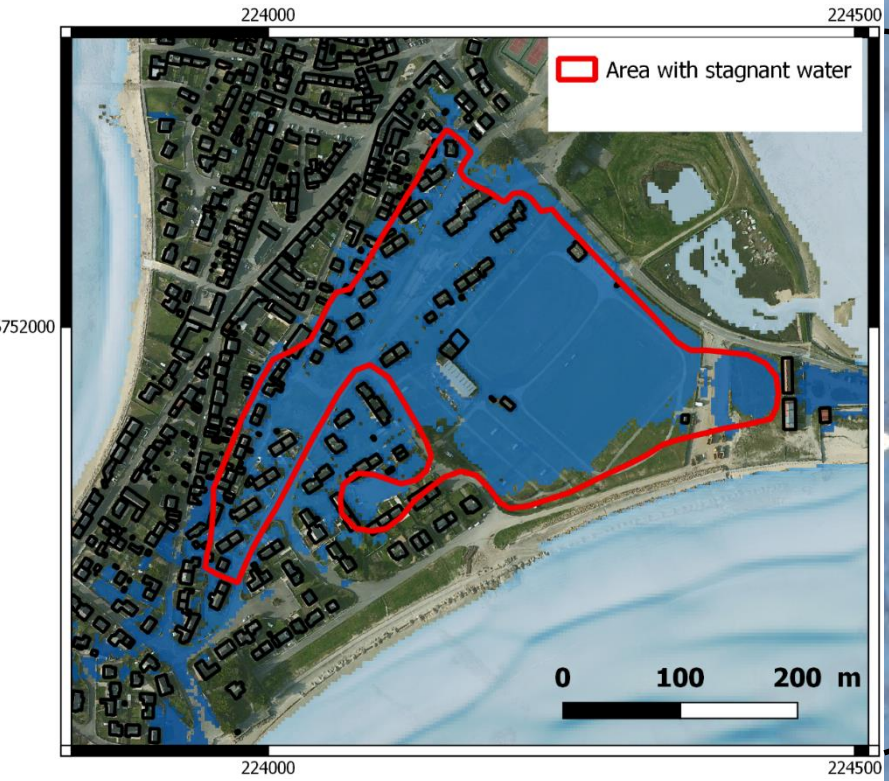
Présentation des méthodes (métamodèles) et résultats Riscope

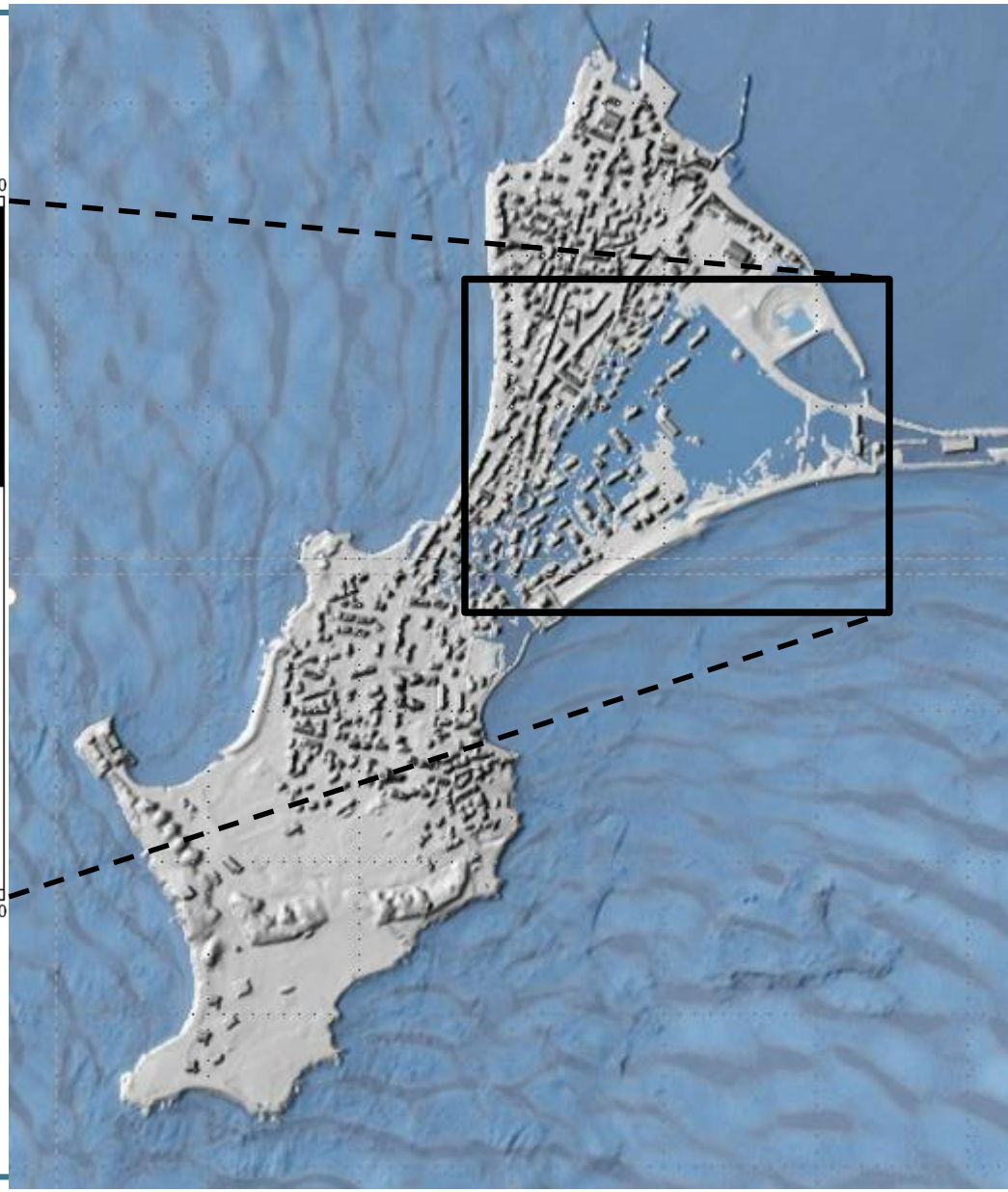
J. Betancourt, A. Lopez-Lopera, J. Rohmer

Séminaire de restitution du projet ANR RISCOPE, 1-2 février 2022

Motivation

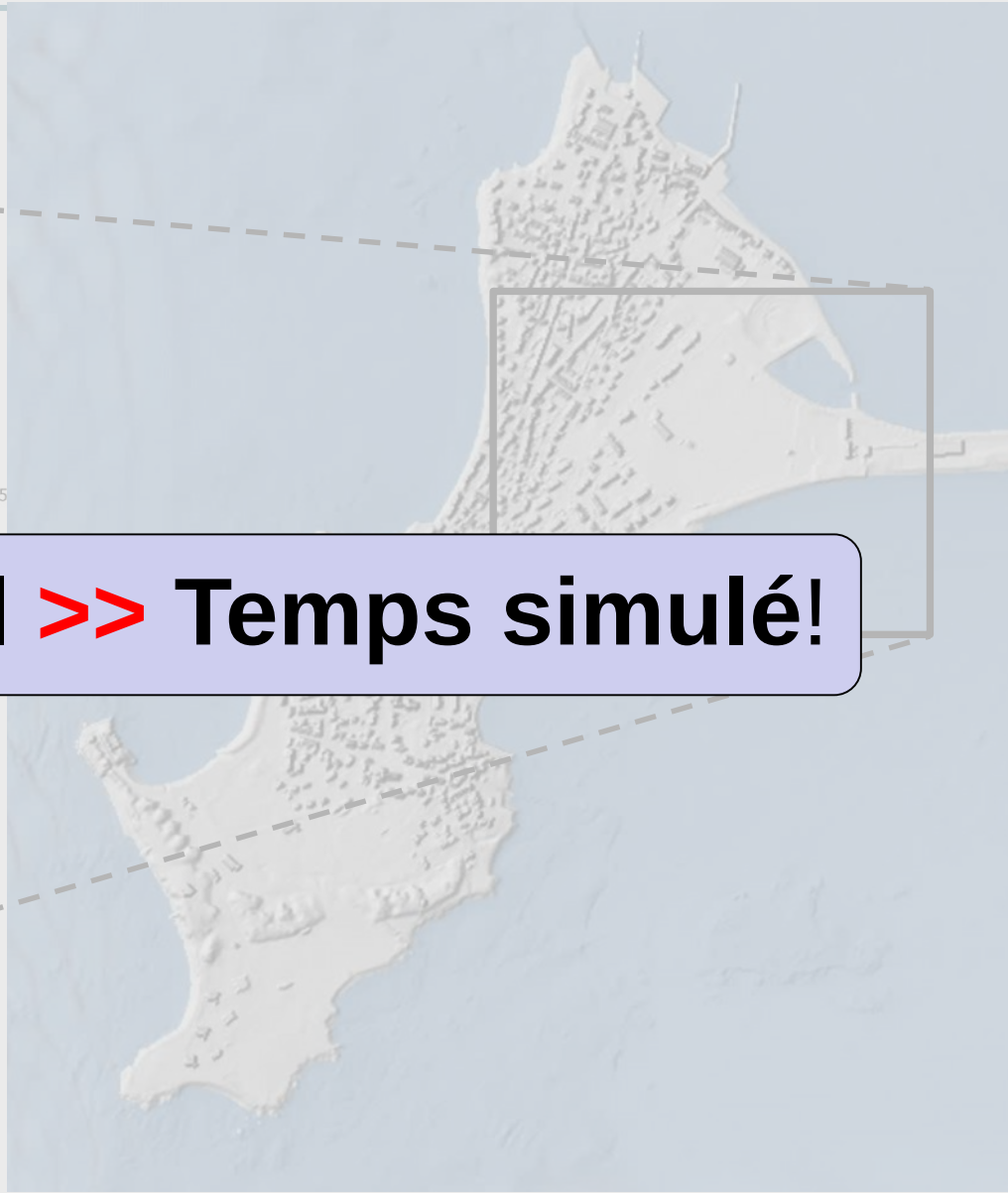
Tempête Johanna, 2008





Motivation

Tempête Johanna, 2008



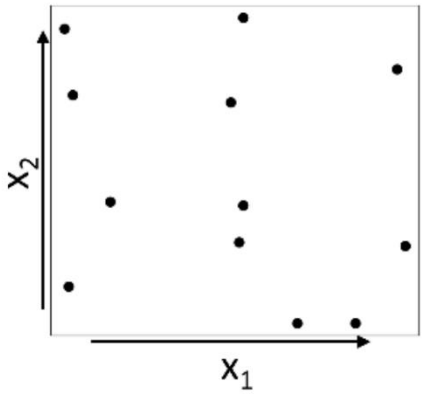
Temps de calcul >> Temps simulé!

Une solution possible (?)

https://www.lebigdata.fr/cluster-definition

Alternative: **métamodélisation**

Design of Experiments (DoE)
Points to perform simulations



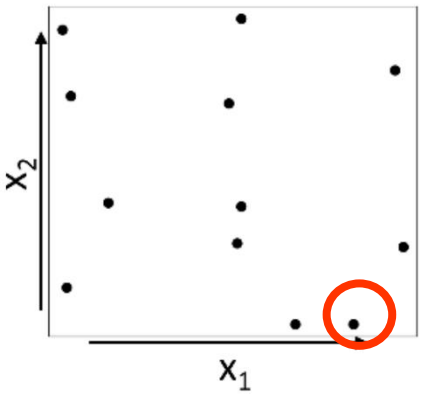
X1, X2 : paramètres d'entrée du simulateur numérique

Alternative: **métamodélisation**

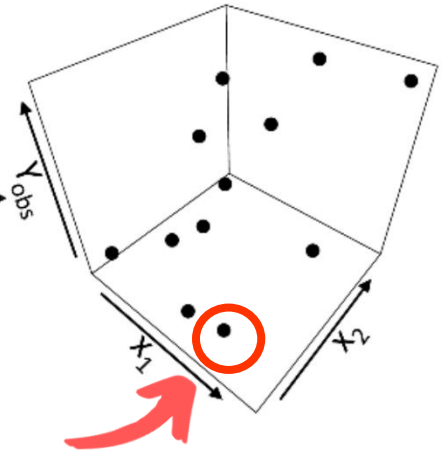
Cout de temps de calcul!

Design of Experiments (DoE)
Points to perform simulations

Simulation
Performing simulations for the points of the DoE



Computer code

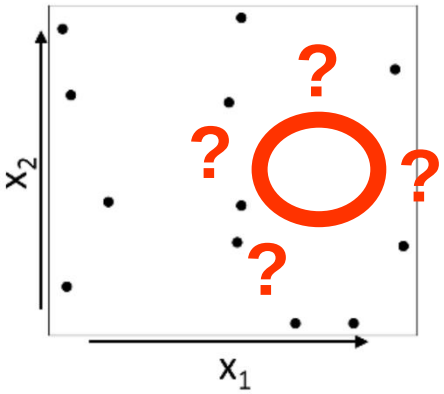


Chaque point =
un résultat de simulation couteuse pour **X1,X2** donné

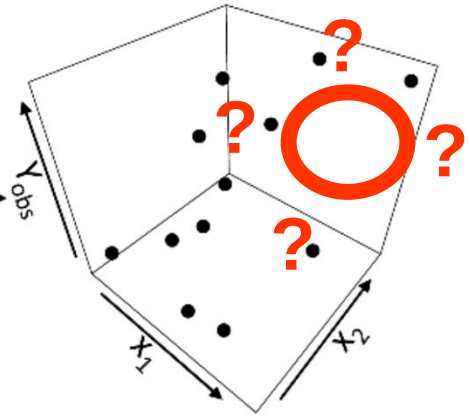
Alternative: **métamodélisation**

Design of Experiments (DoE)
Points to perform simulations

Simulation
Performing simulations for the points of the DoE



Computer code

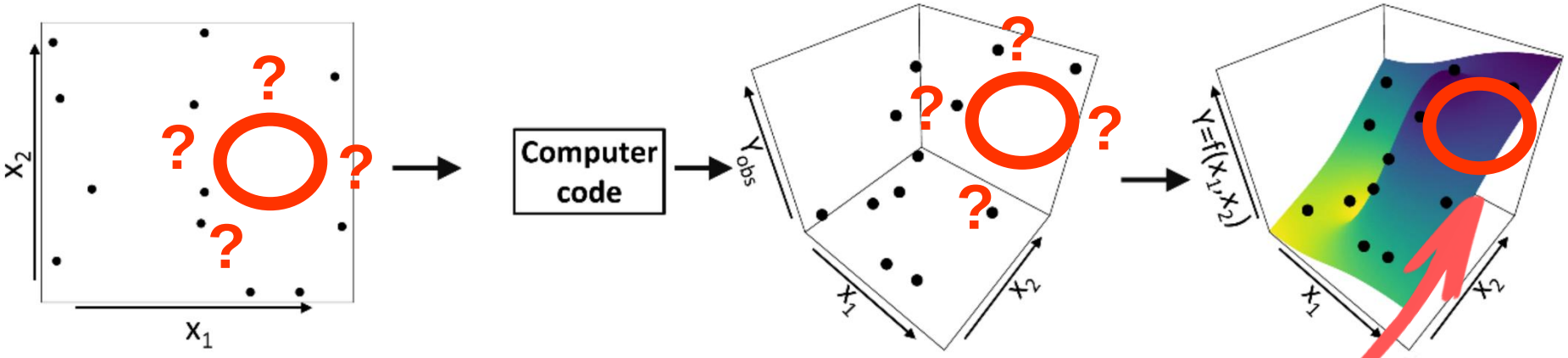


Alternative: **métamodélisation**

Design of Experiments (DoE)
Points to perform simulations

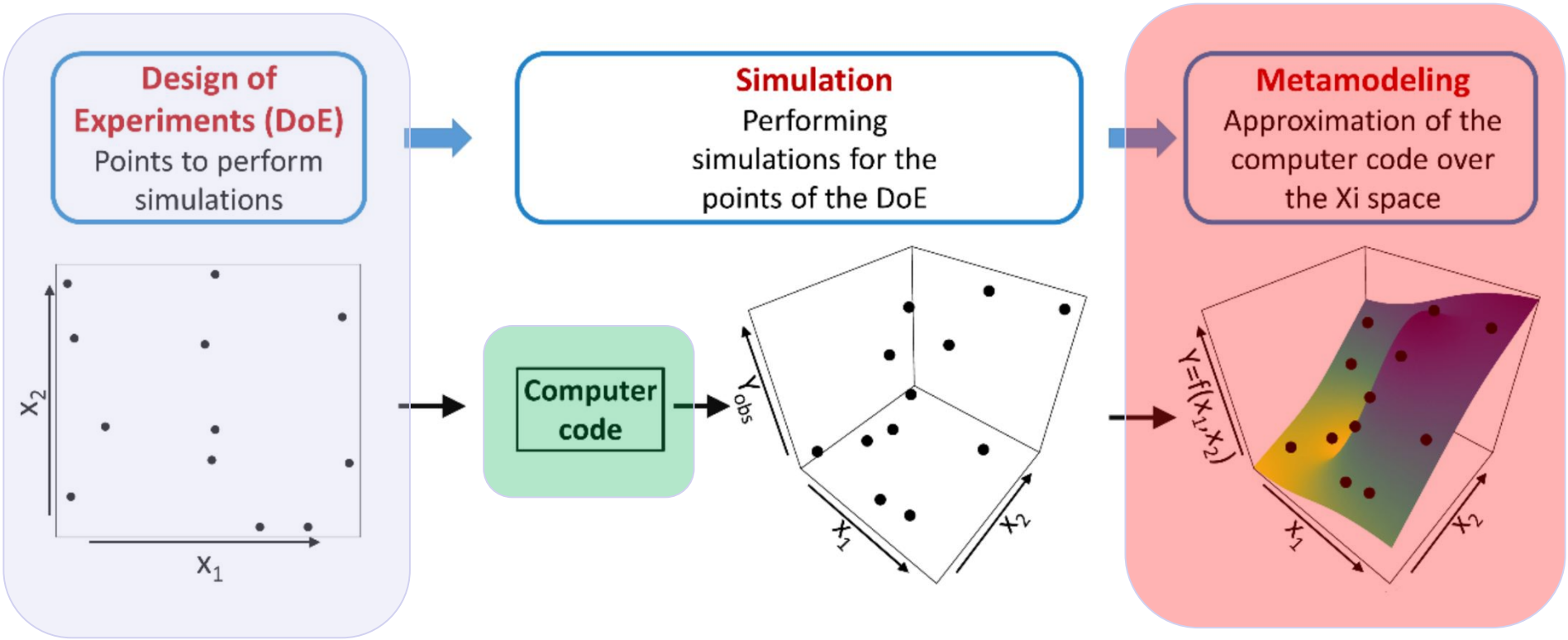
Simulation
Performing simulations for the points of the DoE

Metamodeling
Approximation of the computer code over the Xi space



Remplir les 'trous'
avec un modèle statistique
(=métamodèle)

Alternative: **métamodélisation**



Présentation (15h40, 01/02/22)

Présentation de Rodrigo Pedreros (14h30, 01/02/22)

Présentation aujourd'hui

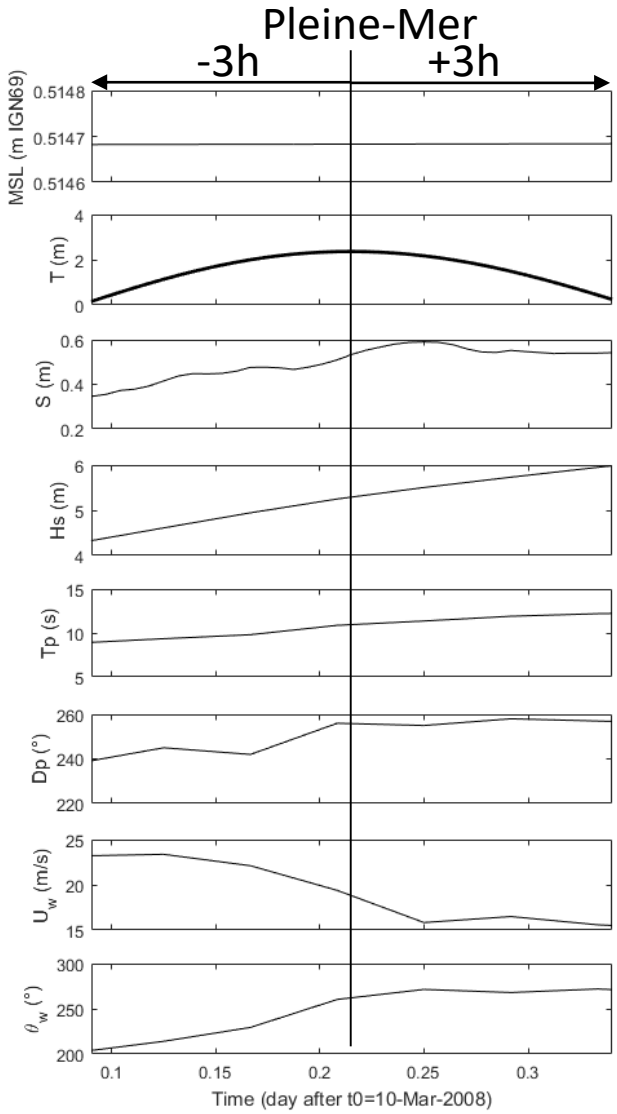
- **Des modèleS de « Méta-Modèle »**
  - Régression linéaire, polynomiale
  - Méthode non-paramétrique « spline »
  - Processus Gaussien (Krigeage)
  - Méthodes de machine learning
  - Etc….

