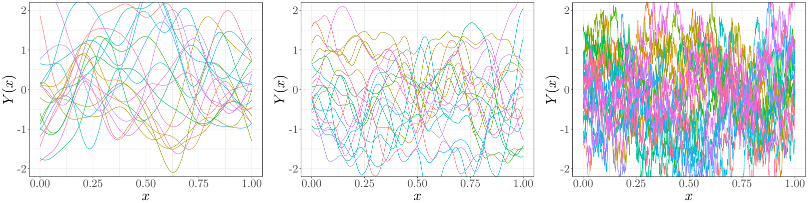
· Regularity assumptions are encoded in the kernel k:

Squared Exponential (SE): $$k_{\sigma^2,\ell}(r) = \sigma^2 \exp\left\{-\frac{1}{2} r_\ell^2\right\},$$

Matérn 5/2: $$k_{\sigma^2,\ell}(r) = \sigma^2 \left(1 + \sqrt{5}|r|_\ell + \frac{5}{3} r_\ell^2\right) \exp\left\{-\sqrt{5}|r|_\ell\right\},$$

Exponential: $$k_{\sigma^2,\ell}(r) = \sigma^2 \exp\left\{-|r|_\ell\right\},$$

with $r_\ell = (x - x')/\ell$, variance σ^2 and length-scale ℓ.



(a) SE (b) Matérn 5/2 (c) Exponential

samples

Effect of different kernels

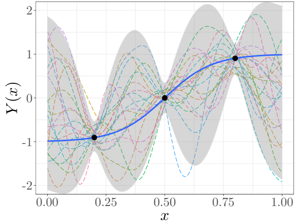
Gaussian process regression

- The distribution of $\mathbf{Y}_*$, knowing $\mathbf{Y}_n = \boldsymbol{y}$ for a given observation vector $\boldsymbol{y} = [y_1, \cdots, y_n]^\top$, is given by

$$\mathbf{Y}_* | \{\mathbf{Y}_n = \boldsymbol{y}\} \sim \mathcal{N}(\boldsymbol{m}, \ \boldsymbol{\Sigma}) \,. \tag{3}$$

with conditional parameters [Rasmussen and Williams, 2005]:

$$\boldsymbol{m} = \mathbf{K}_{\mathbf{Y}_*, \mathbf{Y}_n} \mathbf{K}_{\mathbf{Y}_n, \mathbf{Y}_n}^{-1} \boldsymbol{y}, \quad \boldsymbol{\Sigma} = \mathbf{K}_{\mathbf{Y}_*, \mathbf{Y}_*} - \mathbf{K}_{\mathbf{Y}_*, \mathbf{Y}_n} \mathbf{K}_{\mathbf{Y}_n, \mathbf{Y}_n}^{-1} \mathbf{K}_{\mathbf{Y}_*, \mathbf{Y}_n}^\top . \tag{4}$$



● conditioning observations

■ conditional mean ($\boldsymbol{\mu}$)

■ prediction intervals ($\pm 2\boldsymbol{\sigma} = \pm 2\sqrt{\mathrm{diag}(\boldsymbol{\Sigma})}$)

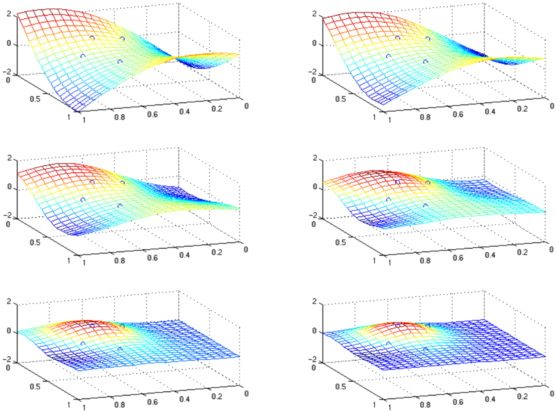
■ ■ ... ■ samples

Note: $\mathbf{K}_{\mathbf{Y}_n, \mathbf{Y}_n}^{-1}$ is challenging when n is large.

