

# Méta-modélisation et analyse de sensibilité pour les modèles avec sortie spatiale. Application aux modèles de submersion marine.

Rencontres MEXICO

Tran Vi-vi Élodie PERRIN

[elodie.perrin@onera.fr](mailto:elodie.perrin@onera.fr)

30 Novembre 2021

---

Directeur de thèse:	Olivier ROUSTANT	INSA Toulouse
Directrice de thèse:	Mireille BATTON-HUBERT	Mines Saint-Étienne
Co-encadrants:	Jérémy ROHMER	BRGM Orléans
	Jean-Philippe NAULIN	CCR Paris

## **BRGM : Bureau de Recherches Géologiques et Minières**

Organisme public français de référence dans le domaine des sciences de la Terre pour la gestion des ressources naturelles et des risques du sol et du sous-sol.

## **CCR : Caisse Centrale de Réassurance**

Société française qui propose aux assureurs une couverture de réassurance dans le cadre du régime d'indemnisation des catastrophes naturelles.



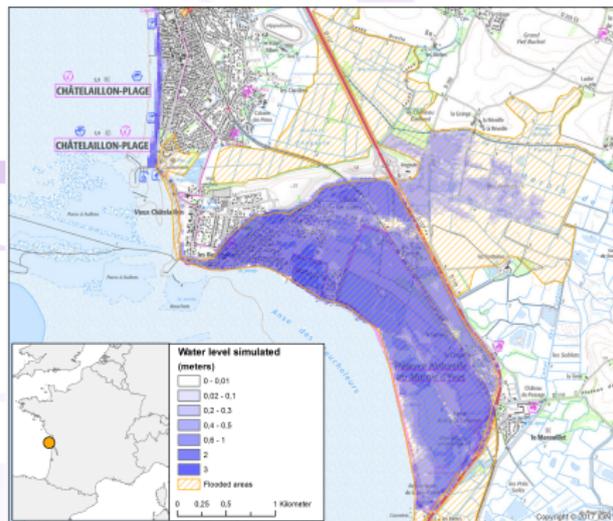
# Introduction

---

## Contexte

### Les submersions marines :

- **Phénomène dévastateur :** cyclone Irma (2017), ouragan Sandy (2012).
- **Tempête Xynthia, 2010 :** 53 morts, 79 blessés et 2.5 milliards d'euros de dommages [Naulin et al., 2015] [FFSA, 2011].
- **Objectif du BRGM et de la CCR :**  
Estimer l'impact et l'étendue spatiale des inondations induites par les tempêtes.

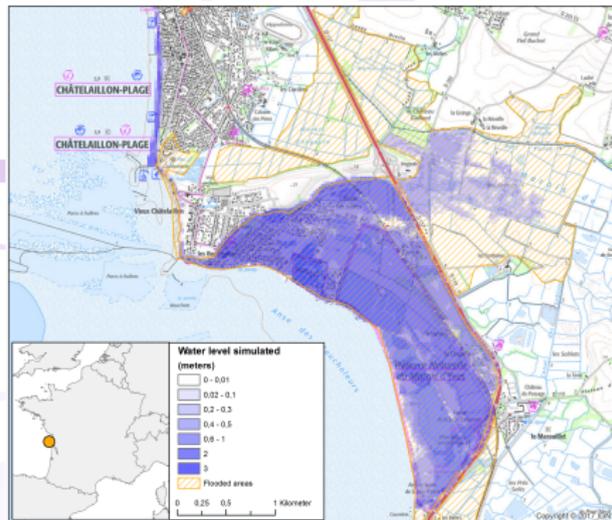


**Figure 1:** Carte de l'inondation des Boucholeurs durant la tempête Xynthia en 2010. La zone inondée est hachurée. Celle estimée par la CCR est en bleu.

# Contexte

## Modèles d'aléa :

- Modèles de type différences finies
- **CCR** : modèle Lisflood-FP [Bates et al., 2005, Bates et al., 2010].
- **BRGM** : modèle MARS [Lazure and Dumas, 2008], avec prise en compte des spécificités de processus d'inondations locaux. (connexions comme les buses, déversoirs, etc. et les phénomènes de brèches)



**Figure 1:** Carte de l'inondation des Boucholeurs durant la tempête Xynthia en 2010. La zone inondée est hachurée. Celle estimée par la CCR est en bleu.

## Contexte

### Sortie des modèles d'aléa :

- Carte de l'inondation estimée.

Inondation quantifiée par **la profondeur d'eau (HE) maximale** atteinte sur la durée d'un événement de tempête donné.

### Problématique

- **Incertitudes** dans les données entrées dans les deux modèles :
  - **Forçages marins au large** : vagues et niveau d'eau des mers ou des océans.
  - **MNT** : altitude du modèle numérique de terrain.
  - **Rugosité du sol.**

# Objectif

## Analyse de sensibilité (AS) des modèles d'aléa

- Étudier l'influence des entrées du modèle sur l'estimation de l'inondation.
- Identifier et hiérarchiser les sources d'incertitudes pouvant être améliorées dans les modèles d'aléa.
- Plusieurs méthodes : régression, méthodes basées sur la variance, nuage de points etc.

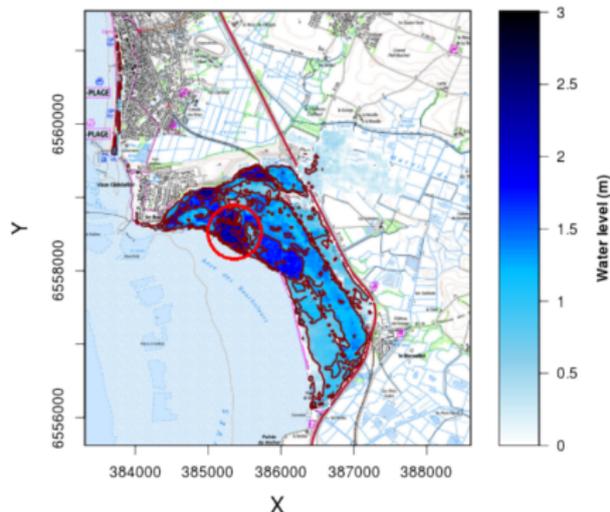
## Problématiques

### Temps de calcul des modèles :

- Nécessité d'un grand nombre de simulations (méthode de Monte Carlo).
- Pour une simulation :
  - BRGM : 30 min-1h
  - CCR  $\approx$  5 min

### Sortie fonctionnelle (spatiale) :

- HE dépend de la localisation.
- Grande dimension :
  - Résolution : 25m  $\times$  25m
  - BRGM : 194  $\times$  202 pixels
  - CCR : 213  $\times$  213 pixels
- Fortes discontinuités et variations.



**Figure 2:** Exemple de carte estimée par le modèle d'aléa du BRGM.

# Méthodologie usuelle

1. **Réduction de dimension de la sortie spatiale** : décomposition sur une base orthonormée :
  - le plus souvent par **analyse en composantes principales (ACP)** [Chen et al., 2011],
  - d'autres **bases de fonctions** : ondelettes (voir [Marrel et al., 2010, Marrel et al., 2015])

## 2. Méta-modélisation et AS :

2.1 Méta-modélisation : un **méta-modèle** est construit pour **chaque coordonnée (ou coefficient)** de la base.

Fonction rapide à évaluer se basant sur un nombre limité de simulations.

## 2.2 Analyse de sensibilité :

- **AS sur chaque coordonnée ou coefficient** [Campbell et al., 2006].
- **Indice de sensibilité généralisé** (Indice de Lamboni) [Lamboni et al., 2011, Xiao and Li, 2016].

# Méthodologie usuelle

1. **Réduction de dimension de la sortie spatiale** : décomposition sur une base orthonormée :
  - le plus souvent par **analyse en composantes principales (ACP)** [Chen et al., 2011],
  - d'autres **bases de fonctions** : ondelettes (voir [Marrel et al., 2010, Marrel et al., 2015])

## 2. **Méta-modélisation et AS** :

2.1 **Méta-modélisation** : un **méta-modèle** est construit pour **chaque coordonnée (ou coefficient)** de la base.

Fonction rapide à évaluer se basant sur un nombre limité de simulations.

2.2 **Analyse de sensibilité** :

- **AS sur chaque coordonnée ou coefficient** [Campbell et al., 2006].
- **Indice de sensibilité généralisé** (Indice de Lamboni) [Lamboni et al., 2011, Xiao and Li, 2016].

# Méthodologie usuelle

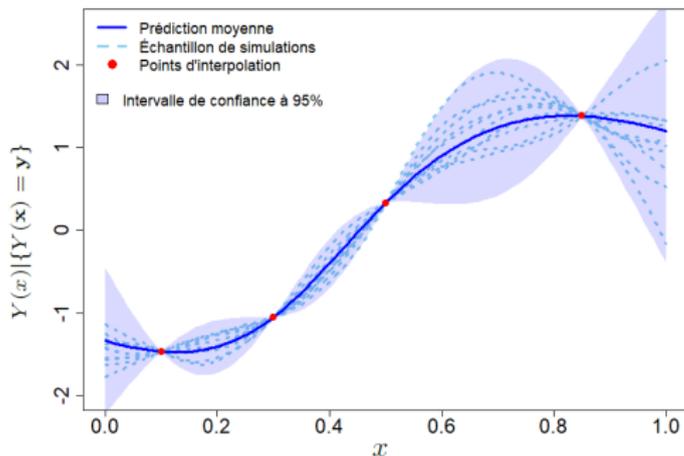
1. **Réduction de dimension de la sortie spatiale** : décomposition sur une base orthonormée :
  - le plus souvent par **analyse en composantes principales (ACP)** [Chen et al., 2011],
  - d'autres **bases de fonctions** : ondelettes (voir [Marrel et al., 2010, Marrel et al., 2015])
2. **Méta-modélisation et AS** :
  - 2.1 **Méta-modélisation** : un **méta-modèle** est construit pour **chaque coordonnée (ou coefficient)** de la base.