Choix dans le projet

- **Validité?**
  - Comparaison à des résultats de simulation **différents** de ceux utilisés pour construire le méta-modèle
  - **Validation croisée**

[e.g. Forrester et al. (2008); Hastie et al. (2009)]

Difficulté (1) : forçages temporels en entrée X(t)



Niveau moyen

Marée(t)

Surcote atmosphérique(t)

Hauteur significative(t)

Période (t)

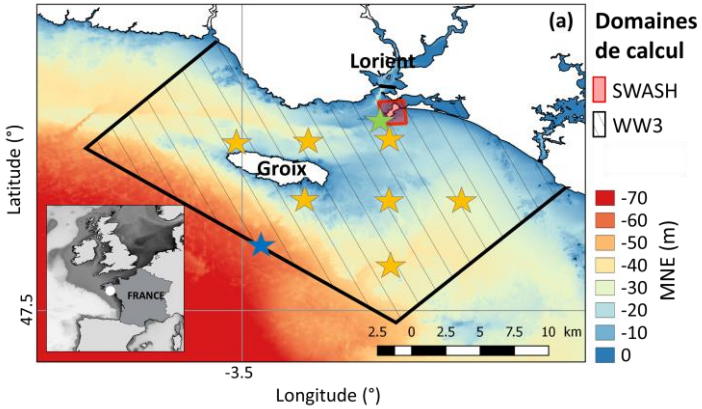
Direction pic (t)

Vagues

Vitesse (t)

Direction (t)

Vent



Difficulté (2) : une information = un métamodèle

- **List of required metamodels**

Global indicators

- Max flooded surface ..... Smax
- Final cumulative volume ..... Vfin
- Mean flow rate ..... Qmn
- Max flow rate ..... Qmx

Location-specific indicators

- Max water height - mairie ..... Hmax_c
- Max water height - gymnase ..... Hmax_g
- Max water height - cuvette ..... Hmax_c
- Trafficability all pixels - road 1 ..... TRt_1
- Trafficability high pixels - road 1 ..... TRh_1

→ Trafficability indicator (TRC_1)

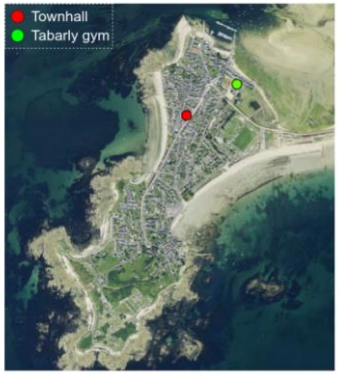
- Trafficability all pixels - road 2 ..... TRt_2
- Trafficability high pixels - road 2 ..... TRh_2

→ Trafficability indicator (TRC_2)

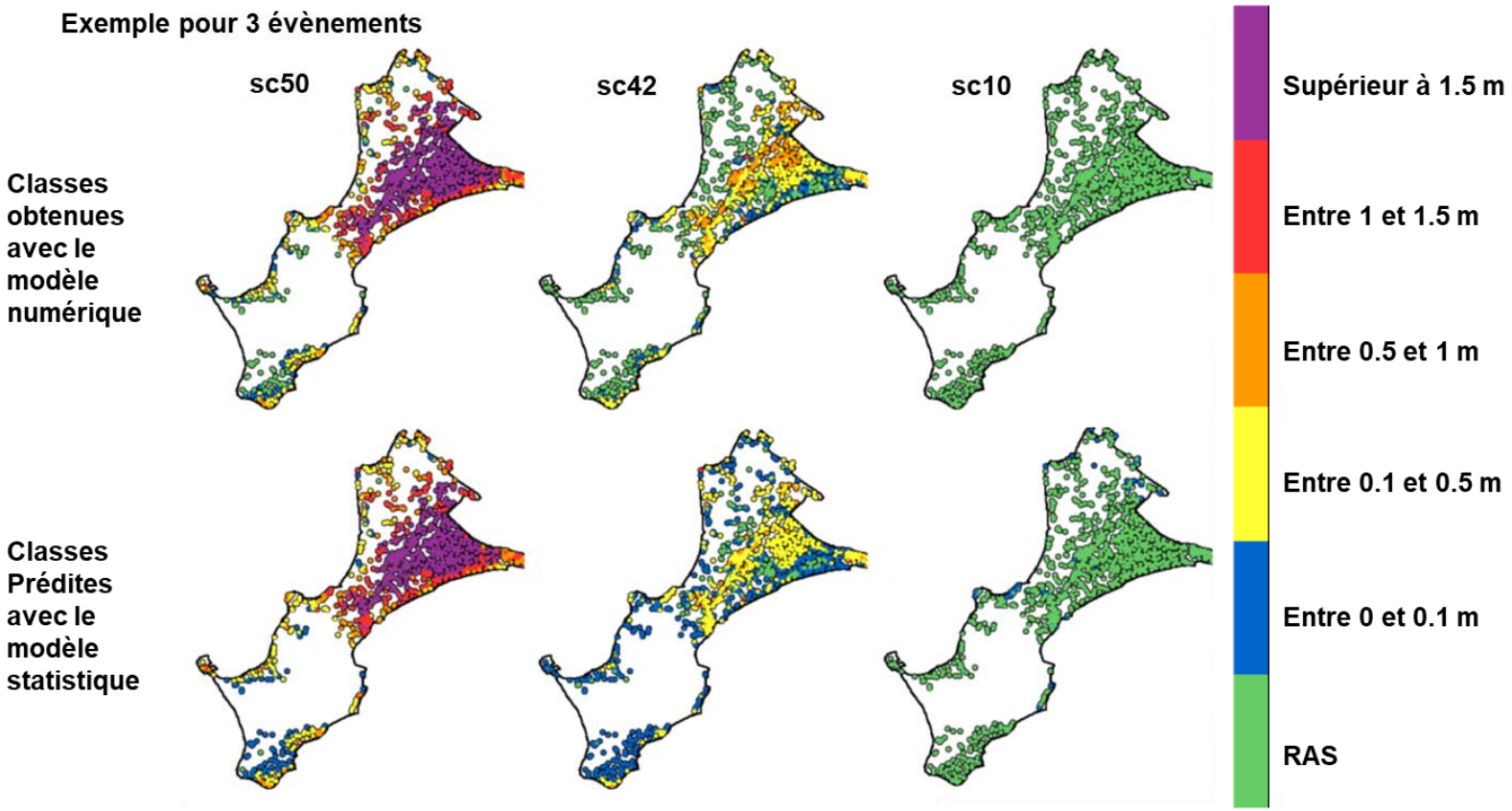
- Trafficability all pixels - road 3 ..... TRt_3
- Trafficability high pixels - road 3 ..... TRh_3

→ Trafficability indicator (TRC_3)

- Flood map
- Flood intensity indicator
- Max water height – monitoring location

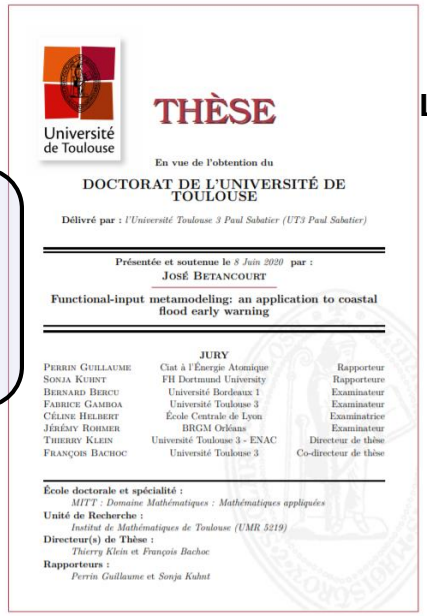


Difficulté (3) : sortie spatiale



Avancées RISCOPE

- Prise en compte des entrées temporelles → **thèse de José Betancourt** (IMT)



Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *8 Juin 2020* par :

JOSÉ BETANCOURT

Functional-input metamodeling: an application to coastal flood early warning

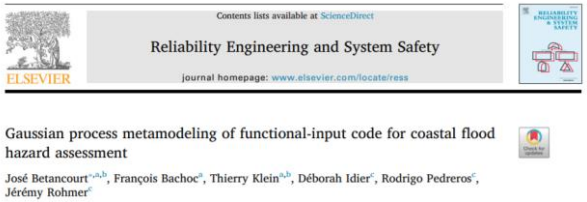
JURY

PERRIN GUILLAUME	Ciat à l'Énergie Atomique	Rapporteur
SONJA KUHNT	FH Dortmund University	Rapporteure
BERNARD BERCU	Université Bordeaux 1	Examinateur
FABRICE GAMBOA	Université Toulouse 3	Examinateur
CÉLINE HELBERT	École Centrale de Lyon	Examinatrice
JÉRÉMY ROHMER	BRGM Orléans	Examinateur
THIERRY KLEIN	Université Toulouse 3 - ENAC	Directeur de thèse
FRANÇOIS BACHOC	Université Toulouse 3	Co-directeur de thèse

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse (UMR 5219)
Directeur(s) de Thèse :
Thierry Klein et François Bachoc
Rapporteurs :
Perrin Guillaume et Sonja Kuhnt

Logiciel libre sous R:

https://djbetancourt-gh.github.io/funGp/



Contents lists available at ScienceDirect

Reliability Engineering and System Safety

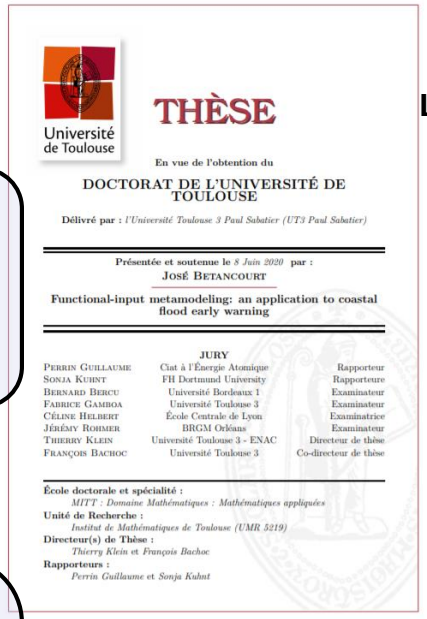
journal homepage: www.elsevier.com/locate/ress

Gaussian process metamodeling of functional-input code for coastal flood hazard assessment

José Betancourt[a,a,b], François Bachoc[c], Thierry Klein[a,b], Déborah Idier[c], Rodrigo Pedreros[c], Jérémy Rohmer[c]

Avancées RISCOPE

- Prise en compte des entrées temporelles → **thèse de José Betancourt** (IMT)



Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *8 Juin 2020* par :

JOSÉ BETANCOURT

Functional-input metamodeling: an application to coastal flood early warning

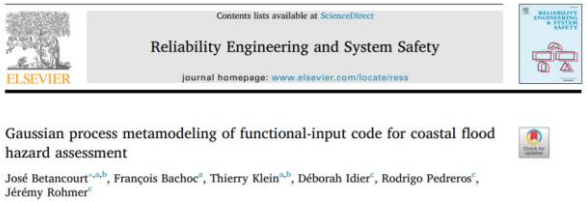
JURY

PERRIN GUILLAUME	Cisat à l'Énergie Atomique	Rapporteur
SONJA KUHNT	FH Dortmund University	Rapporteure
BERNARD BERCU	Université Bordeaux 1	Examinateur
FABRICE GAMBOA	Université Toulouse 3	Examinateur
CÉLINE HELBERT	École Centrale de Lyon	Examinatrice
JÉRÉMY ROHMER	BRGM Orléans	Examinateur
THIERRY KLEIN	Université Toulouse 3 - ENAC	Directeur de thèse
FRANÇOIS BACHOC	Université Toulouse 3	Co-directeur de thèse

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse (UMR 5219)
Directeur(s) de Thèse :
Thierry Klein et François Bachoc
Rapporteurs :
Perrin Guillaume et Sonja Kuhnt

Logiciel libre sous R:

https://djbetancourt-gh.github.io/funGp/



Contents lists available at ScienceDirect

Reliability Engineering and System Safety

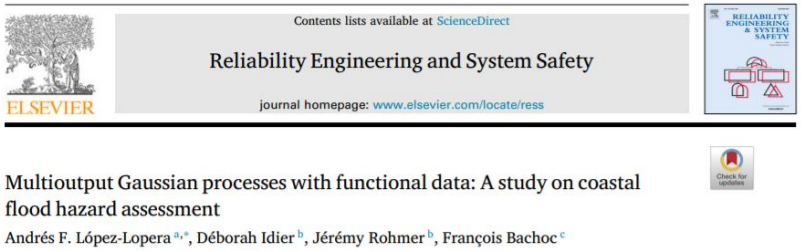
journal homepage: www.elsevier.com/locate/ress

Gaussian process metamodeling of functional-input code for coastal flood hazard assessment

José Betancourt[a,b], François Bachoc[c], Thierry Klein[a,b], Déborah Idier[c], Rodrigo Pedreros[c], Jérémy Rohmer[c]

- Prise en compte à la fois d'entrées temporelles et des sorties cartes → **post-doc d'Andrés F. López-Lopera** (BRGM/IMT)

Logiciel libre sous R: https://github.com/anfelopera/spatfGPs



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]

Avancées RISCOPE

- Prise en compte des entrées temporelles → **thèse de José Betancourt** (IMT)

- Prise en compte à la fois d'entrées temporelles et des sorties cartes → **post-doc d'Andrés F. López-Lopera** (BRGM/IMT)

- Prise en compte du **caractère aléatoire** des vagues

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *8 Juin 2020* par :

JOSÉ BETANCOURT

Functional-input metamodeling: an application to coastal flood early warning

JURY

PERRIN GUILLAUME	Cisat à l'Énergie Atomique	Rapporteur
SONJA KUHNT	FH Dortmund University	Rapporteure
BERNARD BERCU	Université Bordeaux 1	Examinateur
FABRICE GAMBOA	Université Toulouse 3	Examinateur
CÉLINE HELBERT	École Centrale de Lyon	Examinatrice
JÉRÉMY ROHMER	BRGM Orléans	Examinateur
THIERRY KLEIN	Université Toulouse 3 - ENAC	Directeur de thèse
FRANÇOIS BACHOC	Université Toulouse 3	Co-directeur de thèse

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse (UMR 5219)
Directeur(s) de Thèse :
Thierry Klein et François Bachoc
Rapporteurs :
Perrin Guillaume et Sonja Kuhnt

Logiciel libre sous R:

https://djbetancourt-gh.github.io/funGp/

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Gaussian process metamodeling of functional-input code for coastal flood hazard assessment

José Betancourt[a,b], François Bachoc[c], Thierry Klein[a,b], Déborah Idier[c], Rodrigo Pedreros[c], Jérémy Rohmer[c]

Logiciel libre sous R: https://github.com/anfelopera/spatfGPs

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]

Stochastic Environmental Research and Risk Assessment (2020) 34:867–890
https://doi.org/10.1007/s00477-020-01803-2

ORIGINAL PAPER

A nuanced quantile random forest approach for fast prediction of a stochastic marine flooding simulator applied to a macrotidal coastal site

Journal of *Marine Science and Engineering*

MDPI

Article

The Effect of Stochasticity of Waves on Coastal Flood and Its Variations with Sea-level Rise

Déborah Idier *, Rodrigo Pedreros, Jérémy Rohmer and Gonéri Le Cozannet

Métamodèlisation avec entrées temporelles

Collaborateurs

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *8 Juin 2020* par :