[Demo]

Spatial Gaussian processes

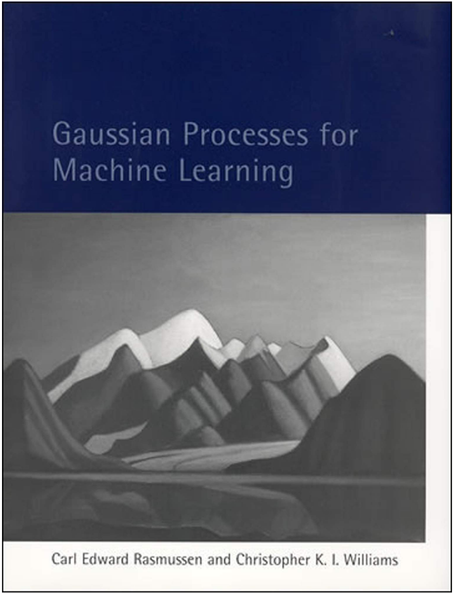
- For $\{Y(\boldsymbol{x}); \boldsymbol{x} \in \mathbb{R}^2\}$ a spatial GP:



- In our case, the spatial process corresponds to a coastal flood one.

Géosciences pour une Terre durable

Gaussian processes for machine learning...



Spatial Gaussian processes with functional inputs

Spatial Gaussian processes with functional inputs

· Let $\{Y(\boldsymbol{x}, \boldsymbol{\mathcal{F}}); \boldsymbol{x} \in \mathbb{R}^2, \boldsymbol{\mathcal{F}} \in \mathcal{F}(\mathcal{T}, \mathbb{R})^Q\}$ be a GP with spatial inputs $\boldsymbol{x} = (x_1, x_2)$ and functional inputs $\boldsymbol{\mathcal{F}} = (f_1, \cdots, f_Q)$.

Spatial Gaussian processes with functional inputs

· Let $\{Y(\boldsymbol{x}, \boldsymbol{\mathcal{F}}); \boldsymbol{x} \in \mathbb{R}^2, \boldsymbol{\mathcal{F}} \in \mathcal{F}(\mathcal{T}, \mathbb{R})^Q\}$ be a GP with spatial inputs $\boldsymbol{x} = (x_1, x_2)$ and functional inputs $\boldsymbol{\mathcal{F}} = (f_1, \cdots, f_Q)$.

· Then, $Y \sim \mathcal{GP}(0, k)$ with kernel k:

$$k((\boldsymbol{x}, \boldsymbol{\mathcal{F}}), (\boldsymbol{x}', \boldsymbol{\mathcal{F}}')) = \mathsf{cov}\left\{Y(\boldsymbol{x}, \boldsymbol{\mathcal{F}}), Y(\boldsymbol{x}', \boldsymbol{\mathcal{F}}')\right\} \tag{5}$$

· k must be defined according to the structure of data (i.e. regularity).

Géosciences pour une Terre durable

Spatial Gaussian processes with functional inputs

· Let $\{Y(\boldsymbol{x}, \boldsymbol{\mathcal{F}}); \boldsymbol{x} \in \mathbb{R}^2, \boldsymbol{\mathcal{F}} \in \mathcal{F}(\mathcal{T}, \mathbb{R})^Q\}$ be a GP with spatial inputs $\boldsymbol{x} = (x_1, x_2)$ and functional inputs $\boldsymbol{\mathcal{F}} = (f_1, \cdots, f_Q)$.

· Then, $Y \sim \mathcal{GP}(0, k)$ with kernel k:

$$k((\boldsymbol{x}, \boldsymbol{\mathcal{F}}), (\boldsymbol{x}', \boldsymbol{\mathcal{F}}')) = \mathsf{cov}\left\{Y(\boldsymbol{x}, \boldsymbol{\mathcal{F}}), Y(\boldsymbol{x}', \boldsymbol{\mathcal{F}}')\right\} \tag{5}$$

· k must be defined according to the structure of data (i.e. regularity).

· Here, we consider that k is a *separable kernel*:

$$k((\boldsymbol{x}, \boldsymbol{\mathcal{F}}), (\boldsymbol{x}', \boldsymbol{\mathcal{F}}')) = k_s(\boldsymbol{x}, \boldsymbol{x}')k_f(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}'), \tag{6}$$

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

Gaussian processes with functional inputs

· For k_f, we need a *measure of "dissimilarity"* [Betancourt et al., 2020], e.g.:

$$d(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}') = \|\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{F}}'\|_{\boldsymbol{\ell}} = \sqrt{\sum_{i=1}^{Q} \|f_i - f_i'\|_{\ell_i}^2}, \tag{7}$$

with the L^2-norm given by

$$\|f_i - f_i'\|_{\ell_i}^2 = \frac{\int_{\mathcal{T}} (f_i(t) - f_i'(t))^2 dt}{\ell_i{}^2}. \tag{8}$$

· **Note:** f_i, f_i' are projected onto *basis functions* (e.g. PCA) for computing (8).