Fonction rapide à évaluer se basant sur un nombre limité de simulations.
  - 2.2 **Analyse de sensibilité** :
    - **AS sur chaque coordonnée ou coefficient** [Campbell et al., 2006].
    - **Indice de sensibilité généralisé** (Indice de Lamboni) [Lamboni et al., 2011, Xiao and Li, 2016].

## Méta-modèle : Processus gaussien (krigeage)

Dans la thèse, les coefficients (ou coordonnées) sont modélisés par des **modèles de régression par processus gaussien (RPG)** [Rasmussen and Williams, 2006] :

- Capacité à donner une prédiction et son incertitude.
- Flexibilité et exploitation de connaissance d'experts.



- Points d'interpolations :

$$\mathcal{D} = \{(x_i, y_i), i = 1, \dots, n\},$$

avec  $n \in \mathbb{N}$  (ici  $n = 4$ ).

- Loi  $Y(x) | \{Y(\mathbf{x}) = \mathbf{y}\}$  :

$$\mathcal{GP}(m(x), c_\theta(x, x'))$$

## Analyse de sensibilité : Indice de Sobol

- Indice entre 0 et 1, basé sur une décomposition de la variance [Sobol, 1993].
- Soit  $\mathbf{X} = (X_1, \dots, X_d)$  un vecteur de variables aléatoires réelles indépendantes. On suppose que  $Y = f(\mathbf{X})$  est une fonction de carré intégrable.
- Le  $i^{\text{ème}}$  indice de sensibilité de premier ordre est défini par :

$$S_i = \frac{\text{Var}[\mathbb{E}[Y|X_i]]}{\text{Var}[Y]}, \quad i \in \{1, \dots, d\}$$

- Indice total :

$$ST_i = 1 - \frac{\text{Var}[\mathbb{E}[Y|\mathbf{X}_{-i}]]}{\text{Var}[Y]}, \quad i \in \{1, \dots, d\}$$

avec  $\mathbf{X}_{-i} = \{X_j, j \in \{1, \dots, d\} \setminus \{i\}\}$ .

## De l'ACP à l'ACP fonctionnelle

### Points à améliorer

- **ACP standard** : dépendance spatiale non prise en compte et coûteuse en temps de calcul pour des cartes de grande dimension.
- **Ondelettes** : un grand nombre de coefficients d'ondelettes peut être nécessaire pour une bonne approximation.

### Solution : ACP fonctionnelle (ACPF) [Ramsay, 2006]

- Une extension de la méthode ACP pour des données fonctionnelles.
- **Objectif** : prendre en compte la dépendance spatiale tout en assurant la réduction de la dimension.

## De l'ACP à l'ACP fonctionnelle

### Points à améliorer

- **ACP standard** : dépendance spatiale non prise en compte et coûteuse en temps de calcul pour des cartes de grande dimension.
- **Ondelettes** : un grand nombre de coefficients d'ondelettes peut être nécessaire pour une bonne approximation.

### Solution : ACP fonctionnelle (ACPF) [Ramsay, 2006]

- Une extension de la méthode ACP pour des données fonctionnelles.
- **Objectif** : prendre en compte la dépendance spatiale tout en assurant la réduction de la dimension.

## ACP fonctionnelle (ACPF)

- Les fonctions sont vues comme des réalisations d'un **processus aléatoire gaussien centré**  $Y$ , avec pour **noyau de covariance**  $k$ .
- La décomposition de Karhunen-Loève (KL) donne une décomposition propre de  $k$  :

$$k(\mathbf{z}, \mathbf{z}') = \sum_{l \in \mathbb{N}} \lambda_l \psi_l(\mathbf{z}) \psi_l(\mathbf{z}').$$

avec  $(\lambda_l)_{l \in \mathbb{N}}$  les valeurs propres et  $(\psi_l(\mathbf{z}))_{l \in \mathbb{N}}$  les fonctions propres orthonormées associées.

- Extension de l'ACP standard qui diagonalise la matrice de covariance empirique.
- **L'ACPF dépend du choix de l'espace fonctionnel  $\mathcal{H}$  associé à  $k$ , appelé RKHS [Rasmussen and Williams, 2006].**

## ACP fonctionnelle (ACPF)

- Les fonctions sont vues comme des réalisations d'un **processus aléatoire gaussien centré**  $Y$ , avec pour **noyau de covariance**  $k$ .
- La décomposition de Karhunen-Loève (KL) donne une décomposition propre de  $k$  :

$$k(\mathbf{z}, \mathbf{z}') = \sum_{l \in \mathbb{N}} \lambda_l \psi_l(\mathbf{z}) \psi_l(\mathbf{z}').$$

avec  $(\lambda_l)_{l \in \mathbb{N}}$  les valeurs propres et  $(\psi_l(\mathbf{z}))_{l \in \mathbb{N}}$  les fonctions propres orthonormées associées.

- Extension de l'ACP standard qui diagonalise la matrice de covariance empirique.
- L'ACPF dépend du choix de l'espace fonctionnel  $\mathcal{H}$  associé à  $k$ , appelé RKHS [Rasmussen and Williams, 2006].

## ACP fonctionnelle (ACPF)

- Les fonctions sont vues comme des réalisations d'un **processus aléatoire gaussien centré**  $Y$ , avec pour **noyau de covariance**  $k$ .
- La décomposition de Karhunen-Loève (KL) donne une décomposition propre de  $k$  :

$$k(\mathbf{z}, \mathbf{z}') = \sum_{l \in \mathbb{N}} \lambda_l \psi_l(\mathbf{z}) \psi_l(\mathbf{z}').$$

avec  $(\lambda_l)_{l \in \mathbb{N}}$  les valeurs propres et  $(\psi_l(\mathbf{z}))_{l \in \mathbb{N}}$  les fonctions propres orthonormées associées.

- Extension de l'ACP standard qui diagonalise la matrice de covariance empirique.
- L'ACPF dépend du choix de l'espace fonctionnel  $\mathcal{H}$  associé à  $k$ , appelé RKHS [Rasmussen and Williams, 2006].

## ACP fonctionnelle (ACPF)

- Les fonctions sont vues comme des réalisations d'un **processus aléatoire gaussien centré**  $Y$ , avec pour **noyau de covariance**  $k$ .
- La décomposition de Karhunen-Loève (KL) donne une décomposition propre de  $k$  :

$$k(\mathbf{z}, \mathbf{z}') = \sum_{l \in \mathbb{N}} \lambda_l \psi_l(\mathbf{z}) \psi_l(\mathbf{z}').$$

avec  $(\lambda_l)_{l \in \mathbb{N}}$  les valeurs propres et  $(\psi_l(\mathbf{z}))_{l \in \mathbb{N}}$  les fonctions propres orthonormées associées.

- Extension de l'ACP standard qui diagonalise la matrice de covariance empirique.
- **L'ACPF dépend du choix de l'espace fonctionnel  $\mathcal{H}$  associé à  $k$ , appelé RKHS [Rasmussen and Williams, 2006].**

## ACPF : définition de $\mathcal{H}$

- **Méthode 1** : choix d'un noyau usuel  $k$ .
- **Méthode 2** : définir  $\mathcal{H}$  comme un espace de dimension finie définie par une base de fonctions donnée  $(\phi_k(\mathbf{z}))_{k=1,\dots,K}$  [Ramsay, 2006].
- La **méthode 2** permet de définir le noyau par rapport à des **non-stationnarités locales** (méthode choisie pour la thèse).

ACP fonctionnelle  $\Leftrightarrow$  Application d'une **ACP aux coefficients** d'une décomposition sur une base de fonctions  $(\phi_k(\mathbf{z}))_{k=1,\dots,K}$ , avec la métrique donnée par la matrice de Gram  $\mathbf{G}$  de la base.

$$\mathbf{G} = \left( \int \phi_k(\mathbf{z}) \phi_{k'}(\mathbf{z}) d\mu(\mathbf{z}) \right)_{1 \leq k, k' \leq K}$$

## ACPF : définition de $\mathcal{H}$

- **Méthode 1** : choix d'un noyau usuel  $k$ .
- **Méthode 2** : définir  $\mathcal{H}$  comme un espace de dimension finie définie par une base de fonctions donnée  $(\phi_k(\mathbf{z}))_{k=1,\dots,K}$  [Ramsay, 2006].
- La **méthode 2** permet de définir le noyau par rapport à des **non-stationnarités locales** (méthode choisie pour la thèse).

**ACP fonctionnelle**  $\Leftrightarrow$  Application d'une **ACP aux coefficients** d'une décomposition sur une base de fonctions  $(\phi_k(\mathbf{z}))_{k=1,\dots,K}$ , avec la **métrique** donnée par la **matrice de Gram  $\mathbf{G}$**  de la base.

$$\mathbf{G} = \left( \int \phi_k(\mathbf{z}) \phi_{k'}(\mathbf{z}) d\mu(\mathbf{z}) \right)_{1 \leq k, k' \leq K}.$$

## Principales contributions

1. Méta-modélisation : ajout à l'ACPF d'une **étape de sélection des coefficients** de la base de fonctions choisie
  - sélection se basant sur la **décomposition de l'énergie (variance spatiale)**.
  - sélection selon une approche de **régression pénalisée (Lasso)**.
2. Analyse de sensibilité
  - **Analyse de sensibilité des composantes principales**.
  - **Indice de sensibilité généralisé : extension de l'indice [Lamboni et al., 2011] valide pour toute base de fonctions**.
3. Développement d'un **package R « GpOutput2D »**  
Pour la méta-modélisation de modèles avec sortie spatiale.
4. Premiers résultats prometteurs sur la perspective de combiner les modèles BRGM et CCR.