JOSÉ BETANCOURT

Functional-input metamodeling: an application to coastal flood early warning

INSTITUT de MATHEMATIQUES de TOULOUSE

Encadrants

- Thierry Klein
- François Bachoc

Directeur d'équipe

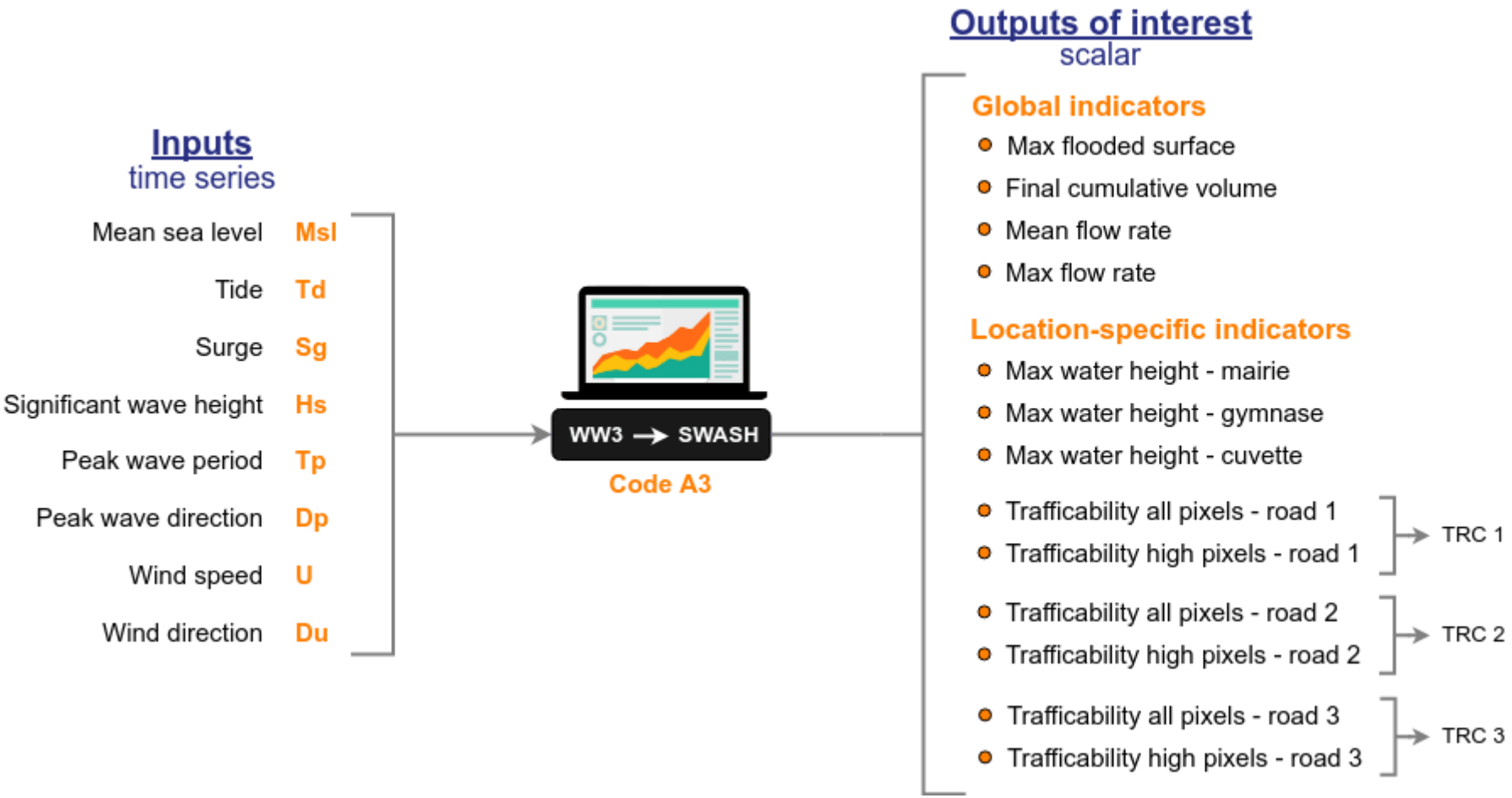
- Fabrice Gamboa

Grands contributeurs

Geoscience for a sustainable Earth brgm

- Déborah Idier
- Jérémy Rohmer
- Rodrigo Pedreros

Contexte de recherche



Contexte de recherche

Entrées fonctionnelles

- Quelles entrées il convient garder actives?
- Quel méthode de réduction de dimension est idéal?
- Quelle est une dimension de projection adéquat?

Métamodèle par processus Gaussiens

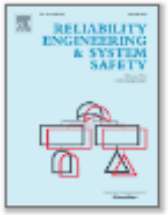
- Quelle fonction de covariance?
- Quelle distance pour quantifier la similarité entre entrées fonctionnelles?

⟶ On appelle ces attribues les **paramètres structurels** du métamodèle

Expérience initiale de métamodèlisation

ELSEVIER

Reliability Engineering & System Safety
Volume 198, June 2020, 106870



Gaussian process metamodeling of functional-input code for coastal flood hazard assessment

José Betancourt [a, b], François Bachoc [a], Thierry Klein [a, b], Déborah Idier [c], Rodrigo Pedreros [c], Jérémy Rohmer [c]

- Première jeu de métamodèles avec entrées fonctionnelles
- Preuve de concept: optimisation structurelle axé sur qualité de prédiction
- Développent des routines R préliminaires

Expérience initiale de métamodèlisation

- Nous divisons chaque entrée fonctionnelle en une représentation scalaire et une fonctionnelle, x et f, respectivement.

- Nous observons un code déterministe de la forme:

$$(\boldsymbol{x}, \boldsymbol{f}) \mapsto f_{code}(\boldsymbol{x}, \boldsymbol{f}),$$

  - $\boldsymbol{x}$ le vecteur des d_s entrées scalaires ($x^{(i)} \in \mathbb{R}, \ \forall\, i = 1, \dots, d_s$)
  - $\boldsymbol{f}$ le vecteur des d_f entrées fonctionnelles ($f^{(i)}: \ T_i \subset \mathbb{R} \to \mathbb{R}, \ \forall\, i = 1, \dots, d_f$)

- Nous voyons f_{code} comme une réalisation d'un processus Gaussien Y:

$$Y \sim \mathcal{GP}(m, k),$$

  - m la fonction moyenne
  - k la fonction de covariance

$$k\left(Y(\boldsymbol{x}, \boldsymbol{f}), (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{f}})\right) = k_s(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \ k_f(\boldsymbol{f}, \tilde{\boldsymbol{f}})$$

Muehlenstaedt et al. (2017)

Expérience initiale de métamodèlisation

Attributs à definir:

- État de chaque entrée scalaire
  - active
  - inactive
- Méthode de projection
  - B-splines Muehlenstaedt et al. (2017)
  - PCA Lataniotis et al. (2018)
- Distance pour les fonctions
  - By_index Nantyet al. (2016)
  - By_group Muehlenstaedt et al. (2017)
- État de c/entrée fonctionnelle
  - active
  - inactive
- Dimension de projection
  - $\{1, \ldots 37\}$
- Type de kernel Rasmussen et al. (2006)
  - Gaussienne
  - Mattérn $5/2$
  - Mattérn $3/2$

Expérience initiale de métamodèlisation

Attributs à definir:

- État de chaque entrée scalaire
  - active
  - inactive
- Méthode de projection
  - B-splines Muehlenstaedt et al. (2017)
  - PCA Lataniotis et al. (2018)
- Distance pour les fonctions
  - By_index Nantyet al. (2016)
  - By_group Muehlenstaedt et al. (2017)
- État de c/entrée fonctionnelle
  - active
  - inactive
- Dimension de projection
  - $\{1, \ldots 37\}$
- Type de kernel Rasmussen et al. (2006)
  - Gaussienne
  - Mattérn $5/2$
  - Mattérn $3/2$

Expérience initiale de métamodèlisation

Attributs à definir:

- État de chaque entrée scalaire
  - active
  - inactive
- Méthode de projection
  - B-splines Muehlenstaedt et al. (2017)
  - PCA Lataniotis et al. (2018)
- Distance pour les fonctions
  - By_index Nantyet al. (2016)
  - By_group Muehlenstaedt et al. (2017)
- État de c/entrée fonctionnelle
  - active
  - inactive
- Dimension de projection
  - $\{1, \ldots 37\}$
- Type de kernel Rasmussen et al. (2006)
  - Gaussienne
  - Mattérn $5/2$
  - Mattérn $3/2$

Expérience initiale de métamodèlisation

Attributs à definir:

- État de chaque entrée scalaire
  - active
  - inactive
- Méthode de projection
  - B-splines Muehlenstaedt et al. (2017)
  - PCA Lataniotis et al. (2018)
- Distance pour les fonctions
  - By_index Nantyet al. (2016)
  - By_group Muehlenstaedt et al. (2017)
- État de c/entrée fonctionnelle
  - active
  - inactive
- Dimension de projection
  - $\{1, \ldots 37\}$
- Type de kernel Rasmussen et al. (2006)
  - Gaussienne
  - Mattérn 5/2
  - Mattérn 3/2

Expérience initiale de métamodèlisation

Attributs à definir:

- État de chaque entrée scalaire
  - active
  - inactive
- Méthode de projection
  - B-splines Muehlenstaedt et al. (2017)
  - PCA Lataniotis et al. (2018)
- Distance pour les fonctions
  - By_index Nantyet al. (2016)
  - By_group Muehlenstaedt et al. (2017)
- État de c/entrée fonctionnelle
  - active
  - inactive
- Dimension de projection
  - $\{1, \dots 37\}$
- Type de kernel Rasmussen et al. (2006)
  - Gaussienne
  - Mattérn 5/2
  - Mattérn 3/2

Expérience initiale de métamodèlisation

Attributs à definir:

- État de chaque entrée scalaire
  - active
  - inactive
- Méthode de projection
  - B-splines Muehlenstaedt et al. (2017)
  - PCA Lataniotis et al. (2018)
- Distance pour les fonctions
  - By_index Nanty et al. (2016)
  - By_group Muehlenstaedt et al. (2017)
- État de c/entrée fonctionnelle
  - active
  - inactive
- Dimension de projection
  - $\{1, \dots 37\}$
- Type de kernel Rasmussen et al. (2006)
  - Gaussienne
  - Mattérn 5/2
  - Mattérn 3/2

Expérience initiale de métamodèlisation

Attributs à definir:

- État de chaque entrée scalaire
  - active
  - inactive
- Méthode de projection
  - B-splines Muehlenstaedt et al. (2017)
  - PCA Lataniotis et al. (2018)
- Distance pour les fonctions
  - By_index Nantyet al. (2016)
  - By_group Muehlenstaedt et al. (2017)
- État de c/entrée fonctionnelle
  - active
  - inactive
- Dimension de projection
  - $\{1, \dots 37\}$
- Type de kernel Rasmussen et al. (2006)
  - Gaussienne
  - Mattérn $5/2$
  - Mattérn $3/2$

Expérience initiale de métamodèlisation

Attributs à definir:

- État de chaque entrée scalaire
  - active
  - inactive
- Méthode de projection
  - B-splines Muehlenstaedt et al. (2017)
  - PCA Lataniotis et al. (2018)
- Distance pour les fonctions
  - By_index Nantyet al. (2016)
  - By_group Muehlenstaedt et al. (2017)
- État de c/entrée fonctionnelle
  - active
  - inactive
- Dimension de projection
  - $\{1, \dots 37\}$
- Type de kernel Rasmussen et al. (2006)
  - Gaussienne
  - Mattérn $5/2$
  - Mattérn $3/2$

Quelle est la combinaison optimale?

Expérience initiale de métamodèlisation

Prototype basé sur la **Méthode de Surface de Réponse**

Montgomery (1991)

- Utilité
  - Optimiser fonction inconnu
  - **N**ombre raisonnable d'observations

- RISCOPE
  - Fonction a optimiser: $Q^2(\boldsymbol{w})$
    - Q^2: corrélation de Pearson (qualité de prédiction du métamodèle)
    - $\boldsymbol{w}$: vecteur des paramètres structurels

Expérience initiale de métamodèlisation

Prototype basé sur la **Méthode de Surface de Réponse**

Montgomery (2017)

- Résultats
  - Configuration optimale dans majorité des scenarios
  - Gap d'optimalité négligeable dans cas restants ($\sim 1e^{-2}\%$)
  - Amélioration du temps vs exploration exhaustive (+28%)

- Conclusions
  - Efficace pour optimisation structurelle
  - Configuration optimale difficile a choisir empiriquement
  - Besoin d'une méthode plus efficient
    - Inclure un majeur nombre d'entrées
    - Permettre majeur diversité des paramètres structurels
    - Eviter stagnation dans des optima locaux

Optimisation structurelle axée sur Fourmis

Ant Colony Based Model Selection for Functional-Input Gaussian Process Regression

Technical Report - April 2020

Deliverable: D3.b
Reference: WP3.2

INSTITUT de MATHEMATIQUES de TOULOUSE

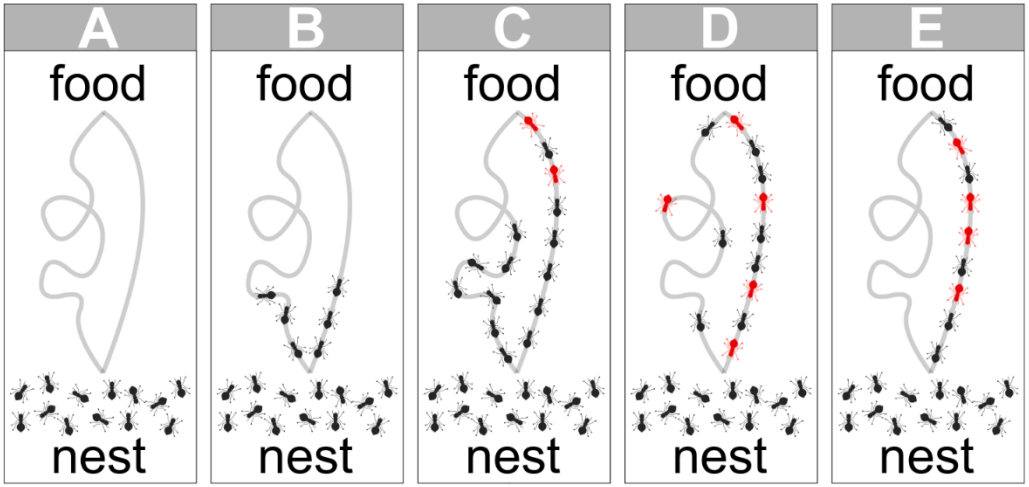
José Betancourt, François Bachoc, Thierry Klein, Fabrice Gamboa

- Exploration plus efficient de l'espace des paramètres structurels

Optimisation structurelle axée sur Fourmis

ACO en bref

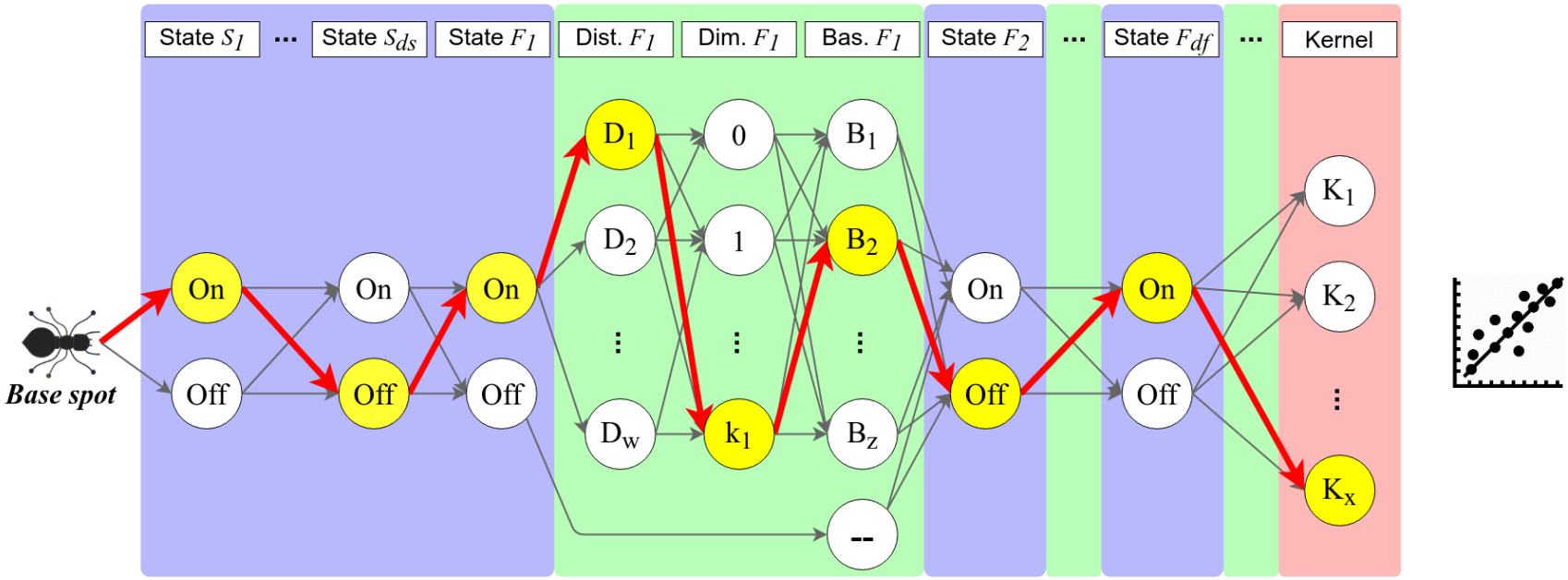
- Méthode heuristique d'optimisation combinatoire Dorigo (1992)
- Basée sur le comportement des fourmis réelles Goss et al. (1989)



- Exploration stratégique pseudo-aléatoirisée
  - Couverture efficiente de la région faisable
  - Convergence relativement rapide vers l'optimum

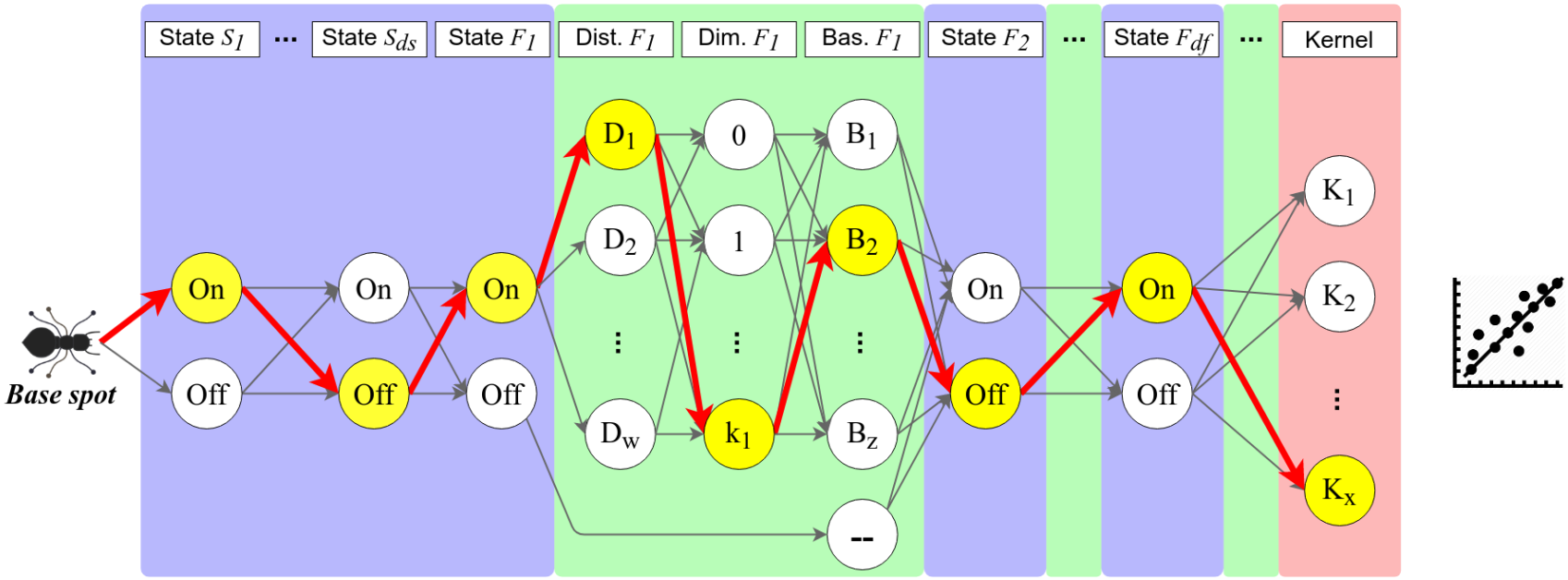
Optimisation structurelle axée sur Fourmis