Gaussian processes with functional inputs

- For k_f, we need a *measure of "dissimilarity"* [Betancourt et al., 2020], e.g.:

$$d(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}') = \|\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{F}}'\|_{\boldsymbol{\ell}} = \sqrt{\sum_{i=1}^{Q} \|f_i - f_i'\|_{\ell_i}^2}, \tag{7}$$

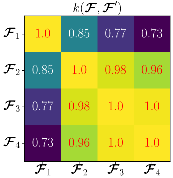
with the L^2-norm given by

$$\|f_i - f_i'\|_{\ell_i}^2 = \frac{\int_{\mathcal{T}} (f_i(t) - f_i'(t))^2 dt}{{\ell_i}^2}. \tag{8}$$

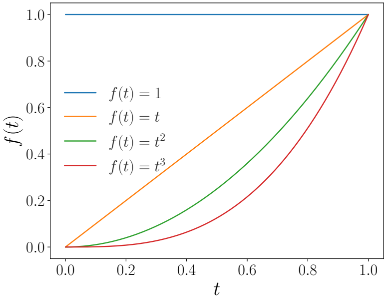
- **Note:** f_i, f_i' are projected onto *basis functions* (e.g. PCA) for computing (8).
- Then, an example of a valid kernel is then given by

$$\text{(Squared Exponential)} \quad k_{f,\sigma^2,\boldsymbol{\ell}}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}') = \sigma^2 \exp\left\{-\frac{\|\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{F}}'\|_{\boldsymbol{\ell}}^2}{2}\right\}. \tag{9}$$

Gaussian processes with functional inputs



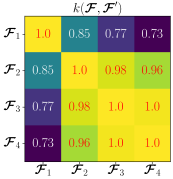
(a) Squared Exponential



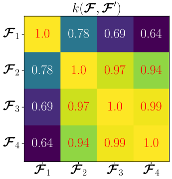
(e) target functions

Effect of the kernels considering $\mathcal{F}_1 = (f(t) = 1)$, $\mathcal{F}_2 = (f(t) = t)$, $\mathcal{F}_3 = (f(t) = t^2)$ and $\mathcal{F}_4 = (f(t) = t^3)$.

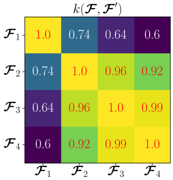
Gaussian processes with functional inputs



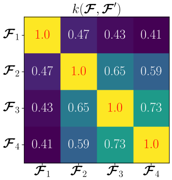
(a) Squared Exponential



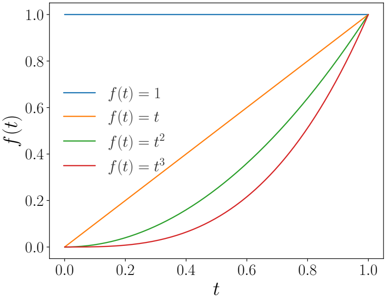
(b) Matérn 5/2



(c) Matérn 3/2



(d) Exponential



(e) target functions

Effect of the kernels considering $\mathcal{F}_1 = (f(t) = 1)$, $\mathcal{F}_2 = (f(t) = t)$, $\mathcal{F}_3 = (f(t) = t^2)$ and $\mathcal{F}_4 = (f(t) = t^3)$.

Link with multi-output Gaussian processes

· Note that Y can be written as a multi-output process Z:

$$Z_i(\boldsymbol{x}) := Y(\boldsymbol{\mathcal{F}}_i, \boldsymbol{x}), \quad \text{for } i = 1, \cdots, R. \tag{10}$$

· In that case, k can be rewritten as:

$$k_{i,j}(\boldsymbol{x}, \boldsymbol{x}') = b_{i,j} \, k_s(\boldsymbol{x}, \boldsymbol{x}'), \tag{11}$$

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

Géosciences pour une Terre durable

Link with multi-output Gaussian processes

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

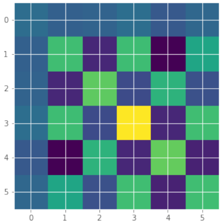
$$Z_i(x) := Y(\mathcal{F}_i, x), \quad \text{for } i = 1, \cdots, R. \tag{10}$$

- In that case, k can be rewritten as:

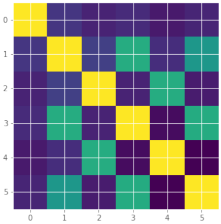
$$k_{i,j}(x, x') = b_{i,j}\, k_s(x, x'), \tag{11}$$