## Principales contributions

1. Méta-modélisation : ajout à l'ACPF d'une **étape de sélection des coefficients** de la base de fonctions choisie
  - sélection se basant sur la **décomposition de l'énergie (variance spatiale)**.
  - sélection selon une approche de **régression pénalisée (Lasso)**.
2. Analyse de sensibilité
  - **Analyse de sensibilité des composantes principales**.
  - Indice de sensibilité généralisé : **extension de l'indice [Lamboni et al., 2011] valide pour toute base de fonctions**.
3. Développement d'un **package R « GpOutput2D »**  
Pour la méta-modélisation de modèles avec sortie spatiale.
4. Premiers résultats prometteurs sur la perspective de combiner les modèles BRGM et CCR.

## Principales contributions

1. Méta-modélisation : ajout à l'ACPF d'une **étape de sélection des coefficients** de la base de fonctions choisie
  - sélection se basant sur la **décomposition de l'énergie (variance spatiale)**.
  - sélection selon une approche de **régression pénalisée (Lasso)**.
2. Analyse de sensibilité
  - **Analyse de sensibilité des composantes principales**.
  - Indice de sensibilité généralisé : **extension de l'indice [Lamboni et al., 2011] valide pour toute base de fonctions**.
3. Développement d'un **package R « GpOutput2D »**  
Pour la méta-modélisation de modèles avec sortie spatiale.
4. Premiers résultats prometteurs sur la perspective de combiner les modèles BRGM et CCR.

## Principales contributions

1. Méta-modélisation : ajout à l'ACPF d'une **étape de sélection des coefficients** de la base de fonctions choisie
  - sélection se basant sur la **décomposition de l'énergie (variance spatiale)**.
  - sélection selon une approche de **régression pénalisée (Lasso)**.
2. Analyse de sensibilité
  - **Analyse de sensibilité des composantes principales**.
  - Indice de sensibilité généralisé : **extension de l'indice [Lamboni et al., 2011] valide pour toute base de fonctions**.
3. Développement d'un **package R « GpOutput2D »**  
Pour la méta-modélisation de modèles avec sortie spatiale.
4. Premiers résultats prometteurs sur la perspective de combiner les modèles BRGM et CCR.

# Plan de la présentation

1. Méta-modélisation de modèle avec sortie spatiale
  - 1.1 Méta-modélisation avec réduction de dimension par ACPF
  - 1.2 Application à un cas analytique
  - 1.3 Application à un cas de submersion marine
  - 1.4 Package R : GpOutput2D
2. Analyse de sensibilité des modèles d'aléa
  - 2.1 Cas d'étude
  - 2.2 Analyse de sensibilité des composantes principales
  - 2.3 Indices de sensibilité généralisés
3. Conclusions et perspectives

# Méta-modélisation de modèle avec sortie spatiale

---

## 1. Méta-modélisation de modèle avec sortie spatiale

1.1 Méta-modélisation avec réduction de dimension par ACPF

1.2 Application à un cas analytique

1.3 Application à un cas de submersion marine

1.4 Package **R** : GpOutput2D

## Formulation mathématique du problème

- On considère le simulateur suivant :

$$\begin{aligned} f &: \mathcal{X} \subseteq \mathbb{R}^d \rightarrow \mathbb{L}^2(\mathcal{Z}) \\ \mathbf{x} &\mapsto y_{\mathbf{x}}(\mathbf{z}) \end{aligned}$$

où  $\mathbf{x}$  est le vecteur des entrées, et  $\mathcal{Z}$  est un domaine spatial.

$y_{\mathbf{x}}(\mathbf{z})$  est la sortie de  $f$ , une fonction de  $\mathbb{L}^2(\mathcal{Z})$ .

- $y_{\mathbf{x}}(\mathbf{z})$  est assimilé à la valeur de la carte à la localisation  $\mathbf{z}$ .
- On connaît  $n \in \mathbb{N}$  simulations de  $f$  :

$$\mathcal{D} = \{(\mathbf{x}_i, y_{\mathbf{x}_i}(\mathbf{z})), i = 1, \dots, n\}$$

- Objectif** : prédiction de  $f(\mathbf{x}^*)$  en un nouveau point  $\mathbf{x}^*$ , sans faire de nouveau appel à  $f$ .

## Étape 1 : Décomposition de la sortie spatiale

1. Décomposition de la sortie  $y_{\mathbf{x}}(\mathbf{z})$  dans une base de fonctions  $\Phi(\mathbf{z}) = (\phi_1(\mathbf{z}), \dots, \phi_K(\mathbf{z}))^\top$  (ondelettes, B-splines, etc.), avec  $K$  la taille de la base. Pour tout  $\mathbf{x} \in \mathcal{X}$ , on a alors :

$$y_{\mathbf{x}}(\mathbf{z}) = \sum_{k=1}^K \alpha_k(\mathbf{x}) \phi_k(\mathbf{z}) = \alpha(\mathbf{x})^\top \Phi(\mathbf{z})$$

où  $\alpha(\mathbf{x}) = (\alpha_1(\mathbf{x}), \dots, \alpha_K(\mathbf{x}))^\top$  est le vecteur des coefficients.

# Étape 1 : Décomposition de la sortie spatiale

## 1. Deux possibilités :

1.1 Si  $\Phi(\mathbf{z})$  est une base orthonormée, alors on a :

$$\|y_{\mathbf{x}}(\mathbf{z})\|_2 = \sum_{k=1}^K \alpha_k(\mathbf{x})^2$$

1.2 (Régression Lasso) [Tibshirani, 1996]

Estimation de  $\alpha(\mathbf{x})$  tel que :

$$\min_{\alpha_1(\mathbf{x}^{(i)}), \dots, \alpha_K(\mathbf{x}^{(i)})} \|y_{\mathbf{x}^{(i)}}(\mathbf{z}) - \sum_{k=1}^K \alpha_k(\mathbf{x}^{(i)}) \phi_k(\mathbf{z})\|^2 + \lambda \sum_{k=1}^K |\alpha_k(\mathbf{x}^{(i)})|, \quad i = 1, \dots, n$$

$\lambda > 0$  est un paramètre de pénalité commun pour tout  $\mathbf{x} \in \mathcal{X}$ .

# Étape 1 : Décomposition de la sortie spatiale

## 1. Deux possibilités :

1.1 Si  $\Phi(\mathbf{z})$  est une base orthonormée, alors on a :

$$\|y_{\mathbf{x}}(\mathbf{z})\|_2 = \sum_{k=1}^K \alpha_k(\mathbf{x})^2$$

1.2 (Régression Lasso) [Tibshirani, 1996]

Estimation de  $\alpha(\mathbf{x})$  tel que :

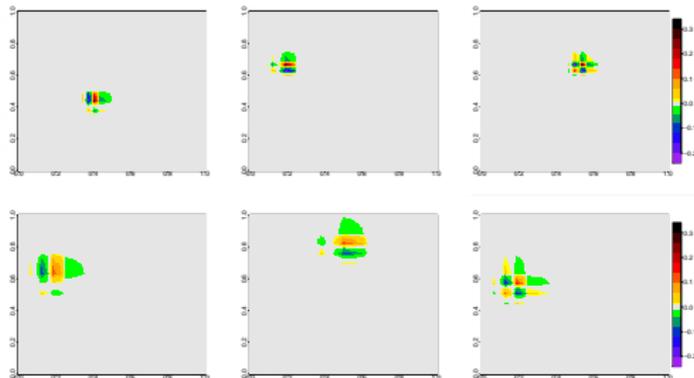
$$\min_{\alpha_1(\mathbf{x}^{(i)}), \dots, \alpha_K(\mathbf{x}^{(i)})} \|y_{\mathbf{x}^{(i)}}(\mathbf{z}) - \sum_{k=1}^K \alpha_k(\mathbf{x}^{(i)}) \phi_k(\mathbf{z})\|^2 + \lambda \sum_{k=1}^K |\alpha_k(\mathbf{x}^{(i)})|, \quad i = 1, \dots, n$$

$\lambda > 0$  est un paramètre de pénalité commun pour tout  $\mathbf{x} \in \mathcal{X}$ .

# Étape 1 : Choix de la base de fonctions $\Phi(\mathbf{z})$

## Ondelettes

- Fonctions oscillantes définies sur un ensemble compact
- (Analyse multi-résolutions) Fonctions de la base : versions translattées et dilatées d'une ondelette « mère ».

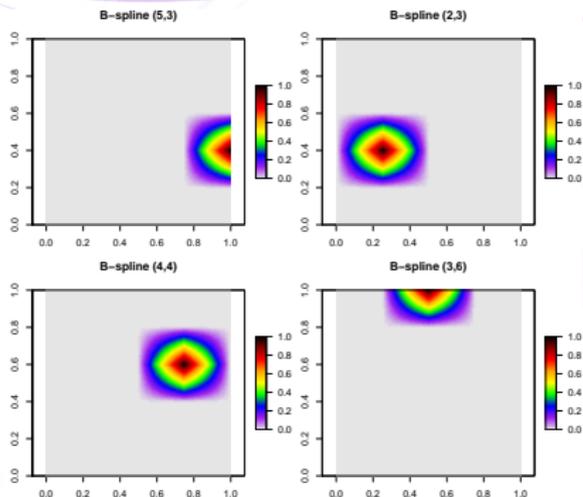


**Figure 3:** Exemple d'ondelettes de Daubechies D4 sur  $[0, 1]^2$ . Les figures du haut sont des ondelettes à l'échelle 3. Celles du bas sont des ondelettes à l'échelle 4.

## Étape 1 : Choix de la base de fonctions $\Phi(\mathbf{z})$

### B-splines

- Fonctions par morceaux définies par des polynômes.
- Dans la thèse : B-splines de degré 1.