ACO en RISCOPE



Optimisation structurelle axée sur Fourmis

ACO en RISCOPE



modification **locale** des phéromones →

← modification **globale** des phéromones

Résultats sur un cas analytique

$$\mathcal{G}_2 : \quad [0,1]^2 \times \mathcal{F}^2 \to \mathbb{R},$$

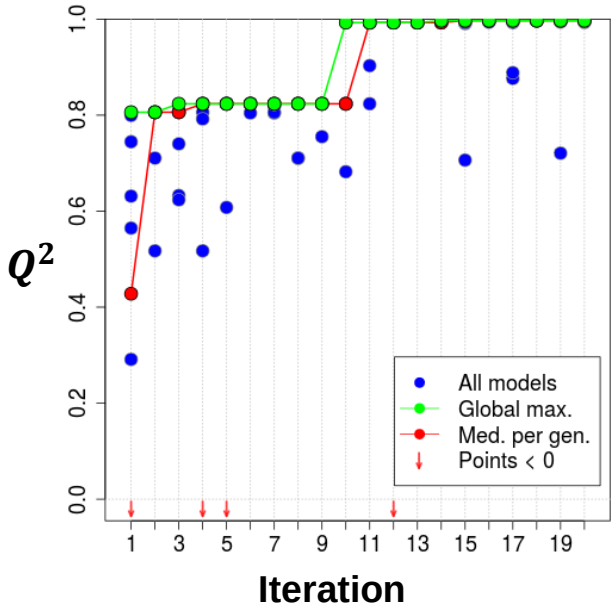
$$(\boldsymbol{x}, \boldsymbol{f}) \mapsto x_1 \sin(x_2) + \int_0^1 \exp(x_1 t_1) f_1 dt_1 - x_2^2 \int_0^1 f_2 t_2 dt_1.$$

Optimisation structurelle axée sur Fourmis

Résultats sur un cas analytique

$$\mathcal{G}_2 : \quad [0,1]^2 \times \mathcal{F}^2 \to \mathbb{R},$$

$$(\boldsymbol{x}, \boldsymbol{f}) \mapsto x_1 \sin(x_2) + \int_0^1 \exp(x_1 t_1) f_1 dt_1 - x_2^2 \int_0^1 f_2 t_2 dt_1.$$

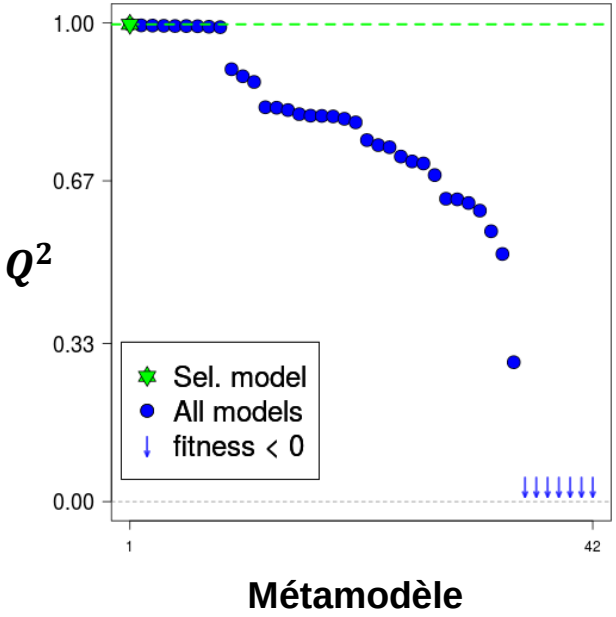


Optimisation structurelle axée sur Fourmis

Résultats sur un cas analytique

$$\mathcal{G}_2 : \quad [0,1]^2 \times \mathcal{F}^2 \to \mathbb{R},$$

$$(\boldsymbol{x}, \boldsymbol{f}) \mapsto x_1 \sin(x_2) + \int_0^1 \exp(x_1 t_1) f_1 dt_1 - x_2^2 \int_0^1 f_2 t_2 dt_1.$$

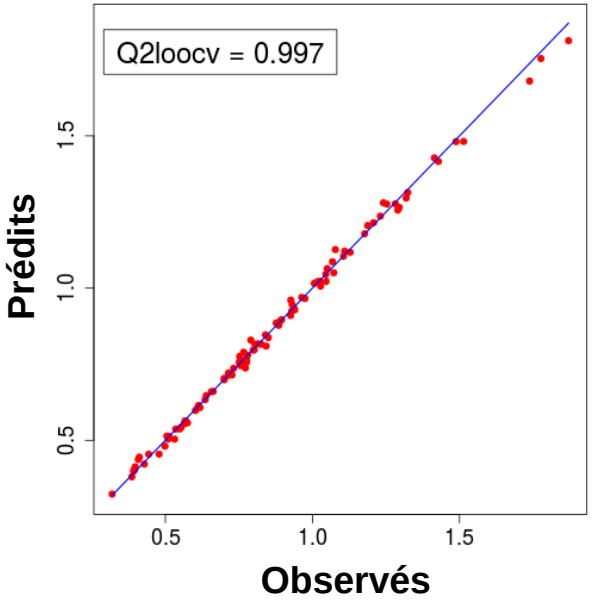


Optimisation structurelle axée sur Fourmis

Résultats sur un cas analytique

$$\mathcal{G}_2 : \quad [0,1]^2 \times \mathcal{F}^2 \to \mathbb{R},$$

$$(\boldsymbol{x}, \boldsymbol{f}) \mapsto x_1 \sin(x_2) + \int_0^1 \exp(x_1 t_1) f_1 dt_1 - x_2^2 \int_0^1 f_2 t_2 dt_1.$$



Optimisation structurelle axée sur Fourmis

Résultats sur RISCOPE

Inputs
time series

- Mean sea level — Msl
- Tide — Td
- Surge — Sg
- Significant wave height — Hs
- Peak wave period — Tp
- Peak wave direction — Dp
- Wind speed — U
- Wind direction — Du

Code A3

Outputs of interest
scalar

Global indicators

- Max flooded surface
- Final cumulative volume
- Mean flow rate
- Max flow rate

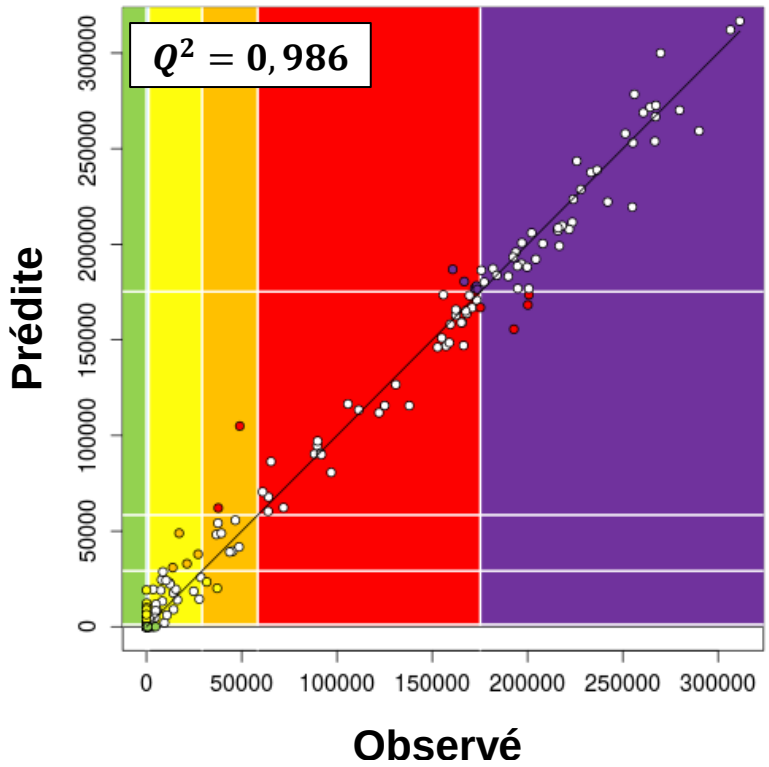
Location-specific indicators

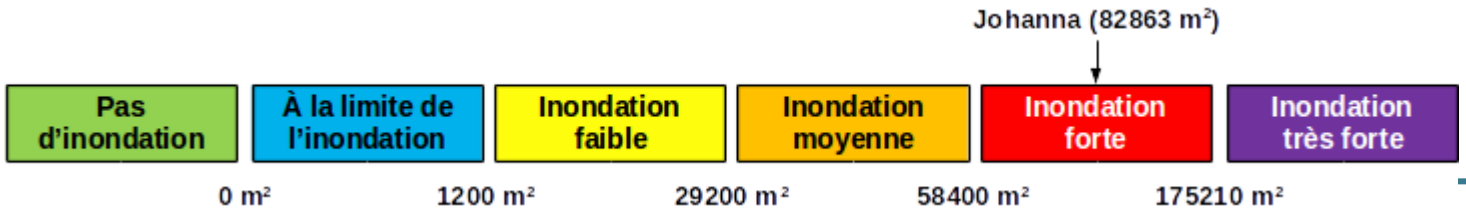
- Max water height - mairie
- Max water height - gymnase
- Max water height - cuvette
- Trafficability all pixels - road 1
- Trafficability high pixels - road 1 → TRC 1
- Trafficability all pixels - road 2
- Trafficability high pixels - road 2 → TRC 2
- Trafficability all pixels - road 3
- Trafficability high pixels - road 3 → TRC 3

Optimisation structurelle axée sur Fourmis

Résultats sur RISCOPE

Surface maximale inondé (m^2)

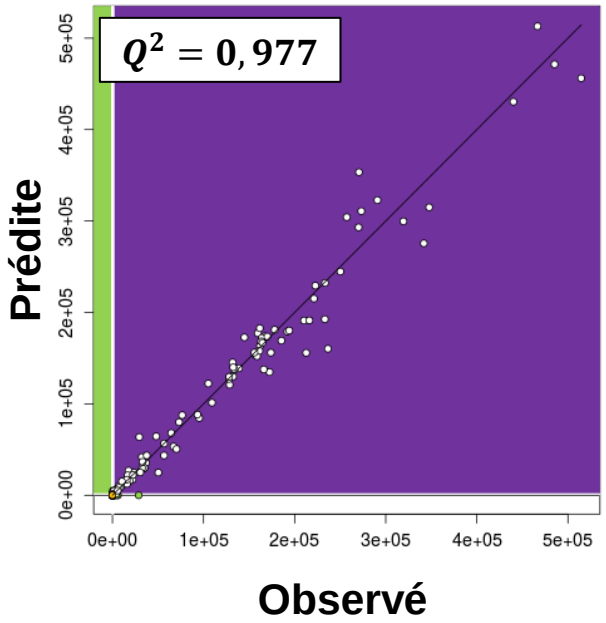




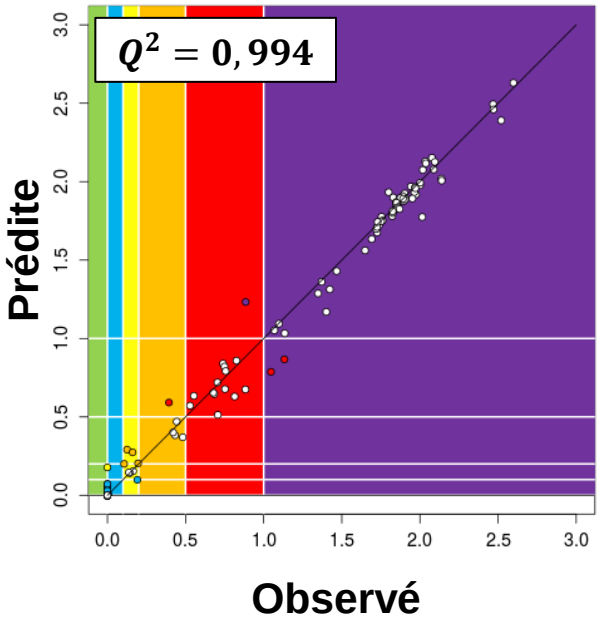
Optimisation structurelle axée sur Fourmis

Résultats sur RISCOPE

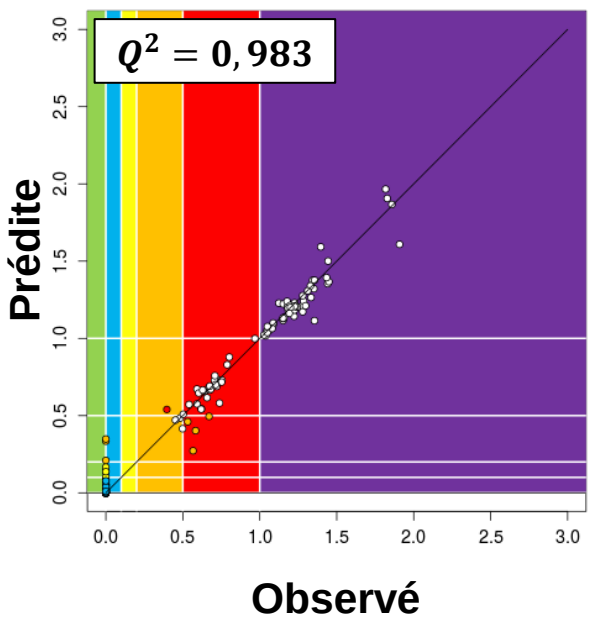
Débit maximale (m^3/h)



Hauteur max_cuv (m)



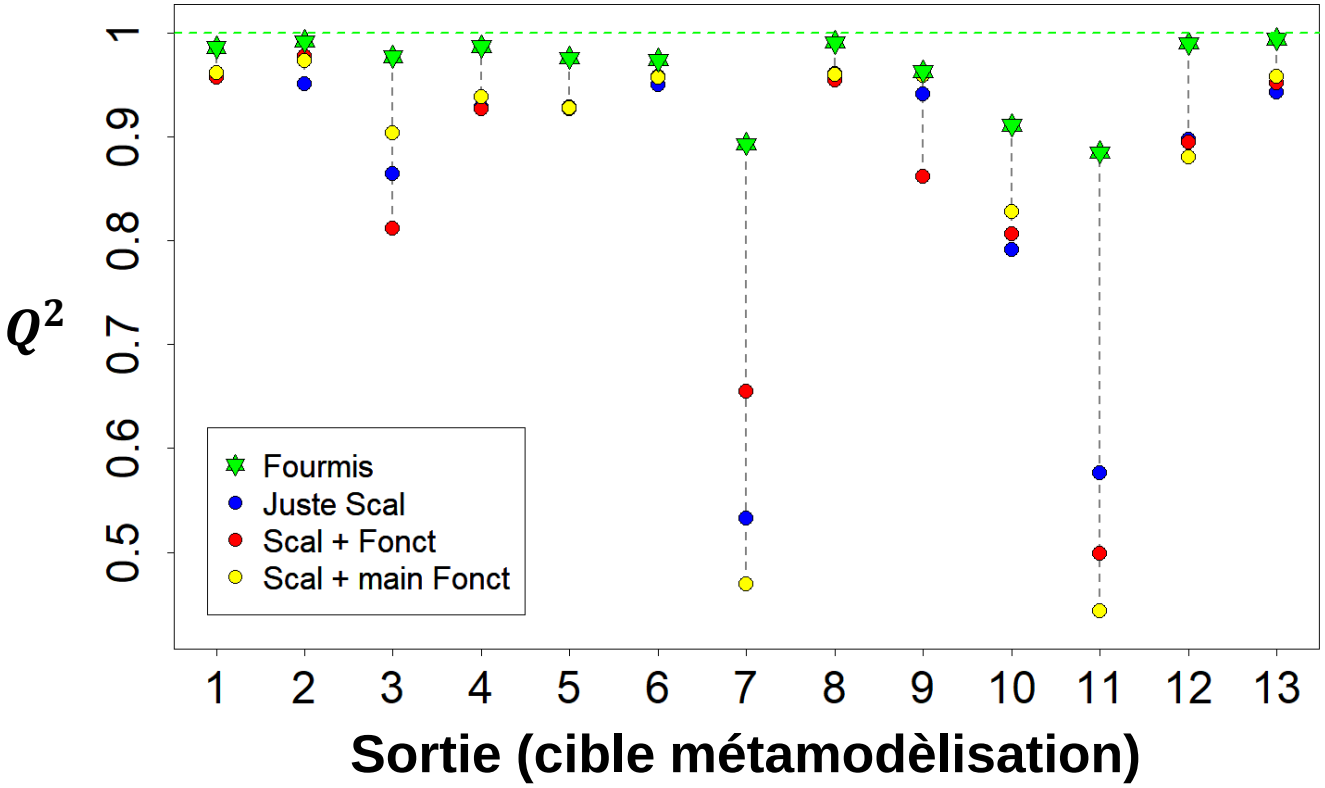
Hauteur max_gym (m)



Optimisation structurelle axée sur Fourmis

Résultats sur RISCOPE

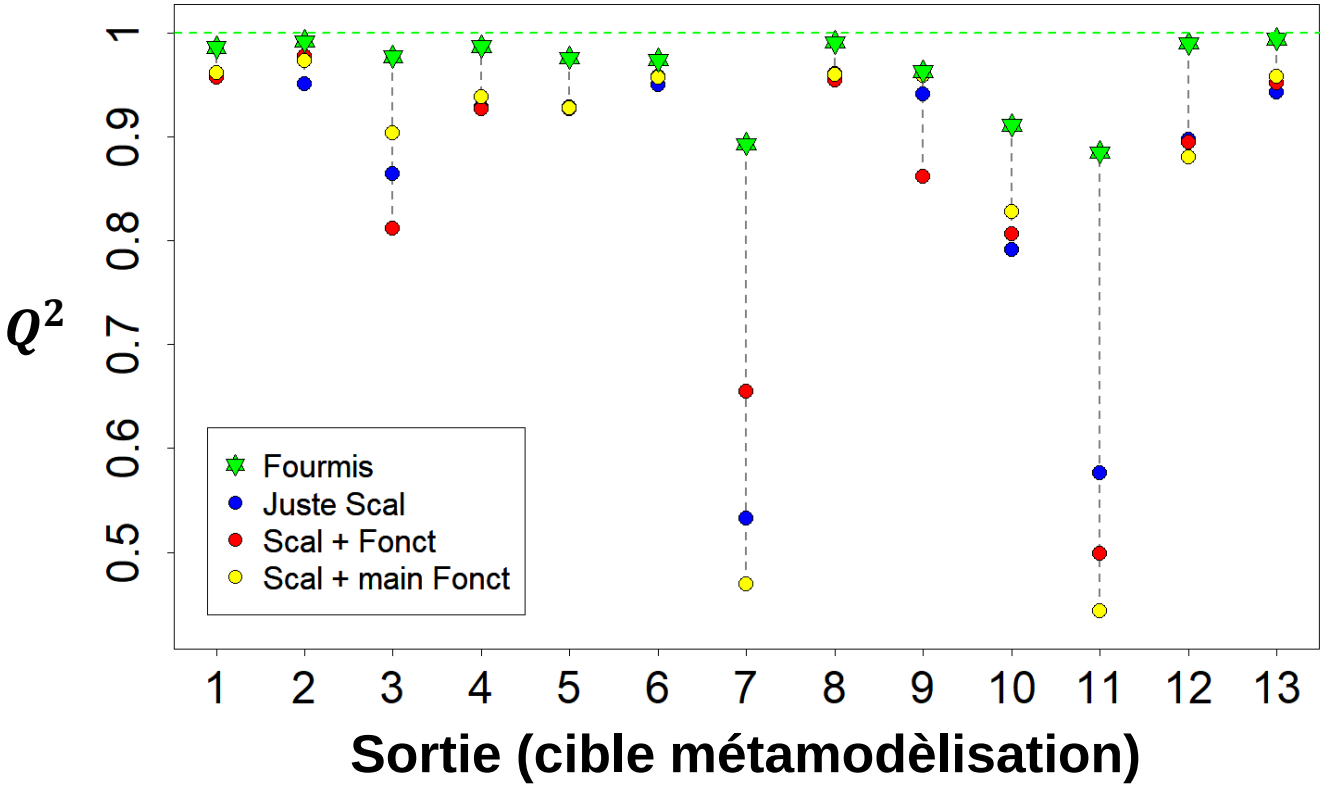
Comparaison **Fourmis vs** configuration **empirique**