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

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



B

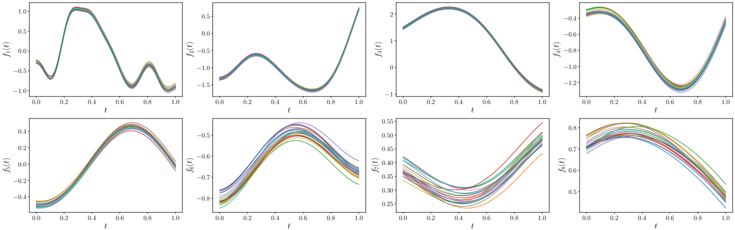


$k_f(\mathcal{F}, \mathcal{F}')$

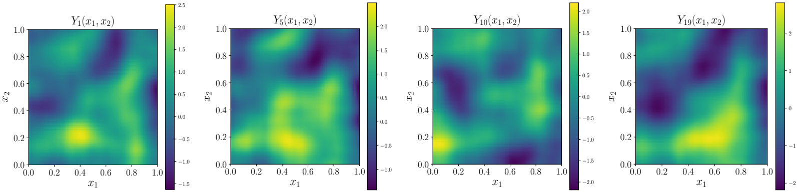
Coregionalisation matrix

Numerical illustration: multi-output task

· Synthetic example with **8 functional inputs** and **20 spatial outputs**.

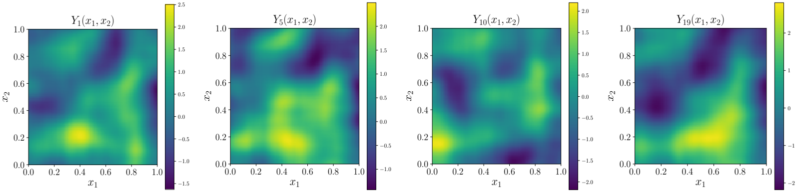


functional replicates

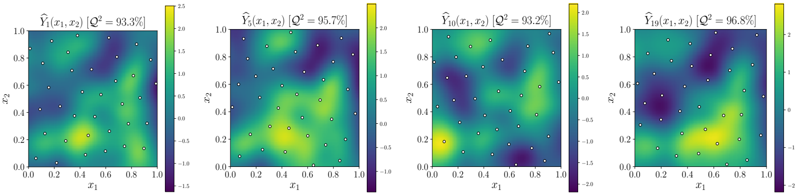


synthetic 100 × 100 maps

Numerical illustration: multi-output task



synthetic 100×100 maps



predictions (35 spatial design points per map)

$$\mathcal{Q}^2 = 1 - \mathrm{SMSE}(\mathbf{Y}, \widehat{\mathbf{Y}})$$

Géosciences pour une Terre durable

Spatial Gaussian processes with functional inputs

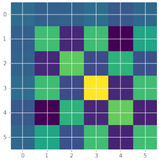
Advantages

Our approach accounts for both spatial outputs and functional inputs

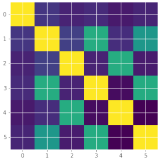
Regularity assumptions are encoded into **kernels**:

$$k((x,\mathcal{F}),(x',\mathcal{F}')) = k_s(x,x')k_f(\mathcal{F},\mathcal{F}').$$

The *coregionalisation matrix* enjoys of **positive semi-definitiveness**



B



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

Disadvantage

The **computational complexity**: $\mathcal{O}(N^3)$ with $N = R \times S$

- R: number of functional inputs
- S: number of spatial design points per map

Extension to large datasets

Extension to large datasets

- To scale the proposed GP framework to large datasets, we can:
  - via *Kronecker-products* [see, e.g., Alvarez et al., 2012],
  - via *dimension reduction* of the output space [see, e.g., Perrin et al., 2020],
  - via *sparse-variational inference* [see, e.g., Van der Wilk et al., 2020].

Extension to large datasets

- To scale the proposed GP framework to large datasets, we can:
  - via *Kronecker-products* [see, e.g., Alvarez et al., 2012],
  - via *dimension reduction* of the output space [see, e.g., Perrin et al., 2020],
  - via *sparse-variational inference* [see, e.g., Van der Wilk et al., 2020].