**Figure 4:** Exemples de fonctions d'une base B-splines de degré 1, définie sur  $[0,1]^2$ .

## Étape 2 : Sélection des coefficients

2. Trier les coefficients  $(\alpha_k(\mathbf{x}))_{k=1,\dots,K}$  « du plus important au moins important ». **Deux possibilités** :

2.1 Si  $\Phi(\mathbf{z})$  est une base orthonormée, trier selon l'ordre décroissant de :

$$\lambda_k = \mathbb{E} \left[ \frac{\alpha_k(\mathbf{x})^2}{\sum_{k'=1}^K \alpha_{k'}(\mathbf{x})^2} \right], \quad k = 1, \dots, K$$

2.2 Après régression Lasso, trier selon l'ordre décroissant de :

$$\lambda_k = \mathbb{E}(\alpha_k(\mathbf{x})) \text{ ou } \lambda_k = \mathbb{P}(\alpha_k(\mathbf{x}) \neq 0), \quad k = 1, \dots, K.$$

On note  $(k)$ ,  $k = 1, \dots, K$ , les indices des coefficients suivant le nouvel ordre.

## Étape 2 : Sélection des coefficients

2. Trier les coefficients  $(\alpha_k(\mathbf{x}))_{k=1,\dots,K}$  « du plus important au moins important ». **Deux possibilités** :

2.1 Si  $\Phi(\mathbf{z})$  est une base orthonormée, trier selon l'ordre décroissant de :

$$\lambda_k = \mathbb{E} \left[ \frac{\alpha_k(\mathbf{x})^2}{\sum_{k'=1}^K \alpha_{k'}(\mathbf{x})^2} \right], \quad k = 1, \dots, K$$

2.2 Après régression Lasso, trier selon l'ordre décroissant de :

$$\lambda_k = \mathbb{E}(\alpha_k(\mathbf{x})) \text{ ou } \lambda_k = \mathbb{P}(\alpha_k(\mathbf{x}) \neq 0), \quad k = 1, \dots, K.$$

On note  $(k)$ ,  $k = 1, \dots, K$ , les indices des coefficients suivant le nouvel ordre.

## Étape 2 : Sélection des coefficients

2. Trier les coefficients  $(\alpha_k(\mathbf{x}))_{k=1,\dots,K}$  « du plus important au moins important ». **Deux possibilités** :

2.1 Si  $\Phi(\mathbf{z})$  est une base orthonormée, trier selon l'ordre décroissant de :

$$\lambda_k = \mathbb{E} \left[ \frac{\alpha_k(\mathbf{x})^2}{\sum_{k'=1}^K \alpha_{k'}(\mathbf{x})^2} \right], \quad k = 1, \dots, K$$

2.2 Après régression Lasso, trier selon l'ordre décroissant de :

$$\lambda_k = \mathbb{E}(\alpha_k(\mathbf{x})) \text{ ou } \lambda_k = \mathbb{P}(\alpha_k(\mathbf{x}) \neq 0), \quad k = 1, \dots, K.$$

On note  $(k)$ ,  $k = 1, \dots, K$ , les indices des coefficients suivant le nouvel ordre.

## Étape 2 : Sélection des coefficients

3. Sélectionner les  $\tilde{K} \ll K$  premiers coefficients  $(\alpha_{(k)}(\mathbf{x}))_{k=1, \dots, K}$ .  
Choix de  $\tilde{K}$  :

- Calibration par validation croisée.
- Si tri selon le critère 2.1 :

$$\lambda_k = \mathbb{E} \left[ \frac{\alpha_k(\mathbf{x})^2}{\sum_{k'=1}^K \alpha_{k'}(\mathbf{x})^2} \right], \quad k = 1, \dots, K$$

On peut fixer  $p \in [0, 1]$  qui correspond à **une proportion de l'énergie moyenne**.  $\tilde{K}$  est choisi tel que

$$\sum_{k=1}^{\tilde{K}} \lambda_{(k)} \leq p$$

$p$  peut aussi être choisi par validation croisée.

## Étape 2 : Sélection des coefficients

3. Sélectionner les  $\tilde{K} \ll K$  premiers coefficients  $(\alpha_{(k)}(\mathbf{x}))_{k=1, \dots, K}$ .

Choix de  $\tilde{K}$  :

- Calibration par validation croisée.
- Si tri selon le critère 2.1 :

$$\lambda_k = \mathbb{E} \left[ \frac{\alpha_k(\mathbf{x})^2}{\sum_{k'=1}^K \alpha_{k'}(\mathbf{x})^2} \right], \quad k = 1, \dots, K$$

On peut fixer  $p \in [0, 1]$  qui correspond à **une proportion de l'énergie moyenne**.  $\tilde{K}$  est choisi tel que

$$\sum_{k=1}^{\tilde{K}} \lambda_{(k)} \leq p$$

$p$  peut aussi être choisi par validation croisée.

## Étape 2 : Sélection des coefficients

3. Sélectionner les  $\tilde{K} \ll K$  premiers coefficients  $(\alpha_{(k)}(\mathbf{x}))_{k=1, \dots, K}$ .  
Choix de  $\tilde{K}$  :

- Calibration par validation croisée.
- Si tri selon le critère 2.1 :

$$\lambda_k = \mathbb{E} \left[ \frac{\alpha_k(\mathbf{x})^2}{\sum_{k'=1}^K \alpha_{k'}(\mathbf{x})^2} \right], \quad k = 1, \dots, K$$

On peut fixer  $p \in [0, 1]$  qui correspond à **une proportion de l'énergie moyenne**.  $\tilde{K}$  est choisi tel que

$$\sum_{k=1}^{\tilde{K}} \lambda_{(k)} \leq p$$

$p$  peut aussi être choisi par validation croisée.

## Étape 3 : prédiction des coefficients

### 4. Prédiction des coefficients $\alpha(\mathbf{x}^*) = (\alpha_1(\mathbf{x}^*), \dots, \alpha_K(\mathbf{x}^*))^\top$ :

- **Les  $\tilde{K}$  Coef. sélectionnés :**

4.1 Appliquer l'ACP dans  $\mathbb{R}^{\tilde{K}}$  aux coefficients  $(\alpha_{(k)}(\mathbf{x}_i))_{k=1, \dots, \tilde{K}}$ .

*Coordonnées (scores) :  $t_l(\mathbf{x}_1), \dots, t_{n_{PC}}(\mathbf{x}_i)$ .*

*Vecteurs propres :  $\omega_1, \dots, \omega_{n_{PC}}$ .*

4.2 Chaque  $t_l(\mathbf{x}^*)$ ,  $l = 1, \dots, n_{PC}$  est prédit par RPF, basé sur les observations  $\{(\mathbf{x}_i, t_l(\mathbf{x}_i)), i = 1, \dots, n\}$ .

4.3 Estimation des coefficients  $k = 1, \dots, \tilde{K}$  :

$$\hat{\alpha}_{(k)}(\mathbf{x}^*) = \sum_{l=1}^{n_{PC}} \hat{t}_l(\mathbf{x}^*) \omega_l.$$

- **Coef. non sélectionnés** estimés par moyenne empirique.

$$\text{Pour } k = (\tilde{K} + 1), \dots, K, \hat{\alpha}_{(k)}(\mathbf{x}^*) = \frac{1}{n} \sum_{i=1}^n \alpha_{(k)}(\mathbf{x}^{(i)}).$$

## Étape 3 : prédiction des coefficients

4. Prédiction des coefficients  $\alpha(\mathbf{x}^*) = (\alpha_1(\mathbf{x}^*), \dots, \alpha_K(\mathbf{x}^*))^\top$  :

- **Les  $\tilde{K}$  Coef. sélectionnés :**

4.1 Appliquer l'ACP dans  $\mathbb{R}^{\tilde{K}}$  aux coefficients  $(\alpha_{(k)}(\mathbf{x}_i))_{k=1, \dots, \tilde{K}}$ .  
 Coordonnées (scores) :  $t_1(\mathbf{x}_i), \dots, t_{n_{PC}}(\mathbf{x}_i)$ .  
 Vecteurs propres :  $\omega_1, \dots, \omega_{n_{PC}}$ .

4.2 Chaque  $t_l(\mathbf{x}^*)$ ,  $l = 1, \dots, n_{PC}$  est prédit par RPF, basé sur les observations  $\{(\mathbf{x}_i, t_l(\mathbf{x}_i)), i = 1, \dots, n\}$ .

4.3 Estimation des coefficients  $k = 1, \dots, \tilde{K}$  :

$$\hat{\alpha}_{(k)}(\mathbf{x}^*) = \sum_{l=1}^{n_{PC}} \hat{t}_l(\mathbf{x}^*) \omega_l.$$

- **Coef. non sélectionnés** estimés par moyenne empirique.

$$\text{Pour } k = (\tilde{K} + 1), \dots, K, \hat{\alpha}_{(k)}(\mathbf{x}^*) = \frac{1}{n} \sum_{i=1}^n \alpha_{(k)}(\mathbf{x}^{(i)}).$$

## Étape 3 : prédiction des coefficients

4. Prédiction des coefficients  $\alpha(\mathbf{x}^*) = (\alpha_1(\mathbf{x}^*), \dots, \alpha_K(\mathbf{x}^*))^\top$  :

- **Les  $\tilde{K}$  Coef. sélectionnés** :

4.1 Appliquer l'ACP dans  $\mathbb{R}^{\tilde{K}}$  aux coefficients  $(\alpha_{(k)}(\mathbf{x}_i))_{k=1, \dots, \tilde{K}}$ .  
 Coordonnées (scores) :  $t_1(\mathbf{x}_i), \dots, t_{n_{PC}}(\mathbf{x}_i)$ .  
 Vecteurs propres :  $\omega_1, \dots, \omega_{n_{PC}}$ .

4.2 Chaque  $t_l(\mathbf{x}^*)$ ,  $l = 1, \dots, n_{PC}$  est prédit par RPG, basé sur les observations  $\{\mathbf{x}_i, t_l(\mathbf{x}_i)\}$ ,  $i = 1, \dots, n$ .