Optimisation structurelle axée sur Fourmis

Résultats sur RISCOPE

Comparaison **Fourmis vs** configuration **empirique**



- Fourmis donne meilleure prédiction pour les 13 sorties

Optimisation structurelle axée sur Fourmis

Conclusions sur RISCOPE

- Tous les métamodèles fournis montrent une **haute capacité prédictive**
- Les **gammes** sont en générale **bien prédites**, surtout en valeurs hauts

Conclusions sur la méthode de Colonie des Fourmis

- **Efficient** – métamodèles performants dans un temps court
- **Robuste** – efficacité montré dans +16 scenarios inclus 13 RISCOPE
- **Scalable** – double des entrées considérés avec Surface de R.
- **Pertinent** – configurations obtenus pas forcement intuitives
- **Code stable** – motivation pour faire le développement accessible

Package R funGp

funGp

Gaussian Process Regression for Scalar and Functional Inputs with funGp

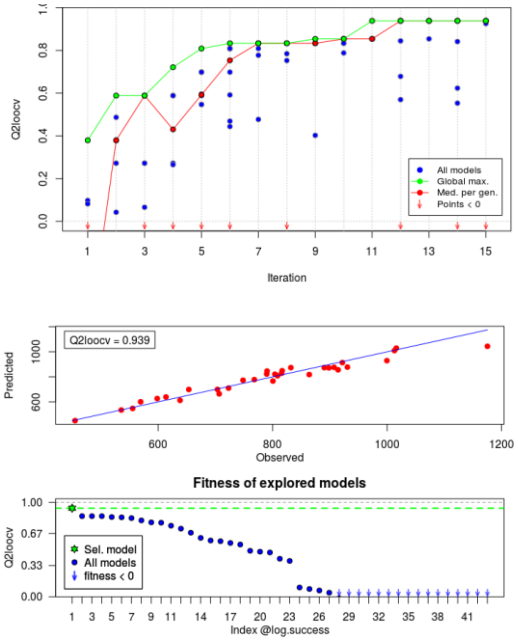
- José Betancourt
- François Bachoc
- Thierry Klein
- Jérémy Rohmer
- Déborah Idier

Package R funGp

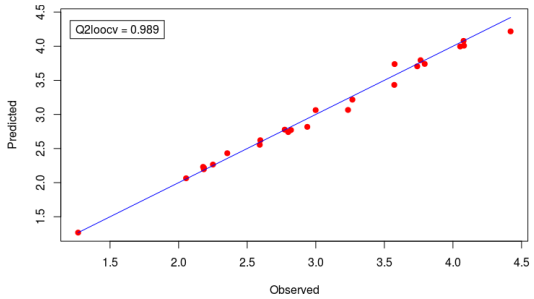
funGp en bref

- Régression avec modèle non-paramétrique
- Traitement des entrées scalaires et fonctionnelles
- Reduction de dimension intégré
- Optimisation structurel efficiente
- Performance renforce par parallélisation
- Implémentation modulaire pour facile extension
- Manuel intégral disponible sur **HAL**
- Code ouvert, disponible sur **CRAN** et **GitHub**

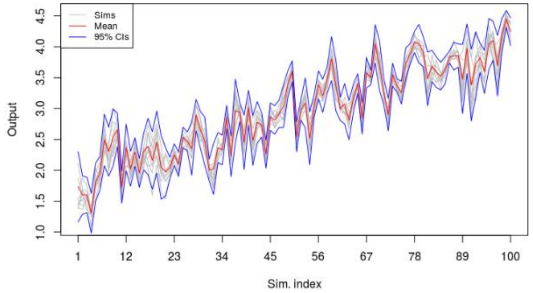
Optimisation structurelle



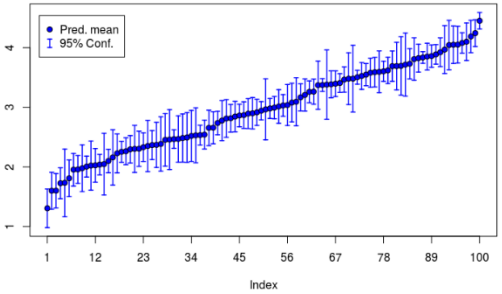
Construction et validation



Simulation

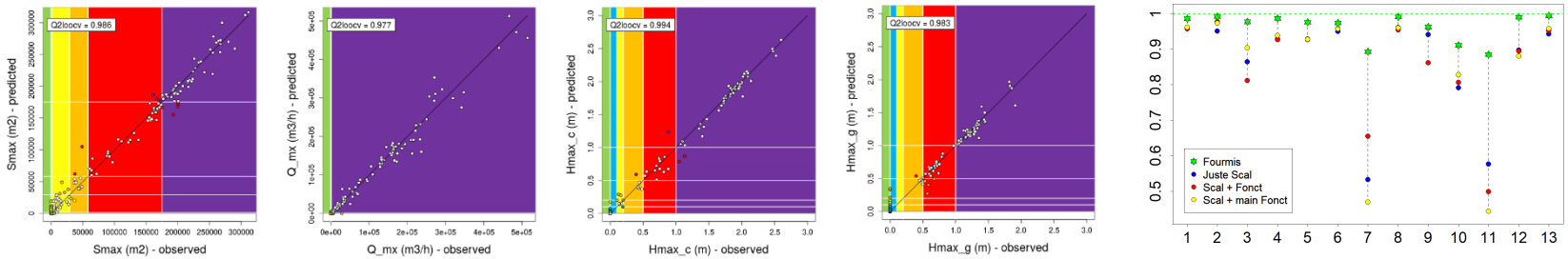


Prédiction

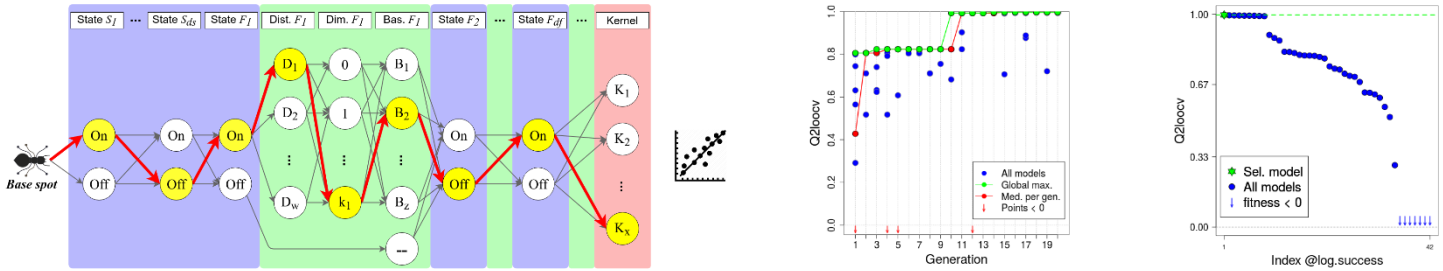


Synthèse

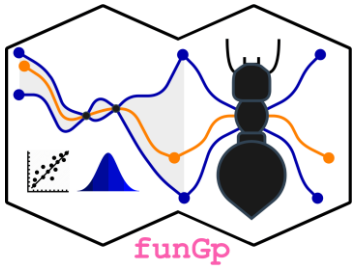
- **Métamodèles pour système d'aide a la décision appliqué en alerte précoce d'inondation**

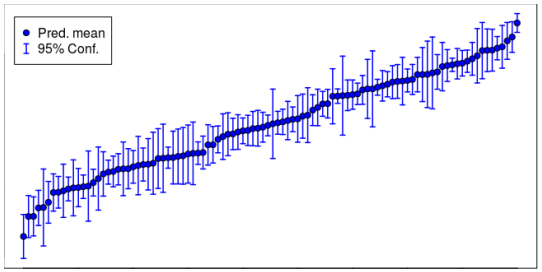


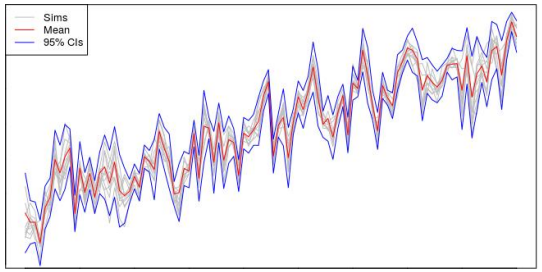
- **Méthode innovateur pour l'optimisation des paramètres structurelles**



- **Package R intégrant des techniques de métamodèlisation utilisées en RISCOPE**







Métamodèlisation avec entrées temporelles et sorties cartes

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

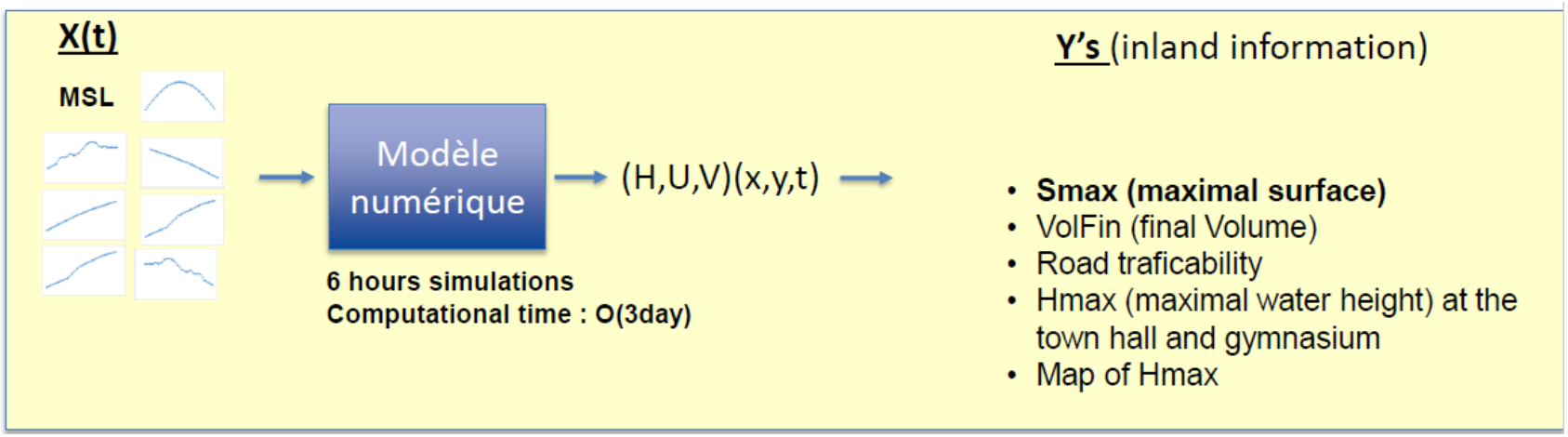
Keywords:
Surrogate model
Spatial flood data
Functional analysis
Dimensionality reduction
Separable kernel
Sparse-variational approximation

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.

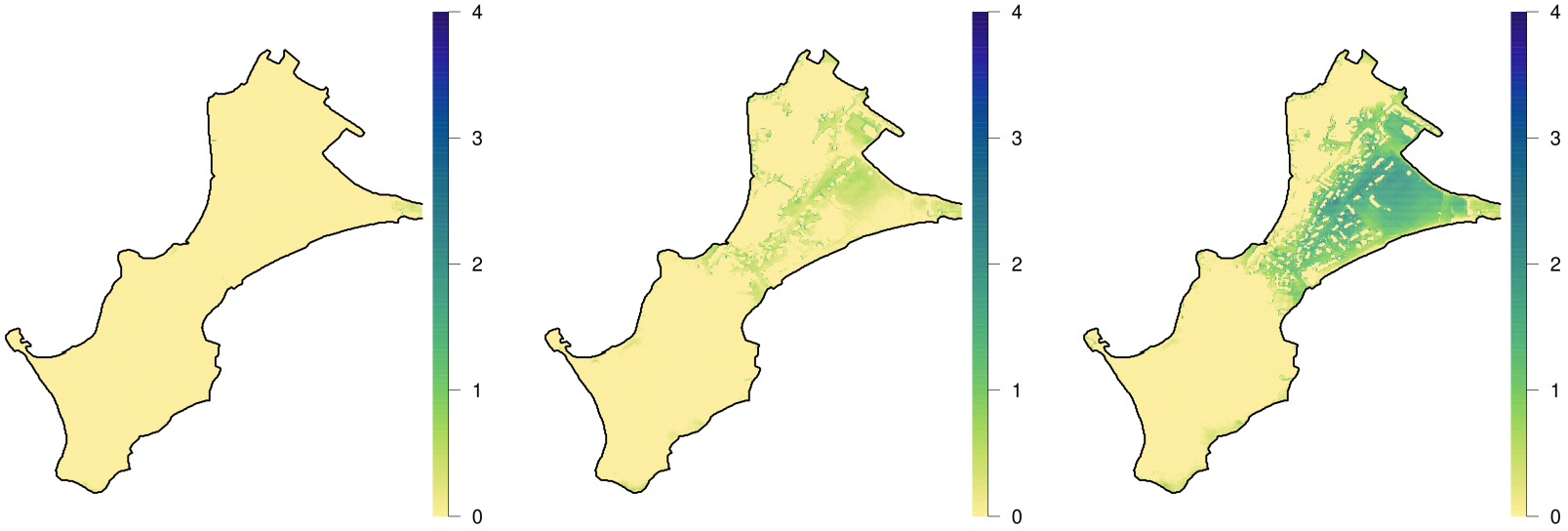
Géosciences pour une Terre durable

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])

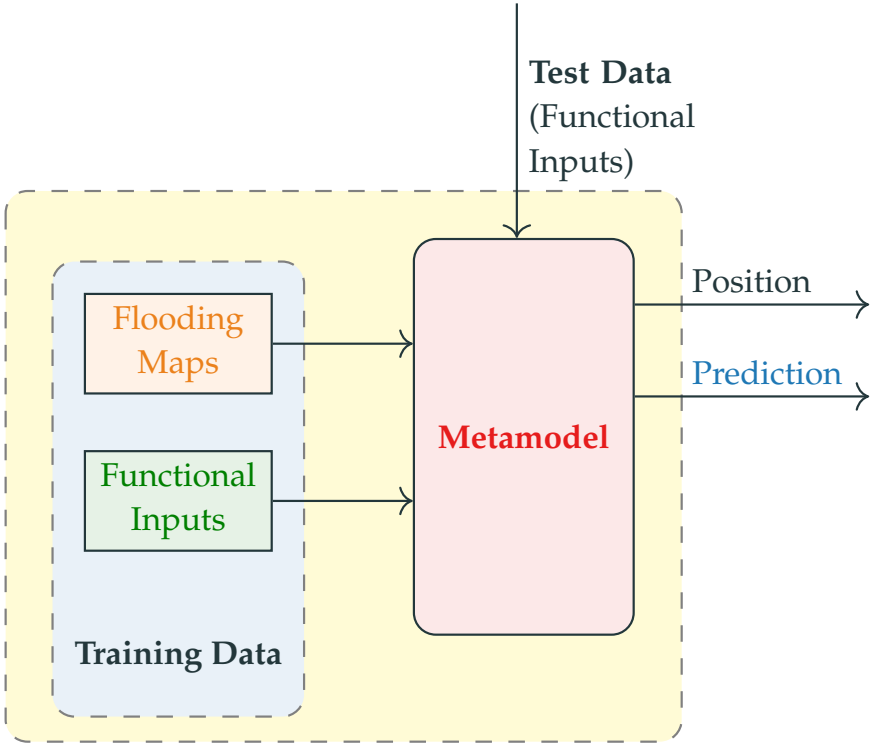


Challenge:

· 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.

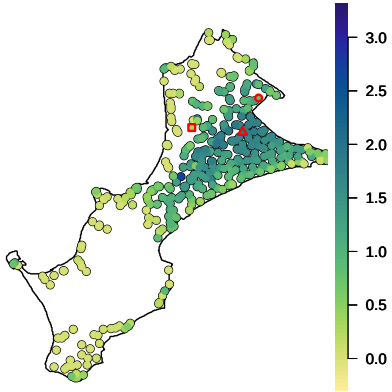
Application: forecasting and early warning systems.