Kronecker-based implementations

· By considering tensor-structured data, we have:

$$\mathbf{K} = \mathbf{K}_s \otimes \mathbf{K}_f. \tag{12}$$

· Using (12) leads to simplifications that can be exploited, e.g.:

$$\mathbf{K}^{-1} = \mathbf{K}_s^{-1} \otimes \mathbf{K}_f^{-1}, \tag{13}$$

$$\mathbf{L} = \mathbf{L}_s \otimes \mathbf{L}_f,$$

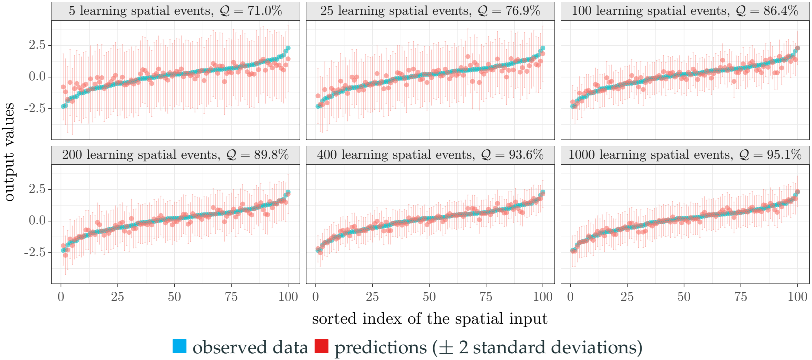
where $\mathbf{L}$, $\mathbf{L}_s$, $\mathbf{L}_f$ are (lower triangular) Cholesky matrices of $\mathbf{K}$, $\mathbf{K}_s$, $\mathbf{K}_f$.

· Further efficient computations can also be proposed for **forwardly (or backwardly) solving systems of linear equations**:

$$z = (\mathbf{L}_s \otimes \mathbf{L}_f)y. \tag{14}$$

Numerical illustration: forecasting task

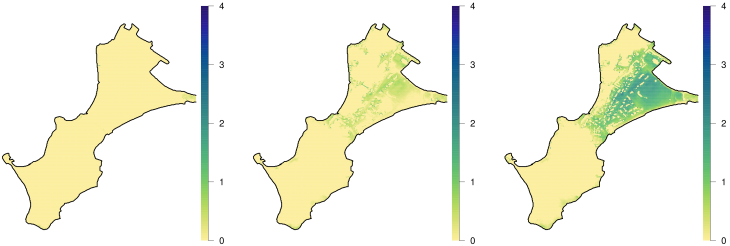
- **Synthetic:** **8 functional inputs** and **1001 spatial outputs**
- **Goal:** predict the map 1001 using data from the other ones



Coastal flooding application

Coastal flooding application

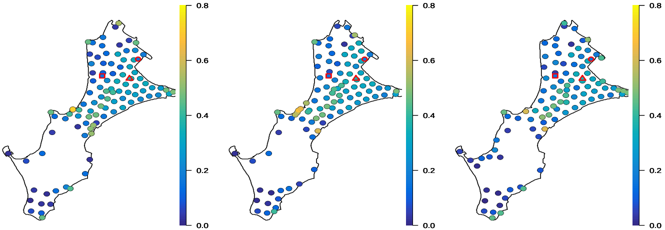
Spatial flood events: maximal inland water level (H_{max} [m])



- **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)

Institut de Mathématiques de Toulouse
brgm Géosciences pour une Terre durable

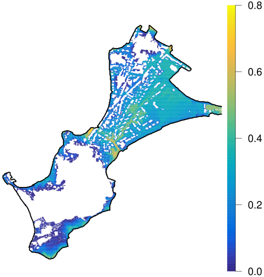
· Due to computational limitations, we focus on a tractable amount of points.

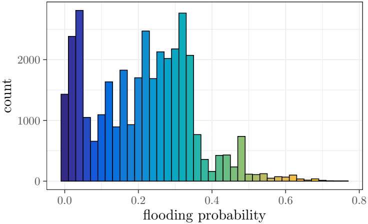


Replicate 1 Replicate 2 Replicate 3

Examples of DoE considering 103 spatial locations

Design of Experiments (DoE)

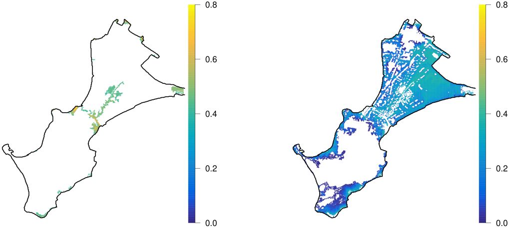




Non-Zero Empirical Flooding Probability (EFP)

- 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).