4.3 Estimation des coefficients  $k = 1, \dots, \tilde{K}$  :

$$\hat{\alpha}_{(k)}(\mathbf{x}^*) = \sum_{l=1}^{n_{PC}} \hat{t}_l(\mathbf{x}^*) \omega_l.$$

- **Coef. non sélectionnés** estimés par moyenne empirique.

$$\text{Pour } k = (\tilde{K} + 1), \dots, K, \hat{\alpha}_{(k)}(\mathbf{x}^*) = \frac{1}{n} \sum_{i=1}^n \alpha_{(k)}(\mathbf{x}^{(i)}).$$

## Étape 3 : prédiction des coefficients

4. Prédiction des coefficients  $\alpha(\mathbf{x}^*) = (\alpha_1(\mathbf{x}^*), \dots, \alpha_K(\mathbf{x}^*))^\top$  :

- **Les  $\tilde{K}$  Coef. sélectionnés** :

4.1 Appliquer l'ACP dans  $\mathbb{R}^{\tilde{K}}$  aux coefficients  $(\alpha_{(k)}(\mathbf{x}_i))_{k=1, \dots, \tilde{K}}$ .  
 Coordonnées (scores) :  $t_1(\mathbf{x}_i), \dots, t_{n_{PC}}(\mathbf{x}_i)$ .  
 Vecteurs propres :  $\omega_1, \dots, \omega_{n_{PC}}$ .

4.2 Chaque  $t_l(\mathbf{x}^*)$ ,  $l = 1, \dots, n_{PC}$  est prédit par RPG, basé sur les observations  $\{\mathbf{x}_i, t_l(\mathbf{x}_i), i = 1, \dots, n\}$ .

4.3 Estimation des coefficients  $k = 1, \dots, \tilde{K}$  :

$$\hat{\alpha}_{(k)}(\mathbf{x}^*) = \sum_{l=1}^{n_{PC}} \hat{t}_l(\mathbf{x}^*) \omega_l.$$

- **Coef. non sélectionnés** estimés par moyenne empirique.

$$\text{Pour } k = (\tilde{K} + 1), \dots, K, \hat{\alpha}_{(k)}(\mathbf{x}^*) = \frac{1}{n} \sum_{i=1}^n \alpha_{(k)}(\mathbf{x}^{(i)}).$$

## Étape 3 : prédiction des coefficients

4. Prédiction des coefficients  $\alpha(\mathbf{x}^*) = (\alpha_1(\mathbf{x}^*), \dots, \alpha_K(\mathbf{x}^*))^\top$  :

- **Les  $\tilde{K}$  Coef. sélectionnés** :

4.1 Appliquer l'ACP dans  $\mathbb{R}^{\tilde{K}}$  aux coefficients  $(\alpha_{(k)}(\mathbf{x}_i))_{k=1, \dots, \tilde{K}}$ .  
 Coordonnées (scores) :  $t_1(\mathbf{x}_i), \dots, t_{n_{PC}}(\mathbf{x}_i)$ .  
 Vecteurs propres :  $\omega_1, \dots, \omega_{n_{PC}}$ .

4.2 Chaque  $t_l(\mathbf{x}^*)$ ,  $l = 1, \dots, n_{PC}$  est prédit par RPG, basé sur les observations  $\{\mathbf{x}_i, t_l(\mathbf{x}_i)\}$ ,  $i = 1, \dots, n$ .

4.3 Estimation des coefficients  $k = 1, \dots, \tilde{K}$  :

$$\hat{\alpha}_{(k)}(\mathbf{x}^*) = \sum_{l=1}^{n_{PC}} \hat{t}_l(\mathbf{x}^*) \omega_l.$$

- **Coef. non sélectionnés** estimés par moyenne empirique.

$$\text{Pour } k = (\tilde{K} + 1), \dots, K, \hat{\alpha}_{(k)}(\mathbf{x}^*) = \frac{1}{n} \sum_{i=1}^n \alpha_{(k)}(\mathbf{x}^{(i)}).$$

## Étape 4 : prédiction de la sortie spatiale

5. Estimation de  $f(\mathbf{x}^*) = y_{\mathbf{x}^*}(\mathbf{z})$ ,  $\mathbf{z} \in \mathcal{Z}$  avec les coefficients prédits  $(\hat{\alpha}_k(\mathbf{x}^*))_{k=1,\dots,K}$  à partir de la formule de décomposition.

$$\hat{f}(\mathbf{x}^*) = \hat{y}_{\mathbf{x}^*}(\mathbf{z}) = \sum_{k=1}^K \hat{\alpha}_k(\mathbf{x}^*) \phi_k(\mathbf{z}) = \hat{\alpha}(\mathbf{x}^*)^\top \Phi(\mathbf{z})$$

## Notations

- $GP^{PCA}$  : méta-modèle utilisant une **ACP standard**.
- Méthode de sélection des coefficients se basant sur le **critère de l'énergie** :
  - $GP_{wavelet}^{FPCA}$  : méta-modèle utilisant une **ACPF sur base d'ondelettes**
  - $GP_{B-splines}^{FPCA}$  : méta-modèle utilisant une **ACPF sur base B-splines après orthonormalisation de la base**.
- $GP_{B-splines}^{Lasso}$  : méta-modèle se basant sur un modèle de **régression Lasso** pour la sélection des coefficients, **ACPF sur base B-splines (sur la base d'origine)**.

## Notations

- $GP^{PCA}$  : méta-modèle utilisant une **ACP standard**.
- Méthode de sélection des coefficients se basant sur le **critère de l'énergie** :
  - $GP^{FPCA}_{wavelet}$  : méta-modèle utilisant une **ACPF sur base d'ondelettes**
  - $GP^{FPCA}_{B-splines}$  : méta-modèle utilisant **une ACPF sur base B-splines après orthonormalisation de la base**.
- $GP^{Lasso}_{B-splines}$  : méta-modèle se basant sur un modèle de **régression Lasso** pour la sélection des coefficients, **ACPF sur base B-splines (sur la base d'origine)**.

## Notations

- $GP^{PCA}$  : méta-modèle utilisant une **ACP standard**.
- Méthode de sélection des coefficients se basant sur le **critère de l'énergie** :
  - $GP^{FPCA}_{wavelet}$  : méta-modèle utilisant une **ACPF sur base d'ondelettes**
  - $GP^{FPCA}_{B-splines}$  : méta-modèle utilisant une **ACPF sur base B-splines après orthonormalisation de la base**.
- $GP^{Lasso}_{B-splines}$  : méta-modèle se basant sur un modèle de **régression Lasso** pour la sélection des coefficients, **ACPF sur base B-splines (sur la base d'origine)**.

## Cas analytique

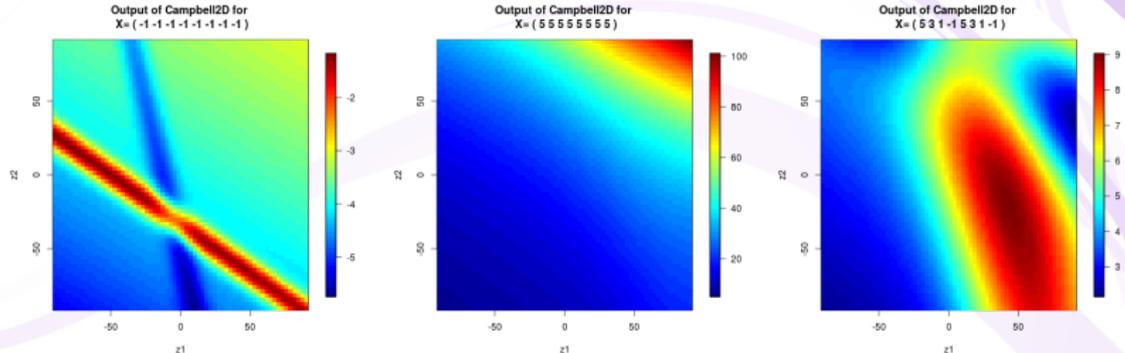
La fonction Campbell2D [Marrel et al., 2010] :

$$f : [-1,5]^8 \rightarrow \mathbb{L}^2([-90,90]^2)$$

$$\mathbf{x} = (x_1, \dots, x_8) \mapsto y_{\mathbf{x}}(\mathbf{z})$$

où  $\mathbf{z} = (z_1, z_2) \in [-90,90]^2$ ,  $x_j \in [-1,5]$  for  $j = 1, \dots, 8$ .

### Exemple de sorties de la fonction Campbell2D



## Cas analytique : contexte

- **Échantillon d'apprentissage** de Campbell2D de taille  $n = 200$ .
- **Plan d'expérience** : échantillonnage par hypercube latin avec critère maximin.
- **Échantillon test** de taille  $n_{test} = 1000$ .
- **Discrétisation** de  $\mathcal{Z} = [-90, 90]^2$  sur une grille de dimension  $64 \times 64$ .

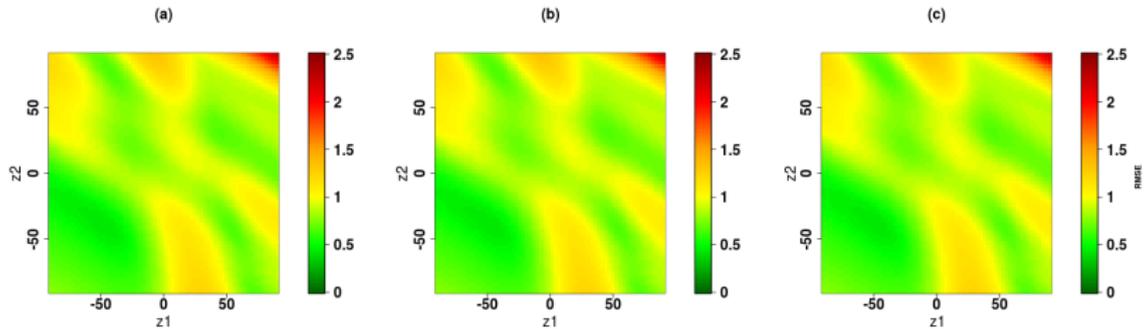
## Calibration de l'ACPF

- $K$  (Taille de la base  $\Phi(\mathbf{z}) = (\phi_1(\mathbf{z}), \dots, \phi_K(\mathbf{z}))^\top$ ) est choisie tel que la racine de l'erreur quadratique moyenne (RMSE) des décompositions des  $(y_{\mathbf{x}_i}(\mathbf{z}))_{i=1, \dots, n}$  soit minimisée.
  - Pour les ondelettes :  $K = 64 \times 64 = 4\ 096$  (profondeur de la résolution  $J = 1$ )
  - Pour les B-splines :  $K = 35 \times 35 = 1\ 225$
- Calibration de  $\tilde{K}$  et  $n_{PC}$  par validation croisée à  $k$  blocs (avec  $k = 10$ ) sur l'échantillon d'apprentissage de taille  $n = 200$ .
  - Pour les trois méthodes :  $n_{PC} = 5$  (98% de l'inertie totale).
  - Pour GP<sup>FPCA</sup><sub>wavelet</sub> :  $\tilde{K} = 1200$ .
  - Pour GP<sup>FPCA</sup><sub>B-splines</sub> :  $\tilde{K} = 1225 (= K)$ .