Research subject

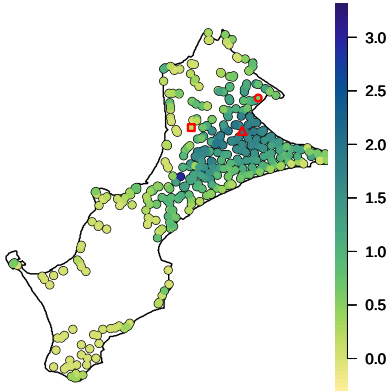
- To do so...
  - We further investigate a metamodel based on **Gaussian process**.
  - This results in a **multi-output process**:
    - correlated spatial flood events are driven by (functional) hydro-meteorological inputs.
  - Efficient implementations are based on **Python and R codes**.
    - Codes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

Research subject

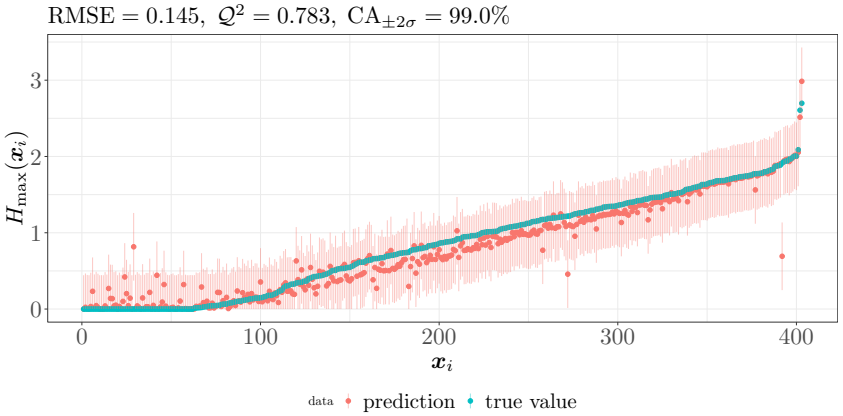
- To do so...
  - We further investigate a metamodel based on **Gaussian process**.
  - This results in a **multi-output process**:
    - correlated spatial flood events are driven by (functional) hydro-meteorological inputs.
  - Efficient implementations are based on **Python and R codes**.
    - Codes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

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

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

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

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:

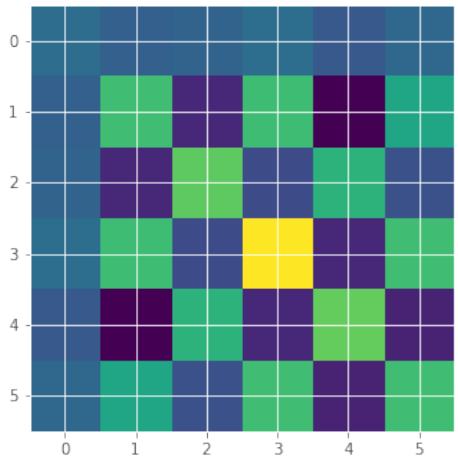
$$Z_i(\boldsymbol{x}) := Y(\boldsymbol{\mathcal{F}}_i, \boldsymbol{x}), \quad \text{for } i = 1, \cdots, R.$$

· In that case, k can be rewritten as:

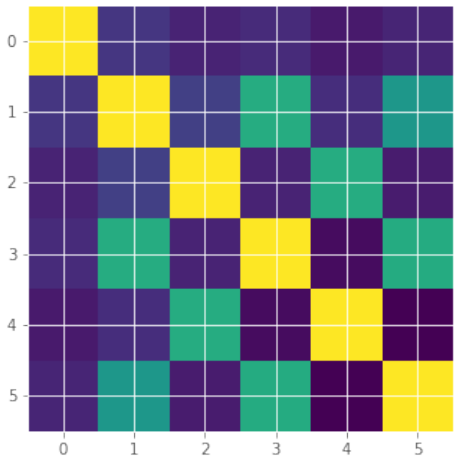
$$k_{i,j}(\boldsymbol{x}, \boldsymbol{x}') = b_{i,j}\, k_s(\boldsymbol{x}, \boldsymbol{x}'), \tag{3}$$

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

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



B

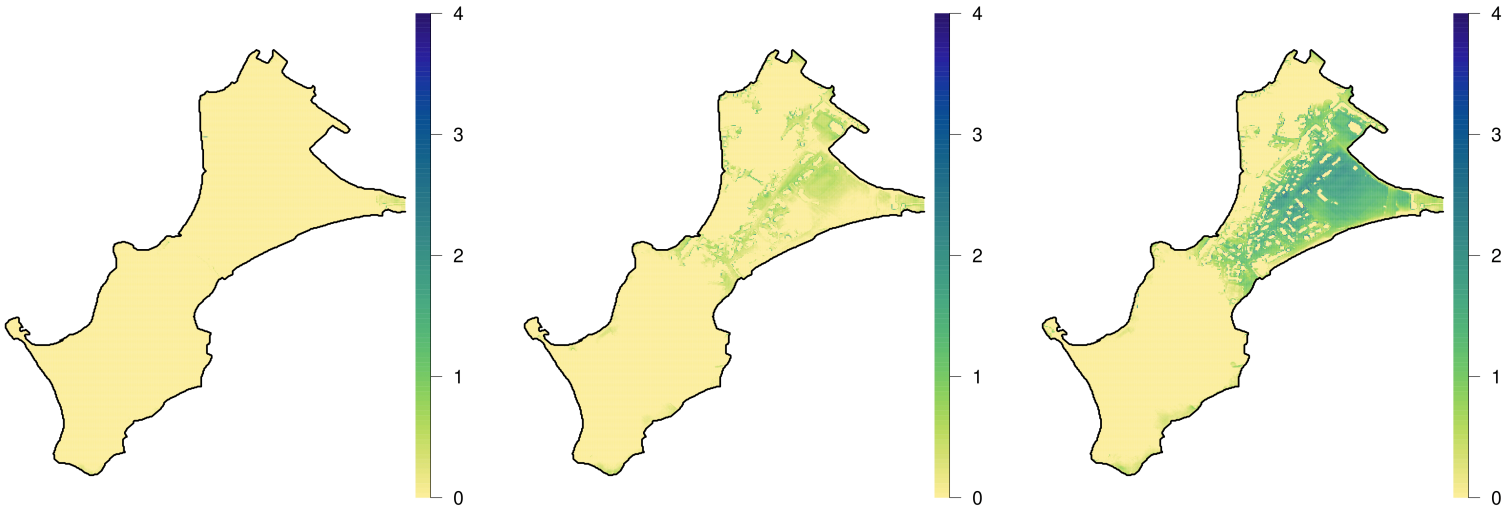


$k_f(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}')$

Coregionalisation matrix

Coastal flooding application

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

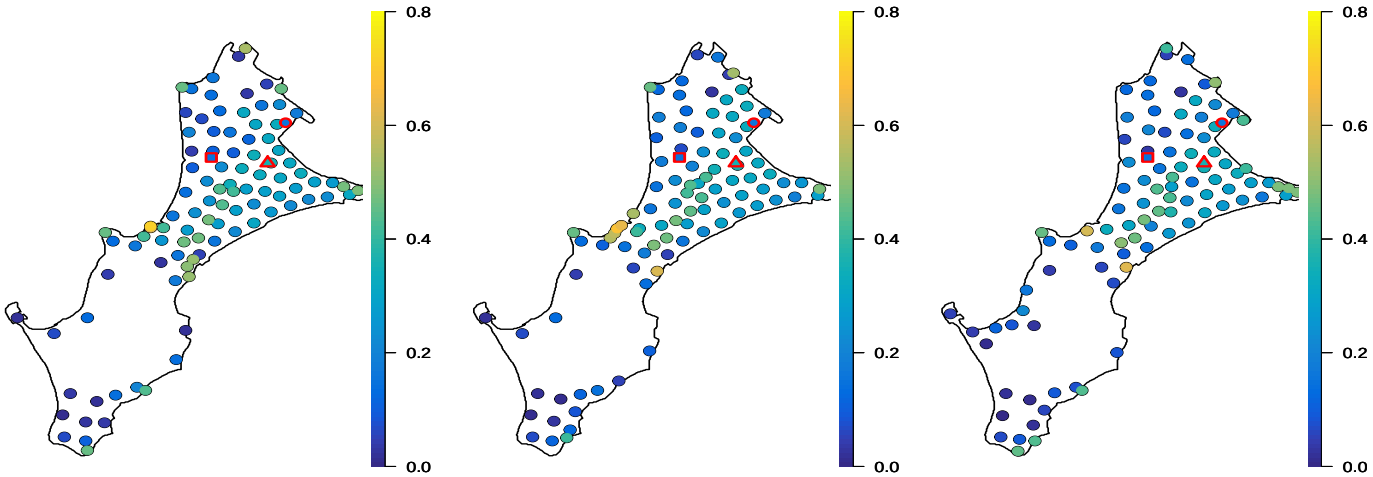


· **174 flood/no flood events** containing 65k inland observations.

- *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

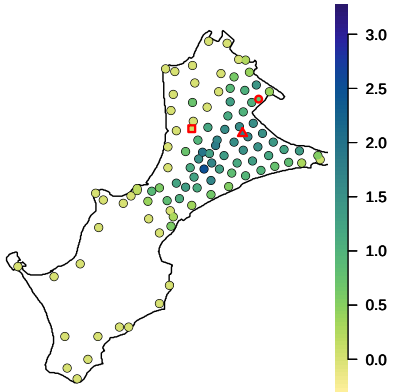
LOO test: each scenario is predicted using data of the other ones.

- 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.