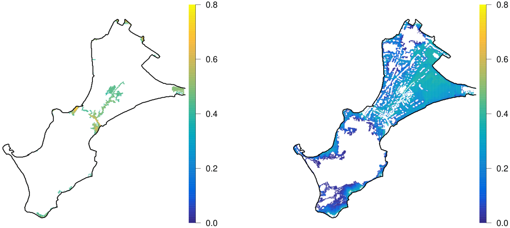
Design of Experiments (DoE)



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

Non-Zero Empirical Flooding Probability (EFP)

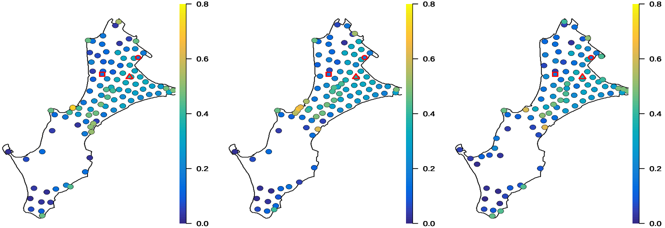
Design of Experiments (DoE)



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

Non-Zero Empirical Flooding Probability (EFP)

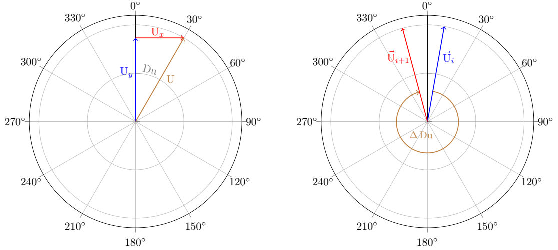
- We then propose DoEs based on k-means clustering: inputs $= (x_1, x_2, \text{EFP})$



Replicate 1 Replicate 2 Replicate 3

Data preprocessing

· Some hydro-meteorological inputs are defined in the *nautical convention*:



· We replace (Hs, Dp) and (U, Du) by the Cartesian tuples:

$$\mathsf{Hs}_x = \mathsf{Hs}\cdot\sin(\mathsf{Dp}),\ \mathsf{Hs}_y = \mathsf{Hs}\cdot\cos(\mathsf{Dp}),\ \mathsf{U}_x = \mathsf{U}\cdot\sin(\mathsf{Du}),\ \mathsf{U}_y = \mathsf{U}\cdot\cos(\mathsf{Du}).$$

· This results in a set of **8 functional inputs**:

Inputs = (MSL, Tide, Surge, Tp, Hs_x, Hs_y, U_x, U_y)

Leave-one-out (LOO) test

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

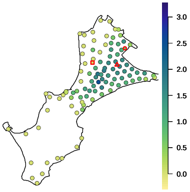
- Therefore, **174 GP metamodels** are trained and tested.
- We fix **103 spatial design points per map**.
  - including the town-hall (■), gym (●) and sports field (▲) at Gâvres.
- We use **Matérn 5/2 kernels** as covariance matrices.
- We use a **PCA representation** of the functional inputs.
  - A 99.9% of the inertia led to:

$$\mathbf{p} = [1, 4, 3, 3, 3, 6, 5, 5] \quad (\mathsf{MSL}, \mathrm{Tide}, \mathrm{Surge}, \mathrm{Tp}, \mathsf{Hs}_x, \mathsf{Hs}_y, \mathsf{U}_x, \mathsf{U}_y)$$