## Calibration de l'ACPF

- $K$  (Taille de la base  $\Phi(\mathbf{z}) = (\phi_1(\mathbf{z}), \dots, \phi_K(\mathbf{z}))^\top$ ) est choisie tel que la racine de l'erreur quadratique moyenne (RMSE) des décompositions des  $(y_{\mathbf{x}_i}(\mathbf{z}))_{i=1, \dots, n}$  soit minimisée.
  - Pour les ondelettes :  $K = 64 \times 64 = 4\ 096$  (profondeur de la résolution  $J = 1$ )
  - Pour les B-splines :  $K = 35 \times 35 = 1\ 225$
- Calibration de  $\tilde{K}$  et  $n_{PC}$  par validation croisée à  $k$  blocs (avec  $k = 10$ ) sur l'échantillon d'apprentissage de taille  $n = 200$ .
  - Pour les trois méthodes :  $n_{PC} = 5$  (98% de l'inertie totale).
  - Pour GP<sup>FPCA</sup><sub>wavelet</sub> :  $\tilde{K} = 1200$ .
  - Pour GP<sup>FPCA</sup><sub>B-splines</sub> :  $\tilde{K} = 1225 (= K)$ .

## Cartes RMSE obtenues avec $GP_{\text{wavelet}}^{FPCA}$ (a), $GP_{B\text{-splines}}^{FPCA}$ (b), et $GP^{PCA}$ (c)



Les trois méthodes sont aussi efficaces. Cependant, la **méthode ACPF** réduit le nombre de variables pour l'ACP à :

- $GP_{\text{wavelet}}^{FPCA}$  : **70.7%**.
- $GP_{B\text{-splines}}^{FPCA}$  : **70%** .

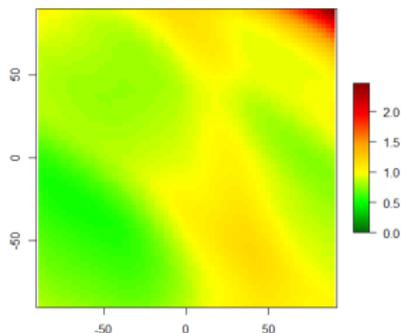
## Coefficient de prédictivité $Q^2$

$$Q^2 = 1 - \frac{\mathbb{E}_{\mathbf{z}} \left\{ \mathbb{E}_{\mathbf{X}} \left[ (Y_{\mathbf{X}}(\mathbf{z}) - \hat{Y}_{\mathbf{X}}(\mathbf{z}))^2 \right] \right\}}{\mathbb{E} [\text{Var} [Y_{\mathbf{X}}(\mathbf{z})]]}.$$

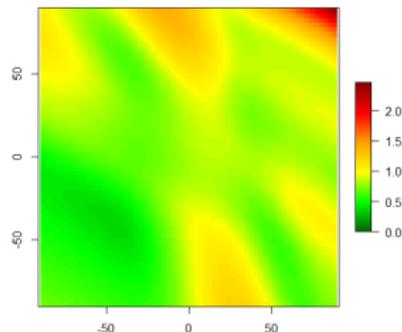
- $\text{GP}_{\text{wavelet}}^{\text{FPCA}}$ ,  $\text{GP}_{\text{B-splines}}^{\text{FPCA}}$ , et  $\text{GP}^{\text{PCA}}$  :  $Q^2 = 96.6\%$  ( $n=200$ ).
- [Marrel et al., 2010] :  $Q^2 = 92.9\%$  ( $n = 200$ ),  $Q^2 = 96.6\%$  ( $n = 500$ )
  - Modèle RPG sur les  $\tilde{K} = 30$  premiers coefficients d'ondelettes, choisis selon leurs variances :  $\lambda_k = \text{Var}[\alpha(\mathbf{X})]$ .
  - Les autres coefficients sont estimés par régression linéaire ou par la moyenne empirique.

# Comparaison $GP_{B-splines}^{Lasso}$ et $GP_{B-splines}^{FPCA}$

RMSE FPCA with Lasso-B-splines  
( $\lambda = 0.01$ )



RMSE with orthonormalization



## Comparaison des cartes RMSE obtenues avec $GP_{B-splines}^{FPCA}$ (en bas) et $GP_{B-splines}^{Lasso}$ (en haut)

- Paramètres estimés par validation à 10 blocs :  
 $\lambda = 0.01$ ,  $\tilde{K} = K = 1\ 225$ ,  $n_{PC} = 5$ .
- $GP_{B-splines}^{Lasso}$  est moins performante en temps de calcul (supérieur à 1 min, comparé à moins 1s pour  $GP_{B-splines}^{FPCA}$ ) et en précision.
- $GP_{B-splines}^{Lasso}$  n'est donc pas appliquée aux modèles d'aléa.

## Cas d'étude : paramètres du forçage marin

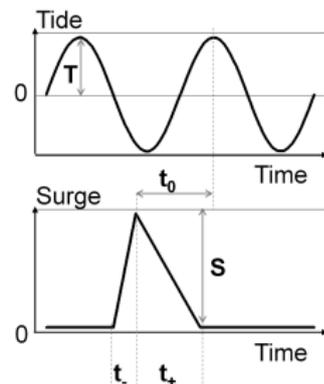
- **(a) Site d'étude** : les Boucholeurs (côte atlantique française).
- Inondation par **débordement**
- **(b) Entrées du modèle** : 5 paramètres du forçage marin (scalaires).

Amplitude de la marée	$T \in [0.95m, 3.70m]$
Pic de surcote	$S \in [0.65m, 2m]$
Différence de phase entre le pic de surcote et l'amplitude de la marée	$t_0 \in [-6h, 6h]$
Durée d'augmentation de la surcote	$t_+ \in [0.5h, 12h]$
Durée de la diminution de la surcote	$t_- \in [0.5h, 12h]$

(a)



(b)



## Cas d'étude : paramètres du forçage marin

Observation de 500 simulations des modèles d'aléa du BRGM et de la CCR

- Plan d'expérience : séquence aléatoire de Sobol.
- **Dimension des cartes** estimées :  $D_{\mathcal{Z}} = 256 \times 256 = 65\,536$ .
- **Échantillon d'apprentissage** de taille  $n = 400$ .
- **Échantillon test** de taille  $n_{test} = 100$ .

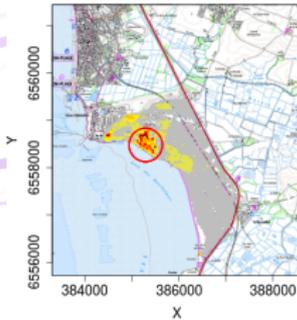
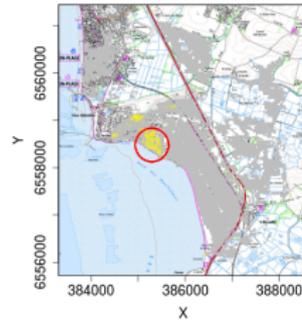
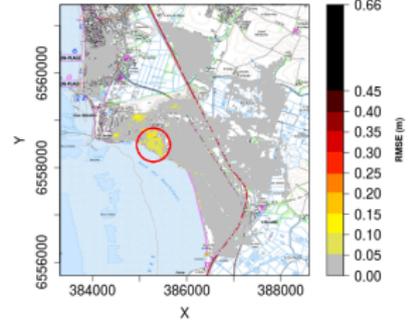
## Paramètres de $GP^{PCA}$ , $GP_{wavelet}^{FPCA}$ , et $GP_{B-splines}^{FPCA}$

- **Calibration de la décomposition :**
  - **Ondelettes** : profondeur de la décomposition  $J = 1$  ( $K = D_{\mathcal{X}}$ ).
  - **B-splines** :  $K = 100 \times 100 = 10\ 000$ .
- **Calibraion de l'ACPF :**
  - Validation croisée à 10 blocs.
  - Pour les trois méthodes :  $n_{PC} = 6$  ( $\approx 96\%$  de l'inertie totale).
  - Pour  $GP_{wavelet}^{FPCA}$  et  $GP_{B-splines}^{FPCA}$  :  
 $\tilde{K}$  tel que  $\sum_{k=1}^{\tilde{K}} \lambda_{(k)} \leq p$ , avec  $p = 0.99$ .