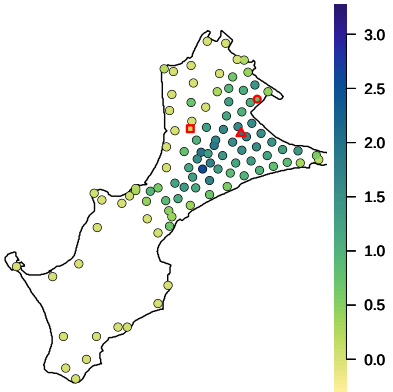
Leave-one-out (LOO) test

Flooding scenario 136: significant flood event

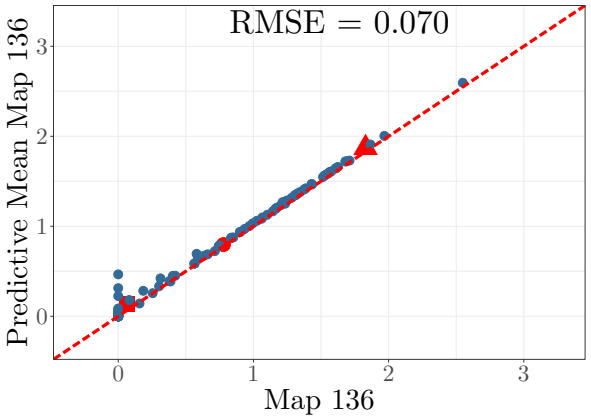
■ Town-Hall ● Gym ▲ Sports Field



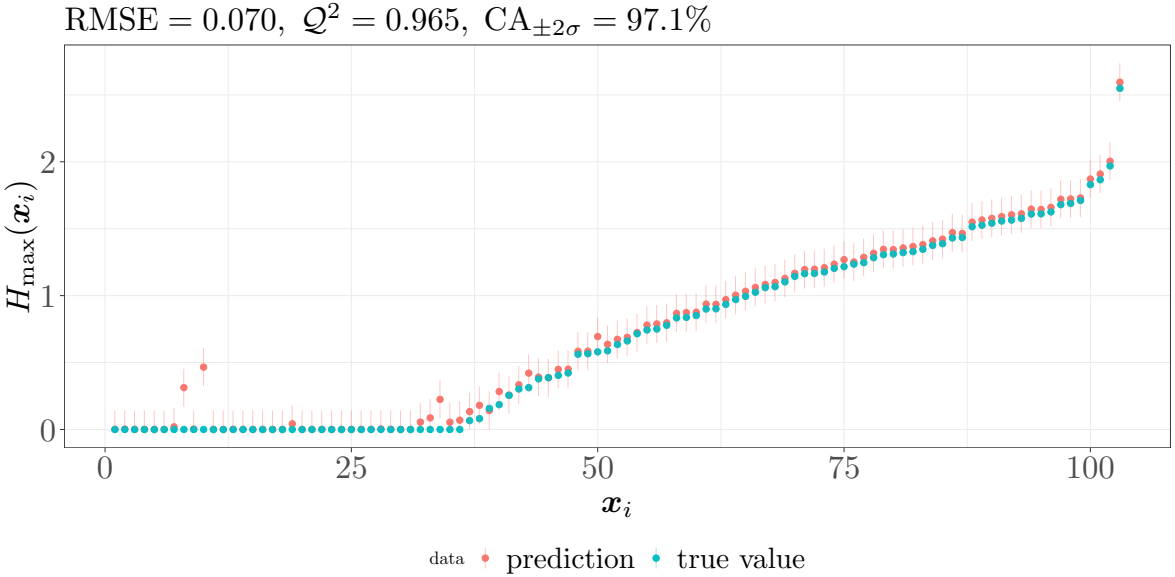
ground truth



prediction



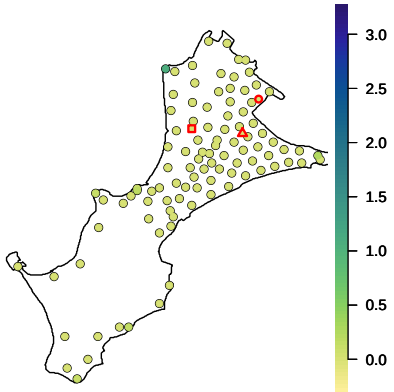
ground truth vs prediction



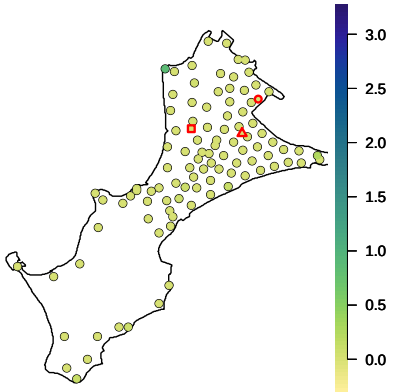
Leave-one-out (LOO) test

Flooding scenario 1: minor flood event

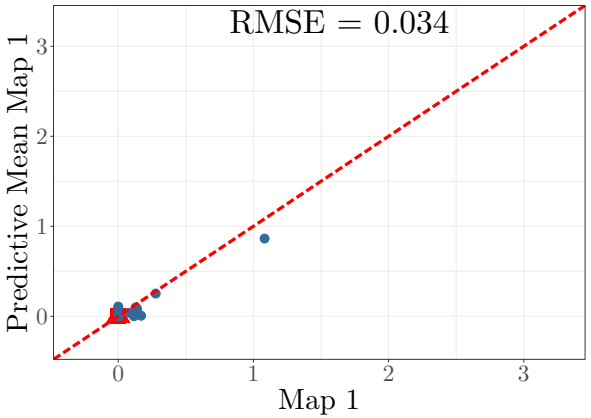
■ Town-Hall ● Gym ▲ Sports Field



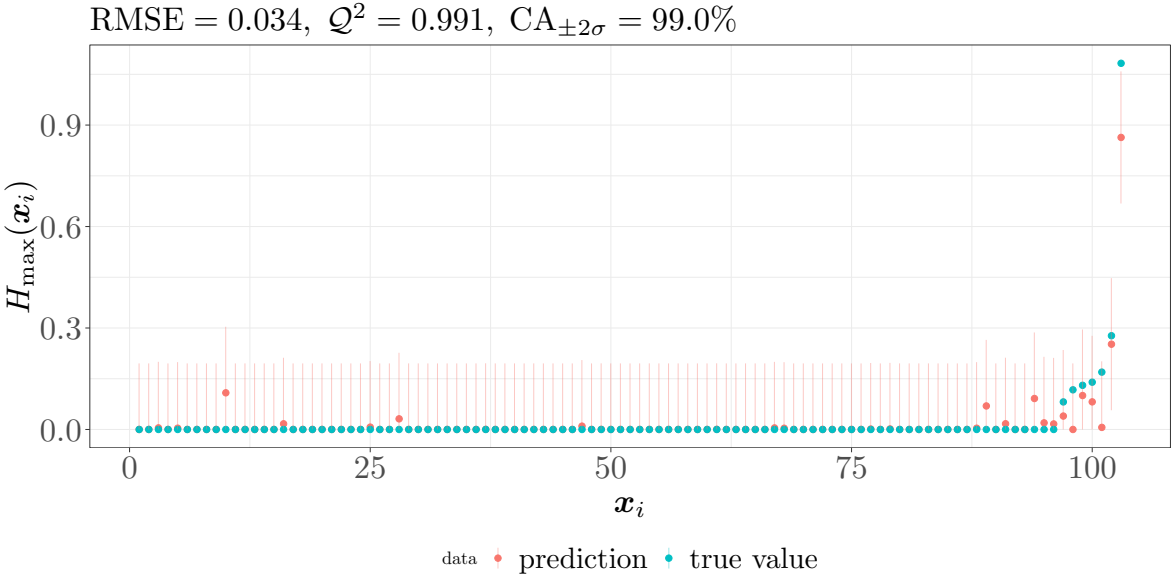
ground truth



prediction



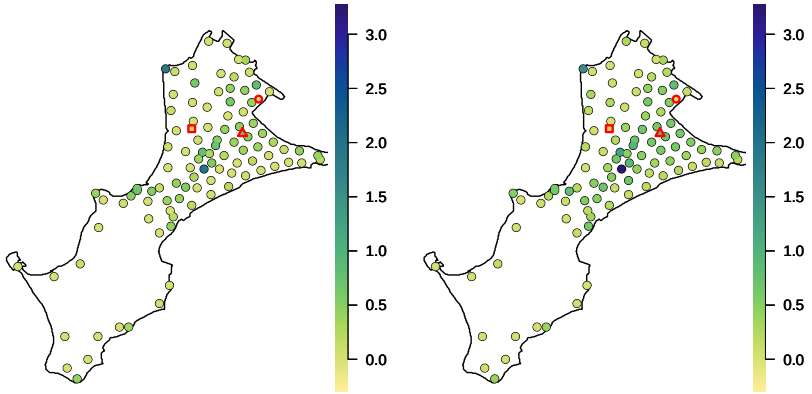
ground truth vs prediction



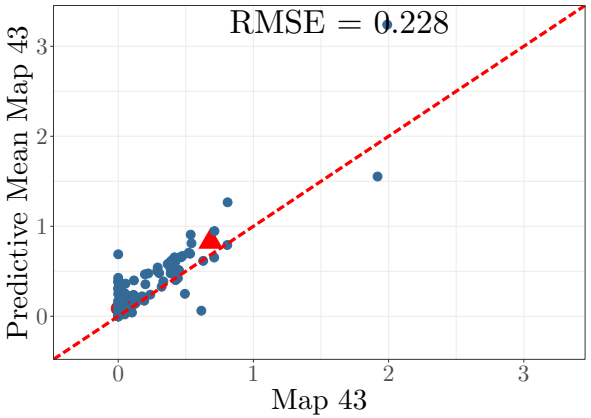
Leave-one-out (LOO) test

Flooding scenario 43: moderate flood event

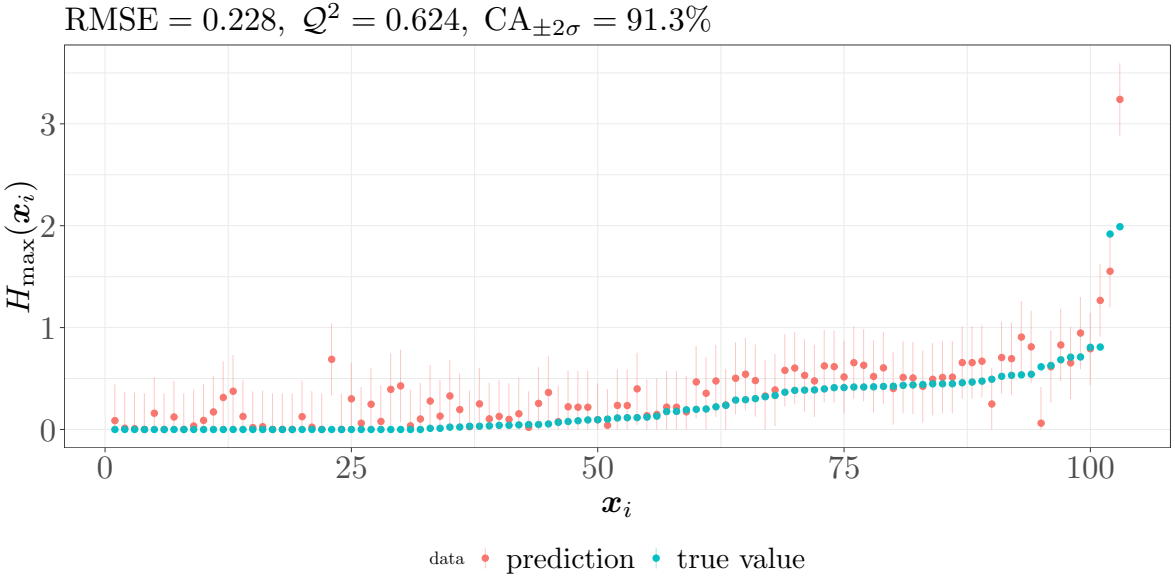
■ Town-Hall ● Gym ▲ Sports Field



ground truth prediction



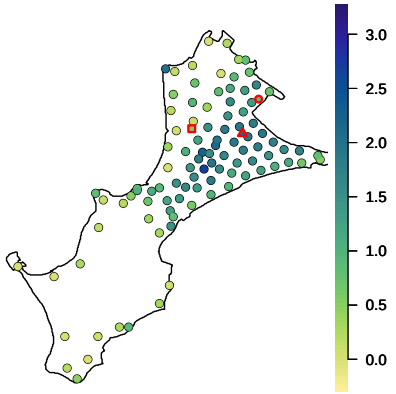
ground truth vs prediction



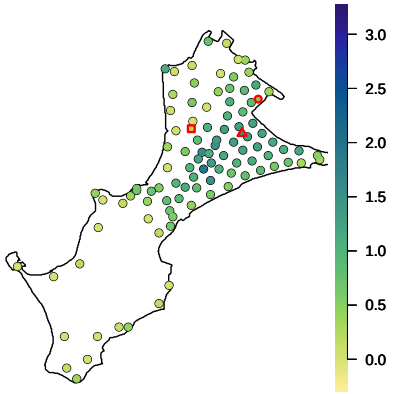
Leave-one-out (LOO) test

Flooding scenario 80: misprediction

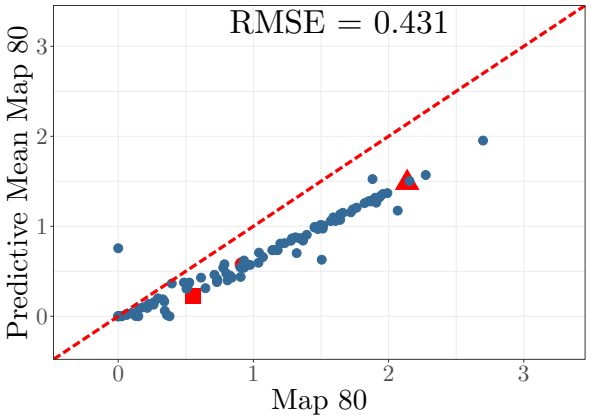
■ Town-Hall ● Gym ▲ Sports Field



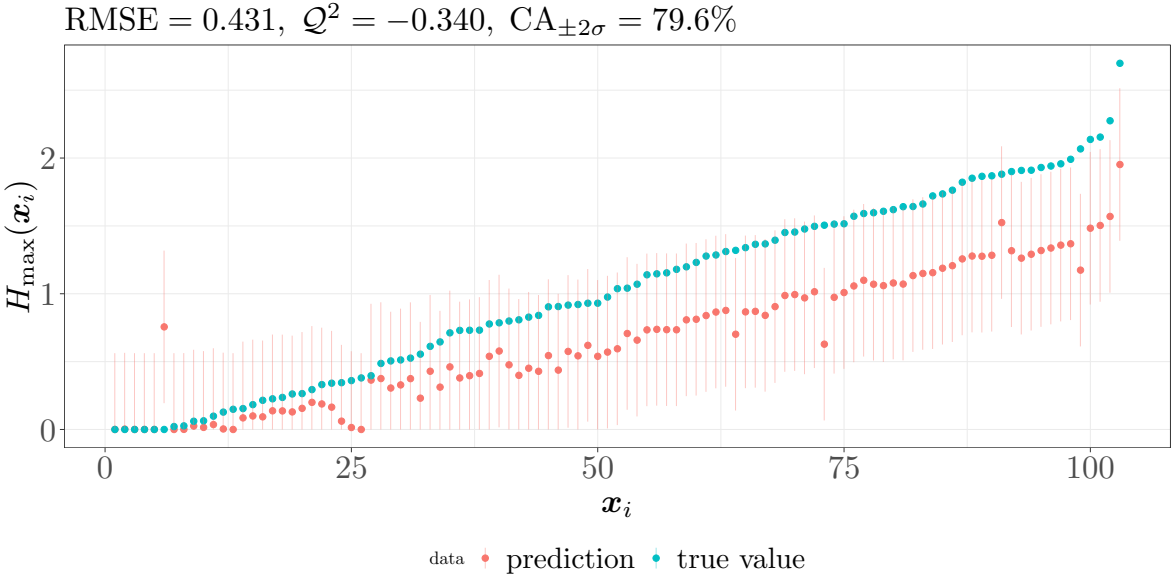
ground truth



prediction

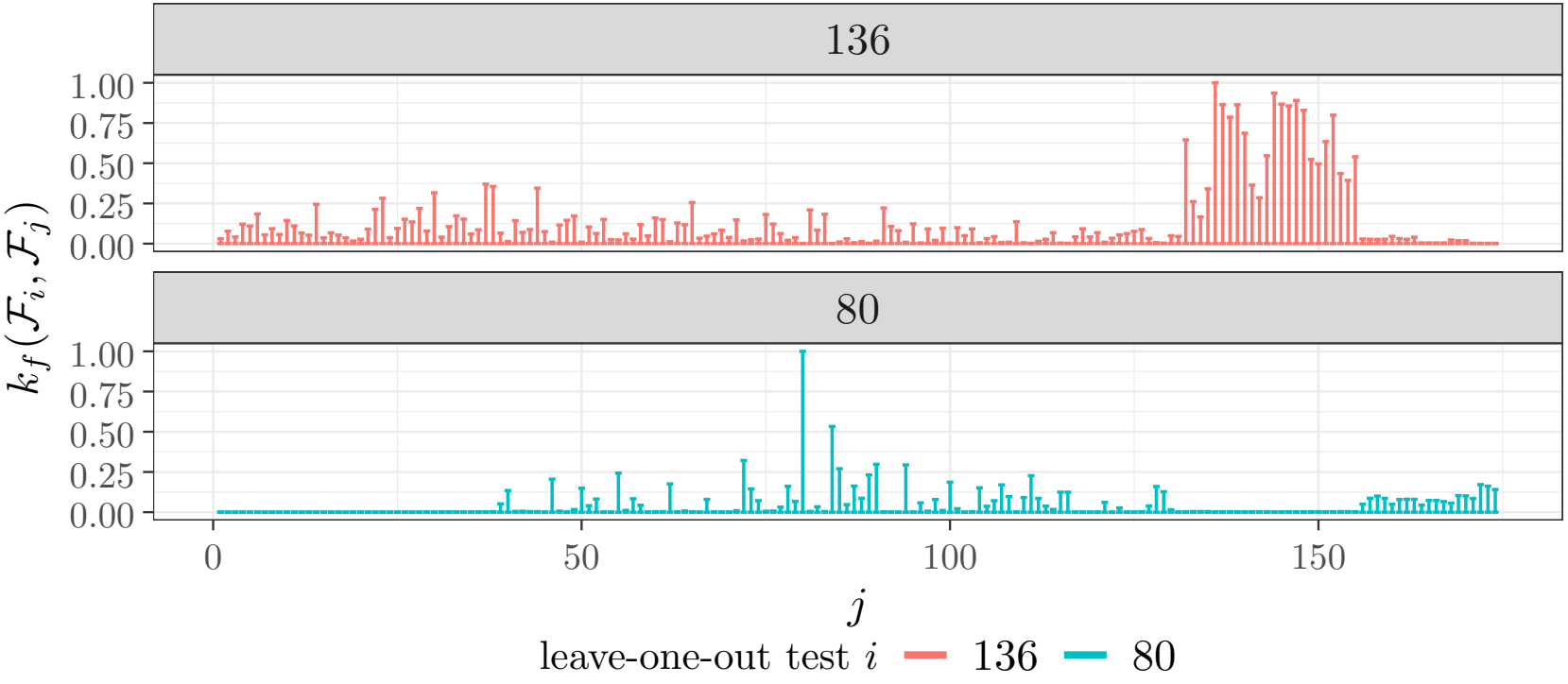


ground truth vs prediction



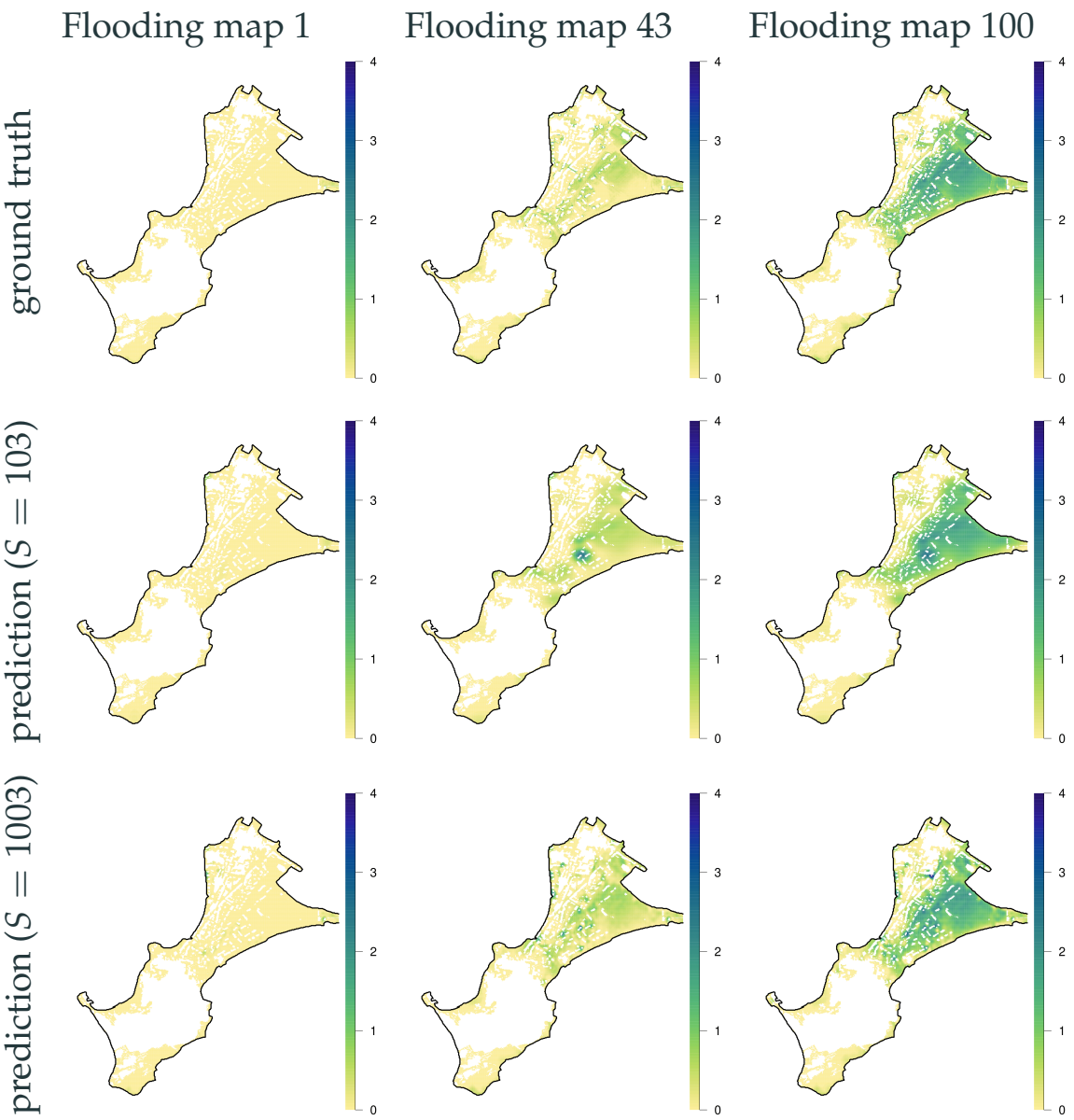
Leave-one-out (LOO) test

Flooding scenario 80: misprediction (discussion)

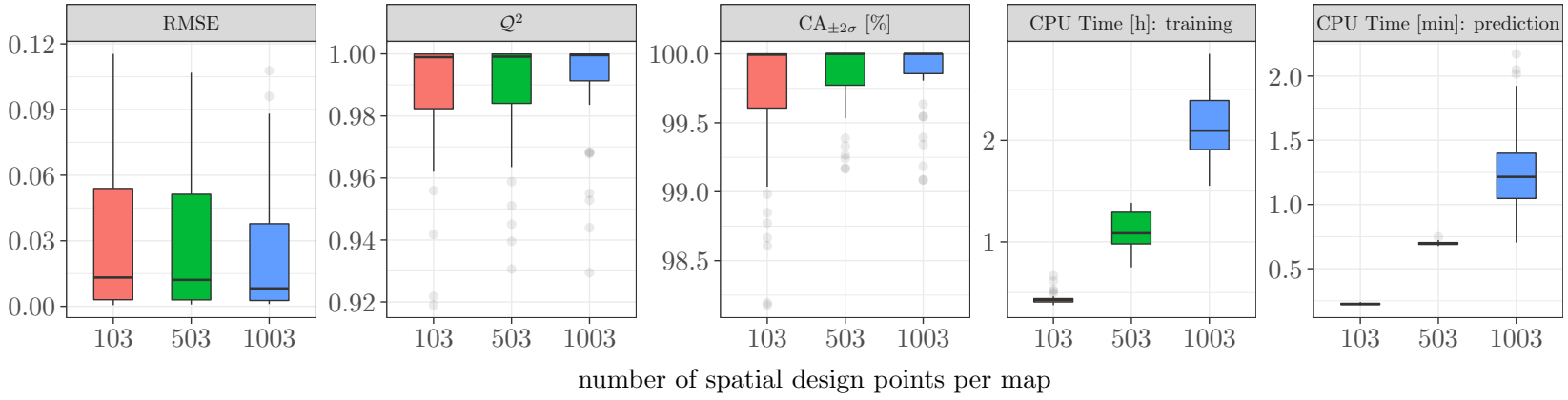


Correlations between functional inputs

Influence of the number of design points



Influence of the number of design points



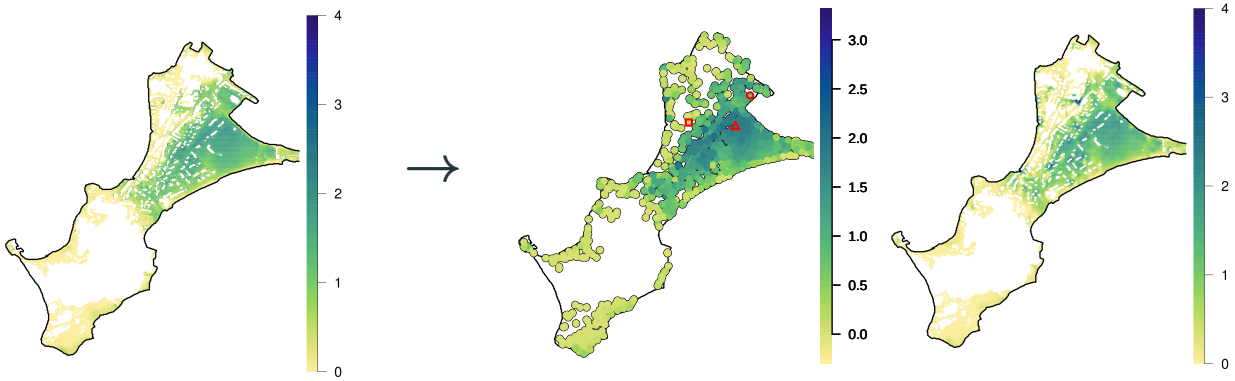
Performance indicators computed on a dataset based on **131 flood events**

Advantages and remaining challenges

Advantages

- Approximate predictions under **smoothness assumptions**.

Numerical simulator | GP metamodel



$\sim$ 3 days (∼ 65k points) | $\sim$ 2 seconds (∼ 1k points) | $\sim$ 1 min (∼ 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*.

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.

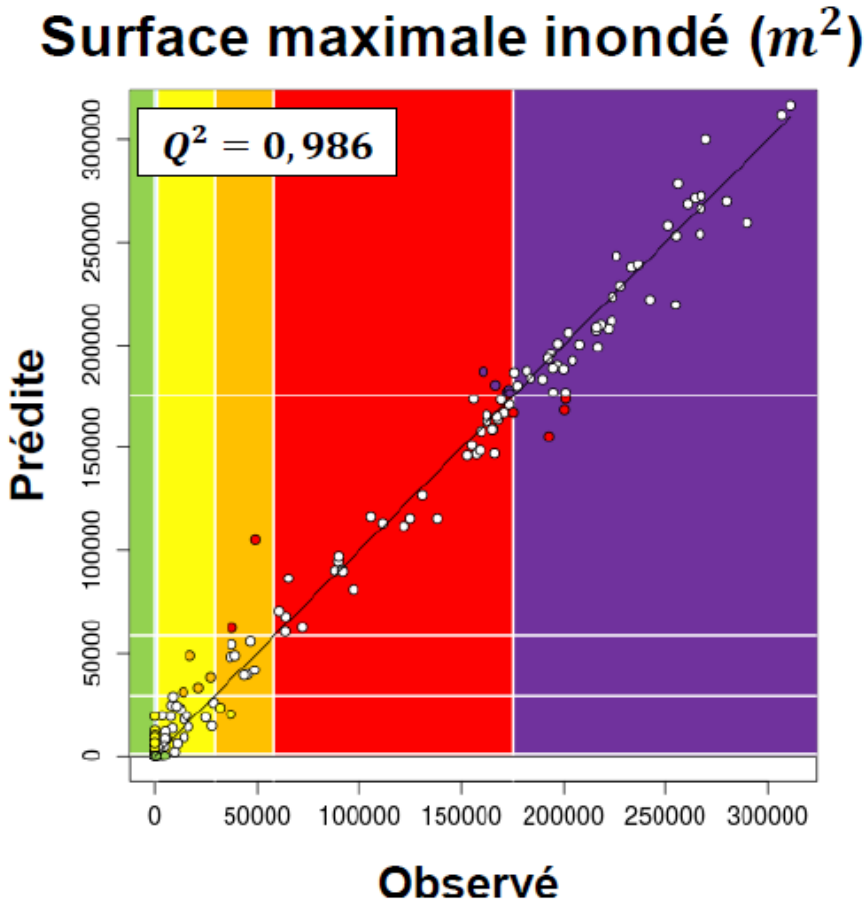
C E Rasmussen and C K I Williams. *GPs for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, Cambridge, MA, 2005.