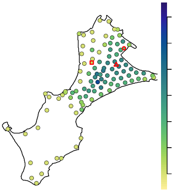
Leave-one-out (LOO) test

Flooding scenario 136: significant flood event

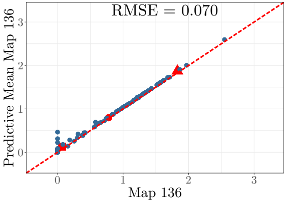
■ Town-Hall ● Gym ▲ Sports Field



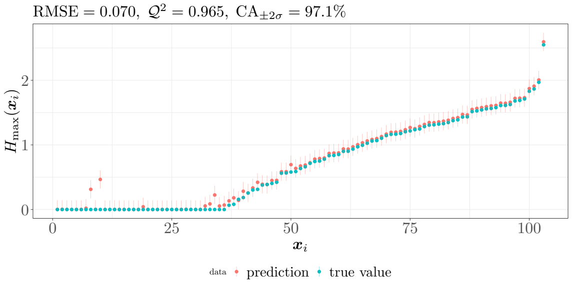
ground truth



prediction



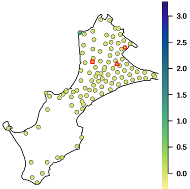
ground truth vs prediction



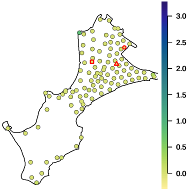
Leave-one-out (LOO) test

Flooding scenario 1: minor flood event

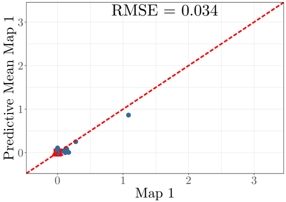
■ Town-Hall ● Gym ▲ Sports Field



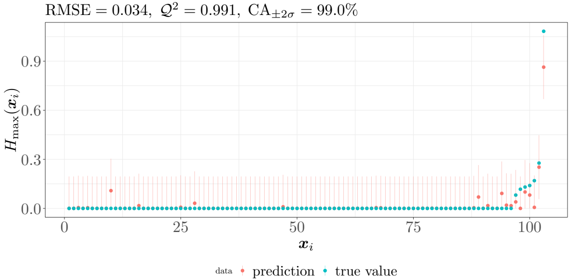
ground truth



prediction

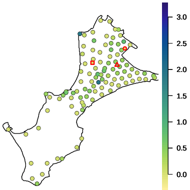


ground truth vs prediction

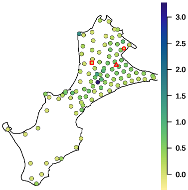


Flooding scenario 43: moderate flood event

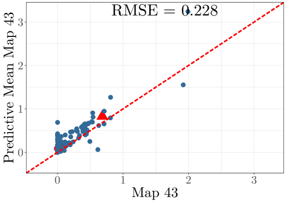
■ Town-Hall ● Gym ▲ Sports Field



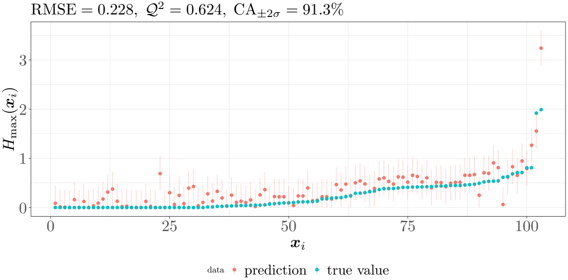
ground truth



prediction



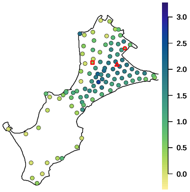
ground truth vs prediction



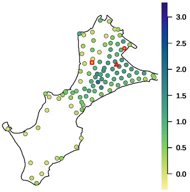
Leave-one-out (LOO) test

Flooding scenario 80: misprediction

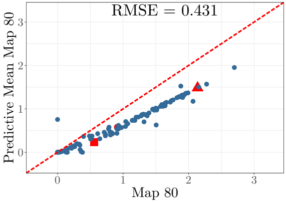
■ Town-Hall ● Gym ▲ Sports Field



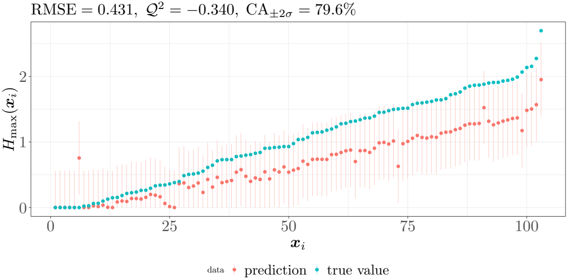
ground truth



prediction

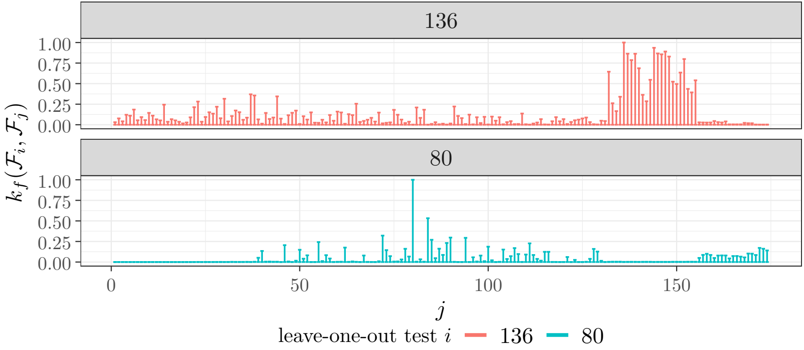


ground truth vs prediction



Géosciences pour une Terre durable

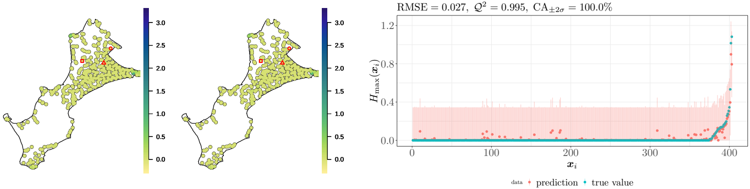
Flooding scenario 80: misprediction (discussion)