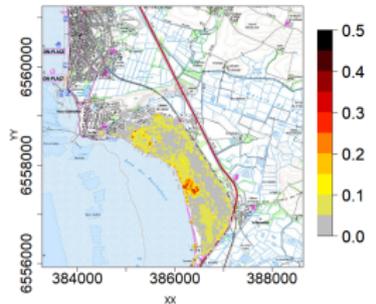
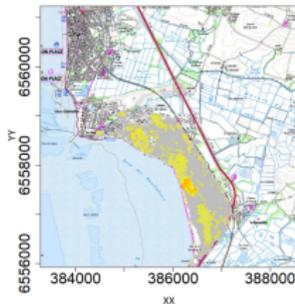
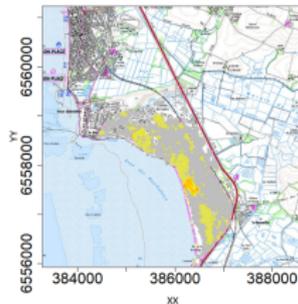
	$GP_{wavelet}^{FPCA}$	$GP_{B-splines}^{FPCA}$
BRGM	$\tilde{K} = 4\ 000$	$\tilde{K} = 1\ 700$
CCR	$\tilde{K} = 1\ 442$	$\tilde{K} = 607$

## Cartes RMSE (en m)

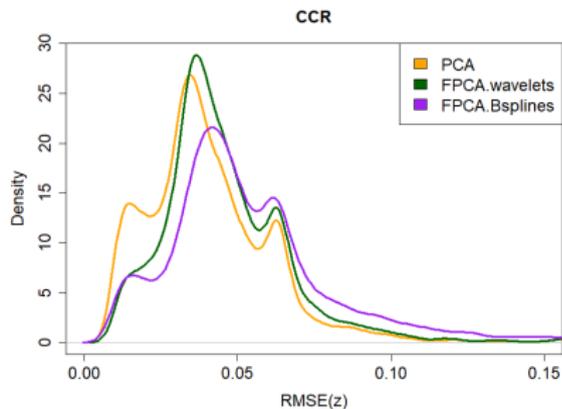
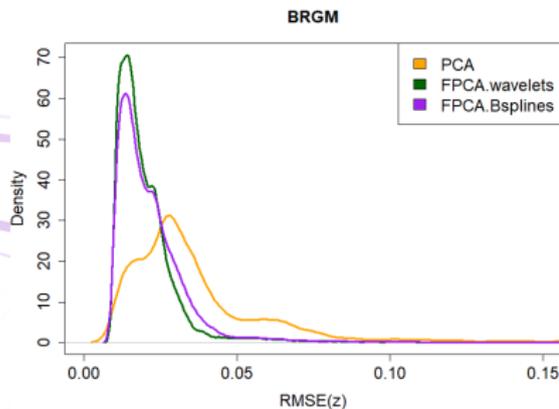
BRGM

GP<sup>PCA</sup>GP<sup>FPCA</sup>  
waveletGP<sup>FPCA</sup>  
B-splines

CCR



## Densité de probabilité, estimée par lissage



- **BRGM** :  $GP^{FPCA}$  *wavelet* est la méthode la plus précise.
- **CCR** :  $GP^{PCA} \approx GP^{FPCA}$  *wavelet*.  
 $GP^{FPCA}$  *B-splines* est la méthode la moins précise.
- Suite des résultats :  $GP^{FPCA}$  *wavelet*.

	BRGM	CCR
$GP^{PCA}$	33 s	19.2s
$GP^{FPCA}$ <i>wavelet</i>	1.5s	1s
$GP^{FPCA}$ <i>B-splines</i>	0.7s	0.5s

## Package R : GpOutput2D

Le package **GpOutput2D** a été développé et partagé sur GitHub (voir <https://github.com/tranvivieli/GpOutput2D>).

Nom de la fonction	Description
Fpca2d	Appliquer une ACP fonctionnelle.
km_Fpca2d, gp_Fpca2d	Construire des modèles de krigeage sur chaque score obtenu avec la fonction <b>Fpca2d</b> (avec le package <b>DiceKriging</b> ou <b>kergp</b> ).
predict	Prédiction de la sortie de la fonction objectif.

**Table 1:** Fonctions principales du package GpOutput2D.

# **Analyse de sensibilité des modèles d'aléa**

---

## 2. Analyse de sensibilité des modèles d'aléa

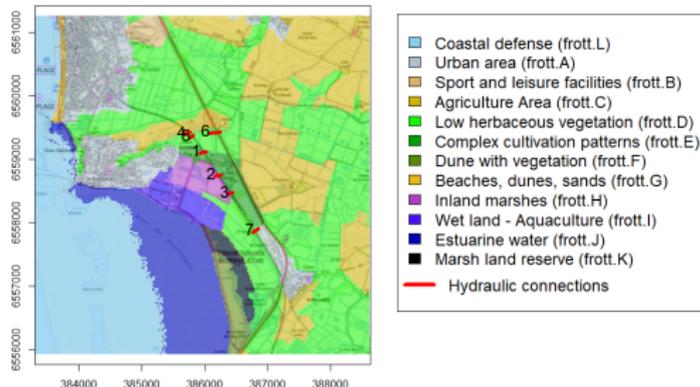
### 2.1 Cas d'étude

### 2.2 Analyse de sensibilité des composantes principales

### 2.3 Indices de sensibilité généralisés

## Contexte du cas d'étude

- Site d'étude : les Boucholeurs, lors de la tempête Xynthia en 2010.
- Entrées des modèles d'aléa :
  - 8 variables catégorielles binaires :
    - Forçage marin : estimation par le "BRGM" ou la "CCR".
    - 7 connexions hydrauliques : "ouvertes" ou "fermées"
  - 12 variables continues : coefficients de rugosité, notées frott.A....frott.L.

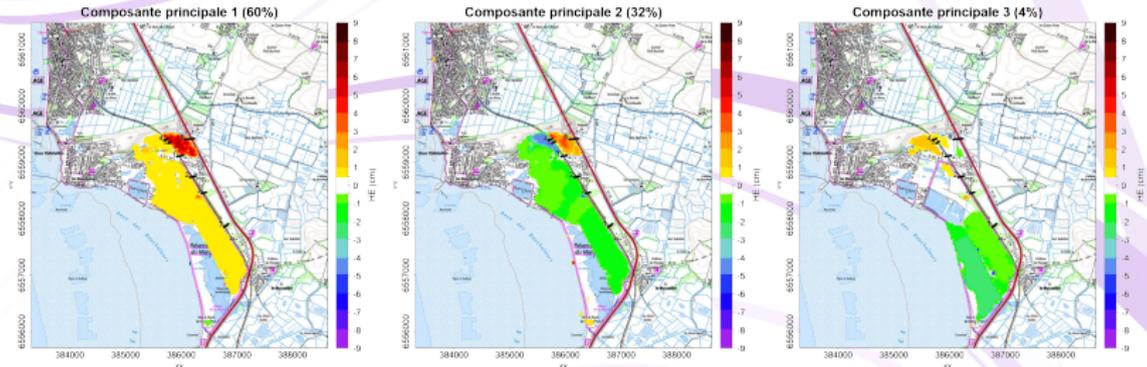


## Application au cas d'étude

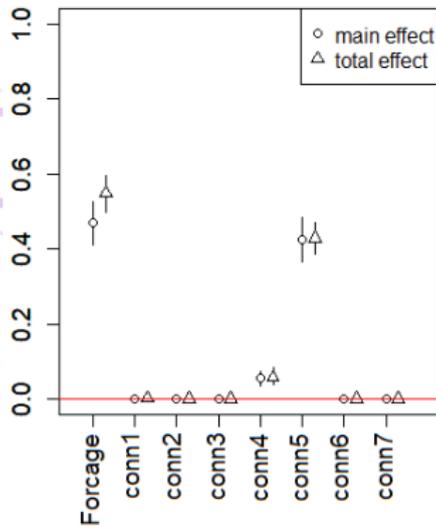
- Analyse de sensibilité des composantes principales : modèle BRGM.
- Indices généralisés : modèles BRGM et CCR.
- Simulations obtenues avec  $GP_{wavelet}^{FPCA}$   
krigeage avec variables continues et catégorielles en entrée :  
package R "kergp" [Roustant et al., 2020].
- Plan d'expérience (PE) de  $GP_{wavelet}^{FPCA}$  :
  - taille **n = 768**.
  - Chacune des  $2^8 = 256$  combinaisons possibles entre les 8 variables catégorielles apparaît 3 fois.
  - PE complété avec un plan hypercube latin des 12 variables continues (rugosité).
- Indices de Sobol du 1<sup>re</sup> ordre et total : calculés à partir d'échantillons aléatoires Monte Carlo de taille initial **n<sub>0</sub> = 10<sup>3</sup>** [Saltelli, 2002].

# ACP : Indices de Sobol

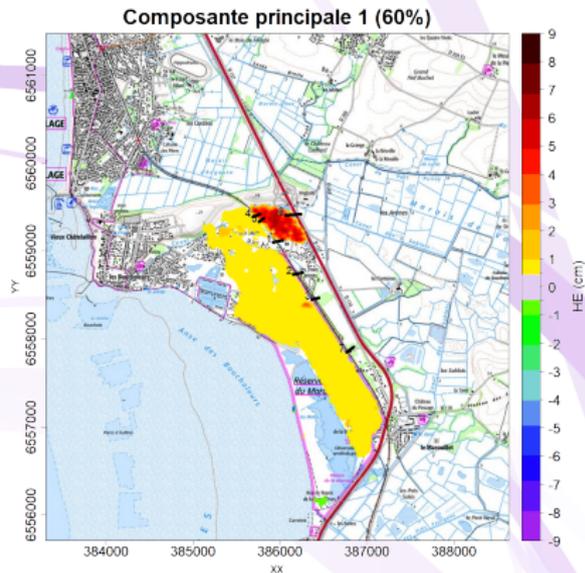
Les indices de Sobol de chaque score de l'ACP sont calculés.



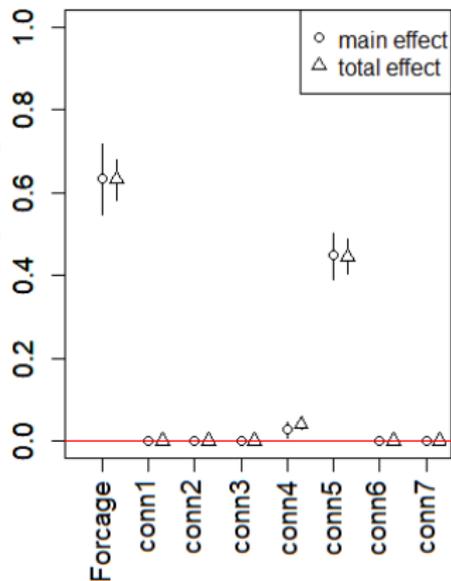
**Figure 6:** Les  $n_{PC} = 3$  premières composantes principales (96% de variance expliquée), obtenues pour le modèle **BRGM**.