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.

Au final…

Prédire avec des entrées fonctionnelles est possible!!

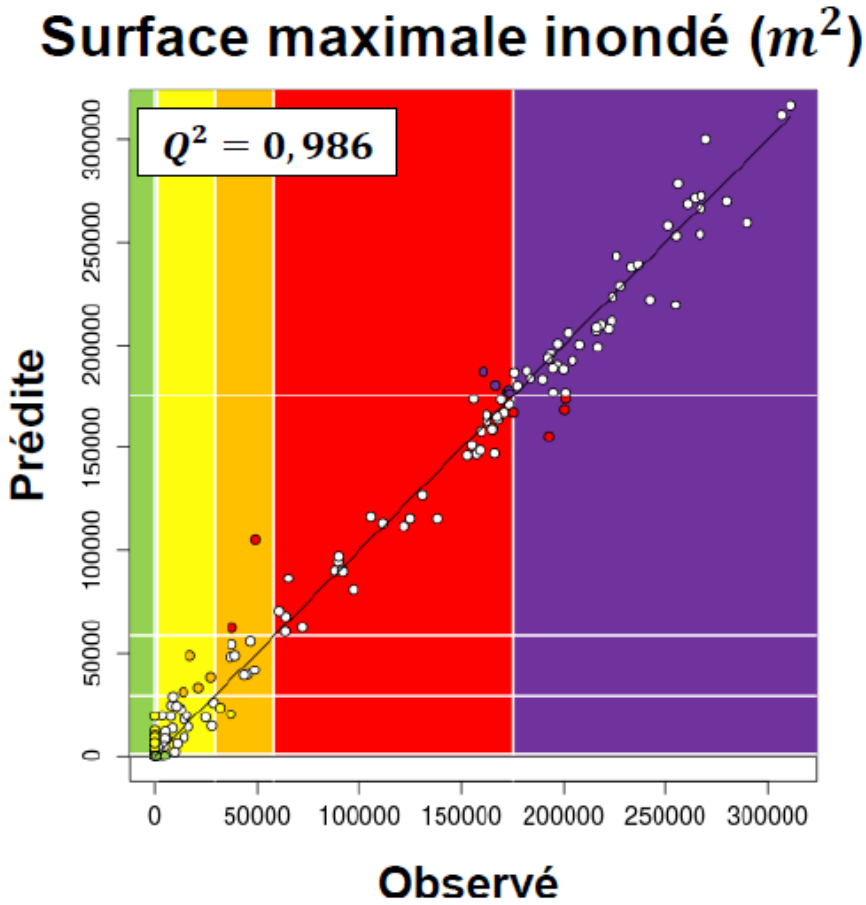


Betancourt et al. (2020)

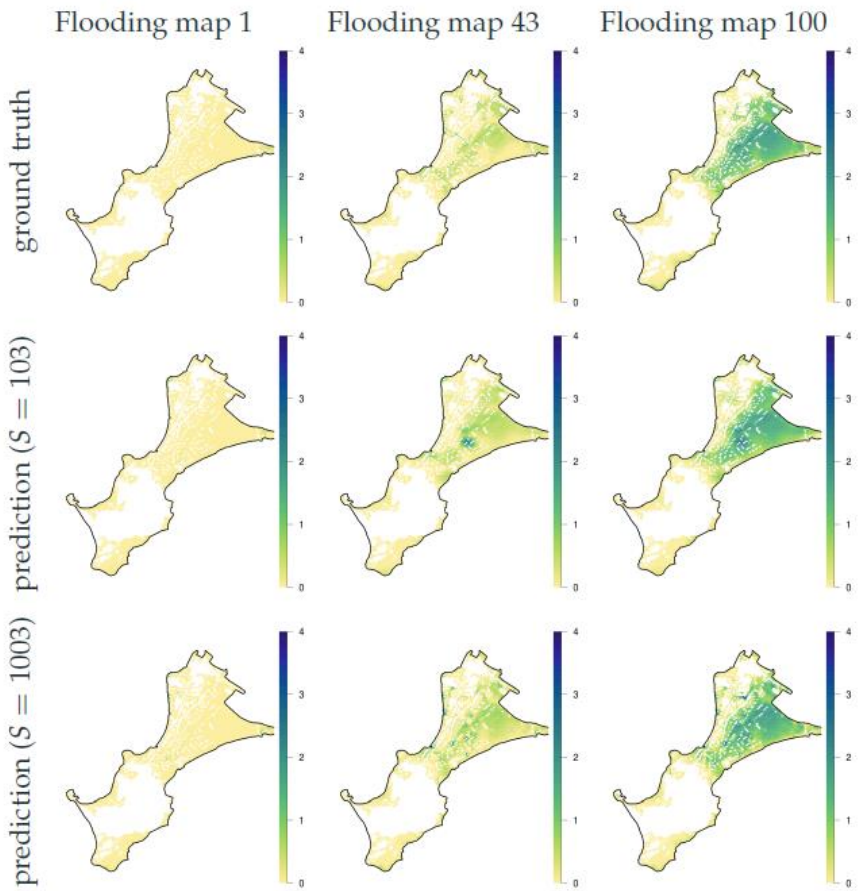
- **Apprentissage avec 174 résultats de simulation**
- **Estimation de la « capacité à prédire » par validation croisée**

Au final…

Prédire des sorties spatiales est possible!!



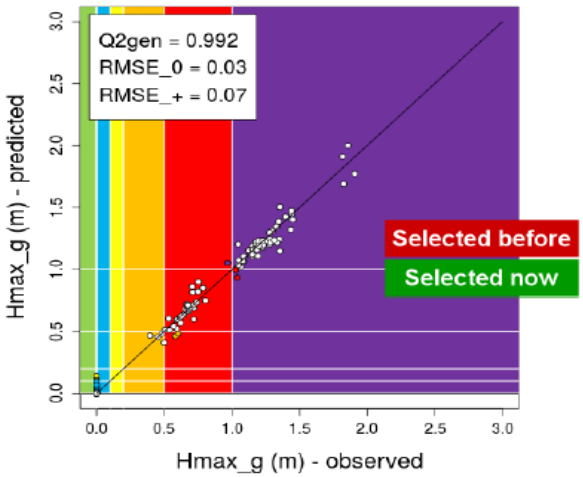
Betancourt et al. (2020)

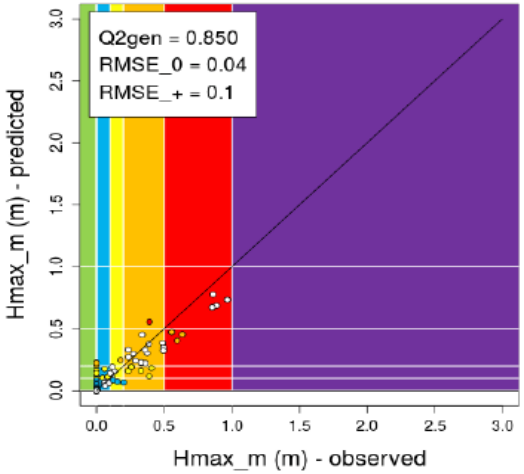


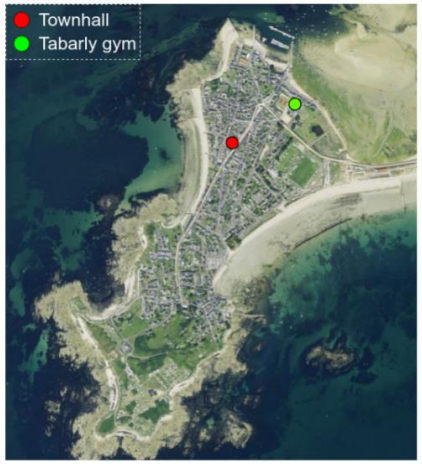
López-Lopera et al. (2020)

Ainsi que les **autres cibles**…

Hauteur d'eau aux points critiques



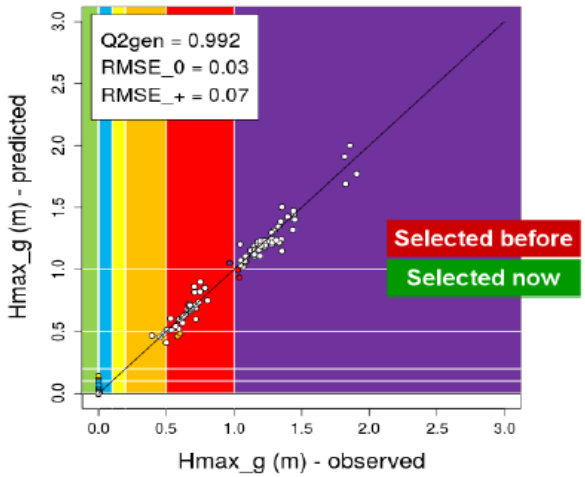


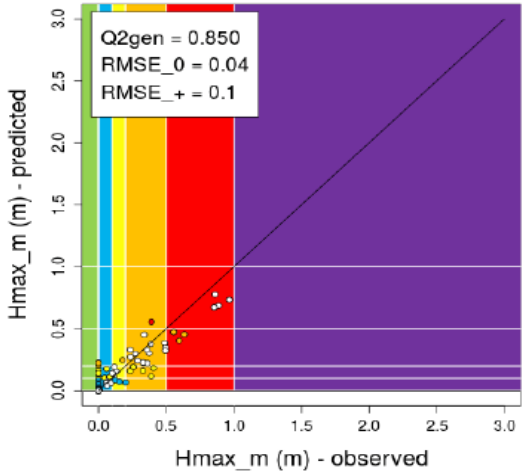


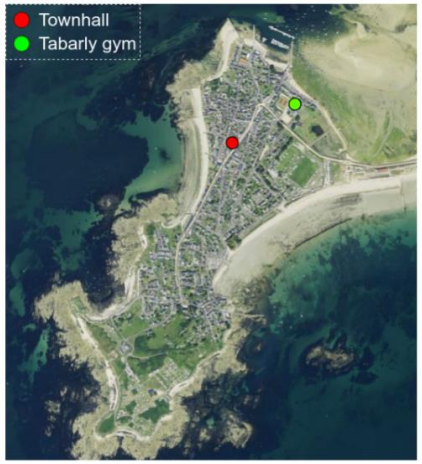
Betancourt (2020)

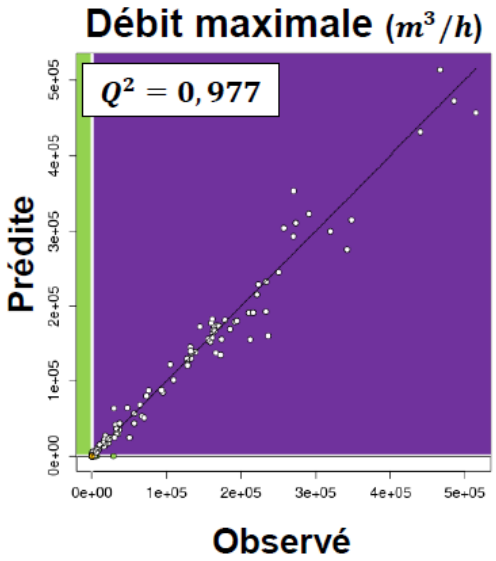
Ainsi que les **autres cibles**…

Hauteur d'eau aux points critiques





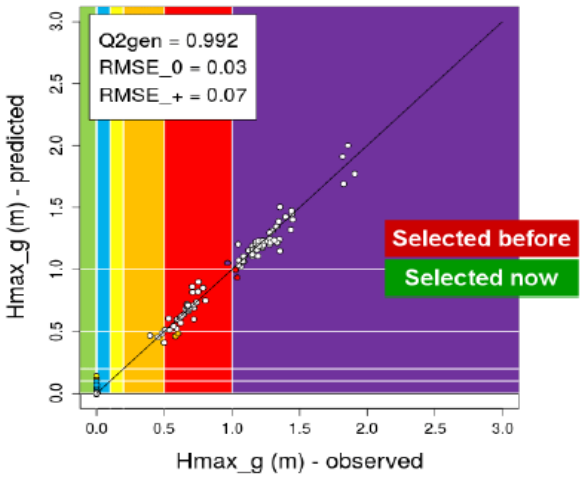


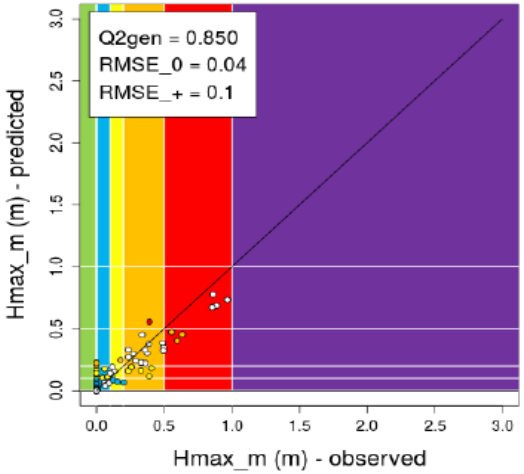


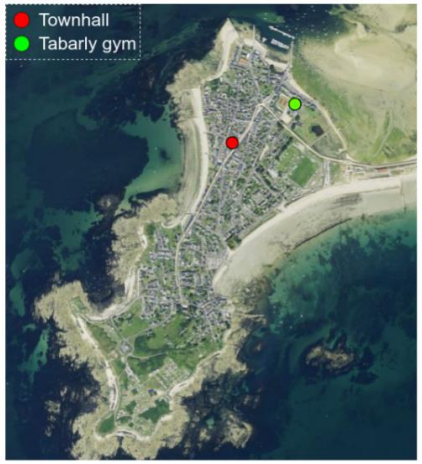
Betancourt (2020)

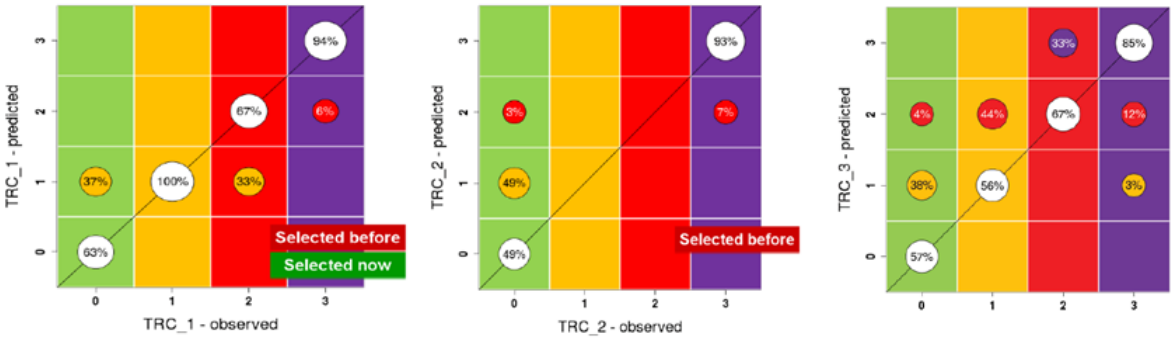
Ainsi que les **autres cibles**…

Hauteur d'eau aux points critiques





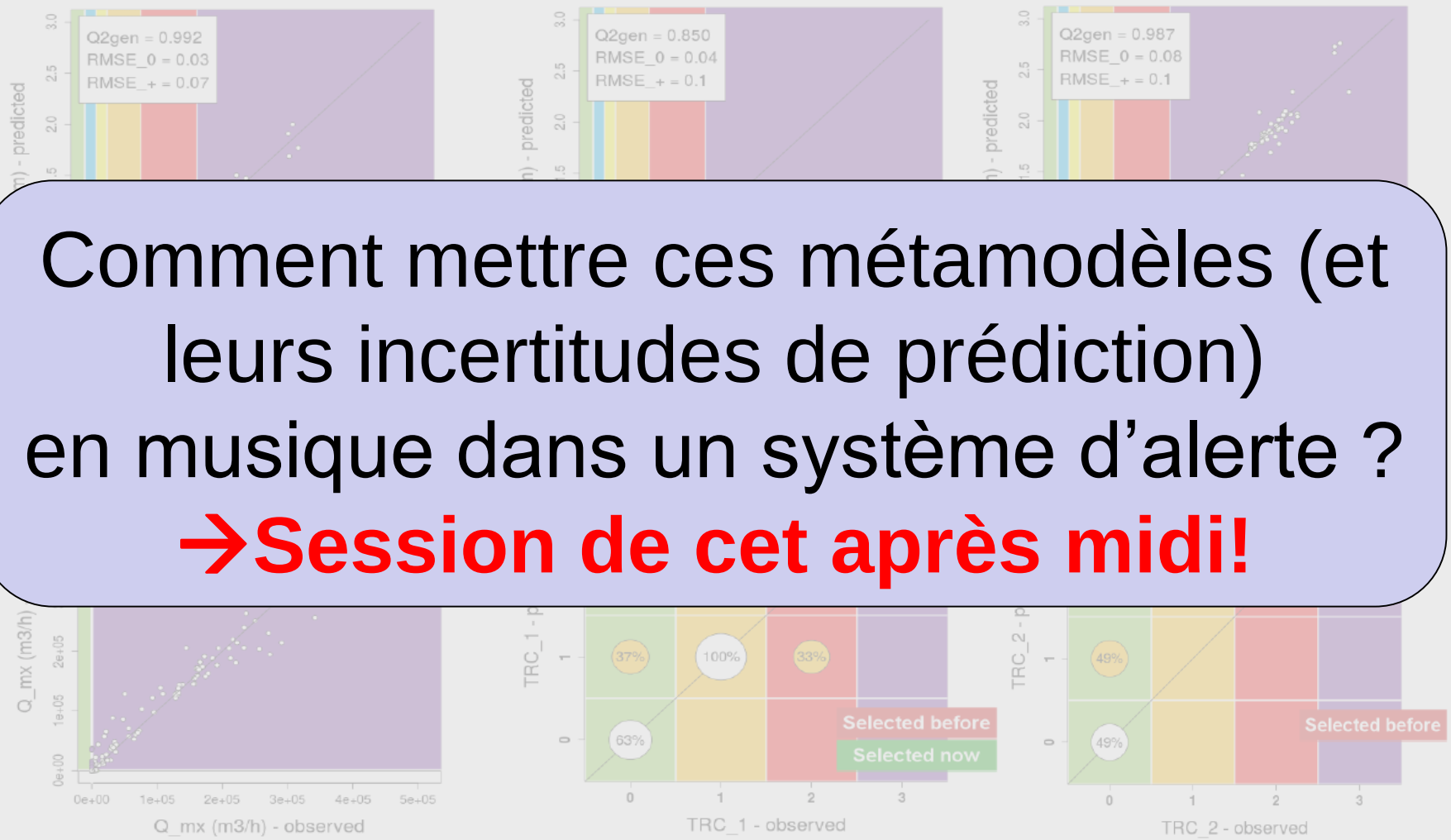




Trafficabilité sur les tronçons de route

Betancourt (2020)

Ainsi que les **autres cibles**…



Comment mettre ces métamodèles (et leurs incertitudes de prédiction) en musique dans un système d'alerte ?

→Session de cet après midi!

Betancourt et al. (2020)