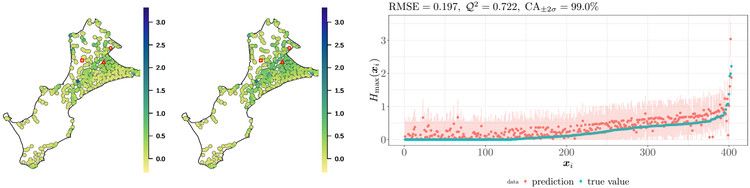
Correlations between functional inputs after covariance parameter estimation via ML

Influence of the number of design points

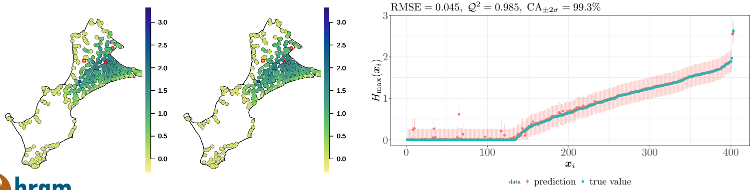
Flooding map 1



Flooding map 43

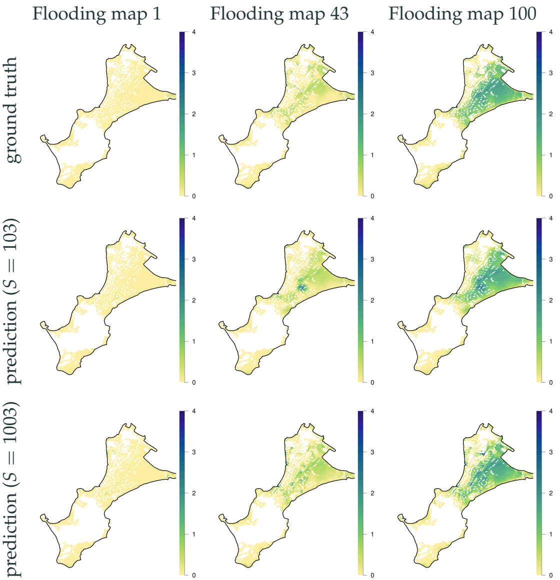


Flooding map 136

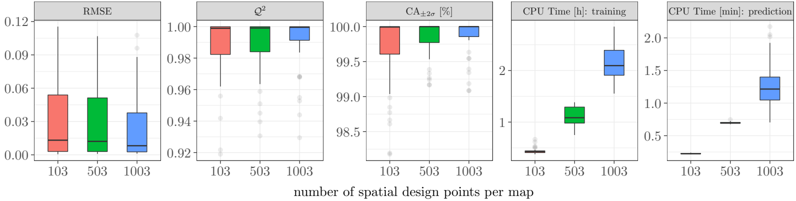


brgm
Géosciences pour une Terre durable

Influence of the number of design points



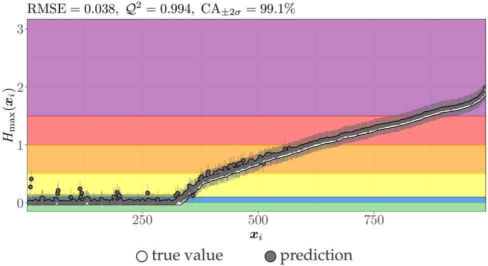
Influence of the number of design points



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

Influence of the number of design points

Assessment using the flood categories suggested by 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**

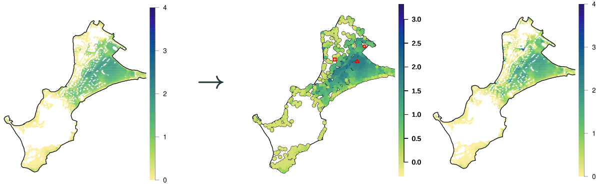
Proportions computed on a dataset based on **131 flood events**

brgm
Géosciences pour une Terre durable

Advantages

Approximate predictions under **smoothness assumptions**.

Numerical simulator GP metamodel



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

Challenges:

Non-stationary kernels for heterogeneous data [e.g., Remes et al., 2017].

Dimension reduction of the output space:

- e.g. via Wavelets decomposition [e.g., Perrin et al., 2020].

Non-negative and/or **zero-inflated GP priors** for flood events.

Conclusions

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 key.
- 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