1<sup>ère</sup> composante principale (60%)

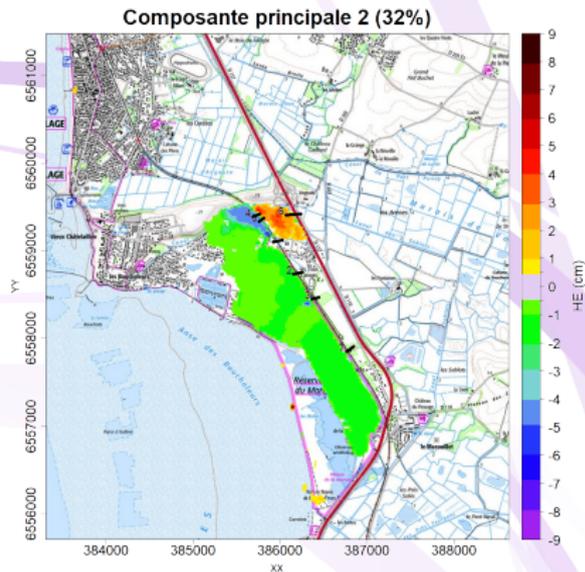
**Figure 7:** Indices de Sobol du score de la 1<sup>ère</sup> composante principale. Les indices des coefficients de rugosité sont égaux à 0.

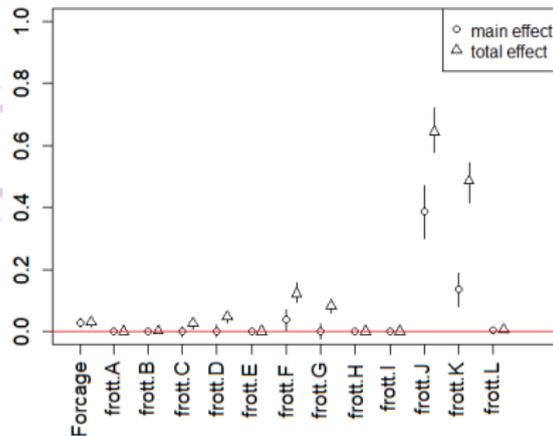


## 2<sup>nd</sup> composante principale (32%)



**Figure 8:** Indices de Sobolj du score de la 2<sup>nd</sup> composante principale. Les indices des coefficients de rugosité sont égaux à 0.



3<sup>ème</sup> composante principale (4%)

Notation	Type de sol
frott.A	Zone urbaine
frott.B	Installations sportives et de loisirs
frott.C	Zone agricole
frott.D	Végétation herbacée basse
frott.E	Modèles de culture complexes
frott.F	Dune avec végétation
frott.G	Plages, dunes, sables
frott.H	Marais intérieurs
frott.I	Terres humides - Aquaculture
frott.J	Eau estuarienne
frott.K	Réserve marécageuse
frott.L	Défense côtière

**Figure 9:** Indices de Sobol du score de la 3<sup>ème</sup> composante principale. Les indices des connexions hydrauliques sont égaux à 0.

## Calcul des indices de sensibilité généralisés

- Avec l'AS sur les composantes principales : influence des entrées sur la structure spatiale de la sortie.
- Difficile de faire une hiérarchie bien définie entre les variables d'entrée.
- Avec les indices de Sobol calculés sur chaque composante principale, des **indices de sensibilité généralisés** peuvent être calculés.
- **Indice entre 0 et 1**, résumant l'influence des entrées sur l'ensemble de la sortie spatiale.

## Indice généralisé

$GSI_{\omega}$  : indice de sensibilité généralisé de  $y_{\mathbf{x}}(\mathbf{z})$  par rapport à  $\mathbf{x}_{\omega}$ , avec  $\omega \subseteq \{1, \dots, d\}$ .

### Propriété [Lamboni et al., 2011]

Pour tout  $\omega \subseteq \{1, \dots, d\}$ ,  $GSI_{\omega}$  peut être calculé de la manière suivante :

$$GSI_{\omega} = \frac{\sum_{l=1}^{n_{PC}} \lambda_l SI_{\omega,l}}{\sum_{l=1}^{n_{PC}} \lambda_l}$$

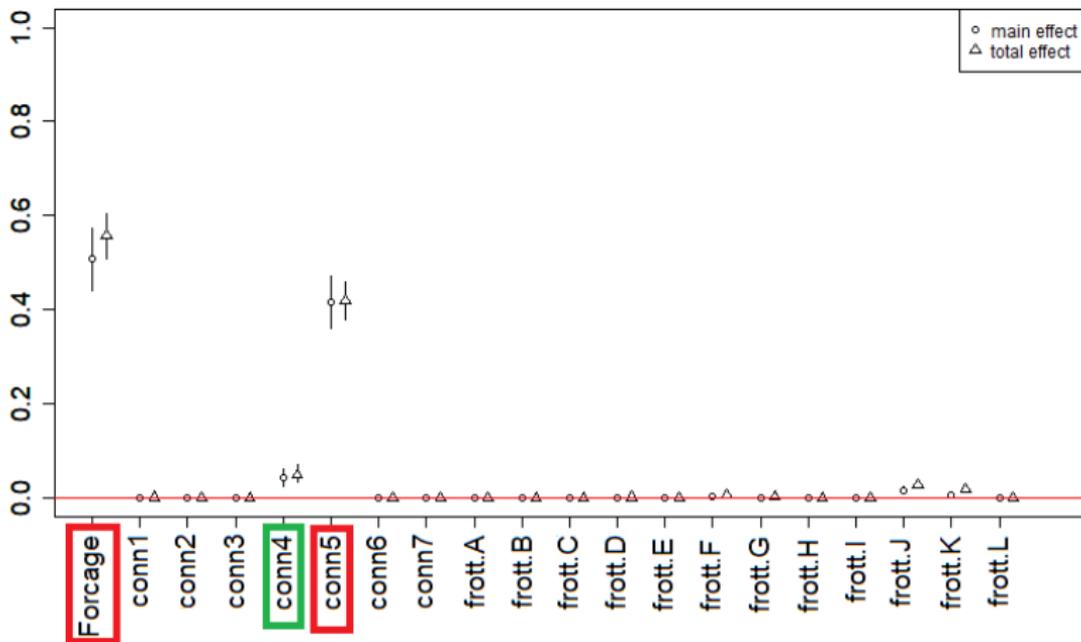
où  $\lambda_l$  est la  $l$ -ème valeur propre,  $SI_{\omega,l}$  est l'indice de Sobol de la  $l$ -ème composante principale.

Formule pour une base de fonctions quelconque :

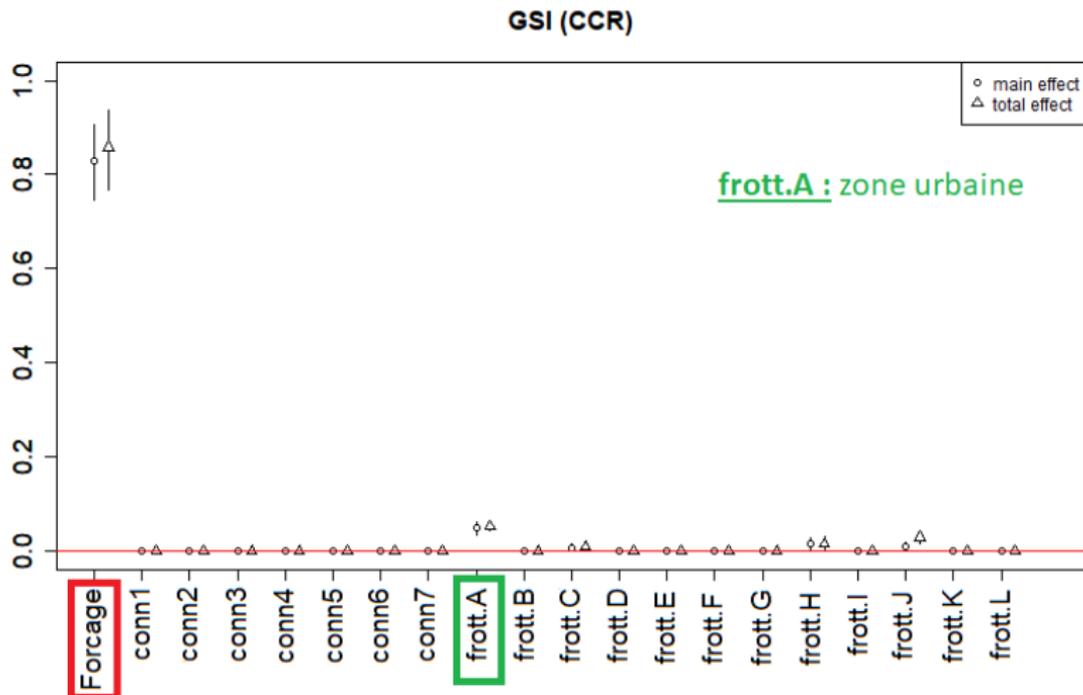
[Perrin et al., 2021] (article : *Reliability Engineering and System Safety*)

[Perrin, 2021] (Manuscrit de thèse)

# Indices généralisés du modèle BRGM

**GSI (BRGM)**

# Indices généralisés du modèle CCR



## Conclusions et perspectives

---

# Conclusions

- **Motivation** : Analyse de sensibilité de modèle de submersion marine.
- Méta-modélisation et l'analyse de sensibilité de modèles avec **sorties spatiales** :
  - à grande dimension
  - incluant de fortes discontinuités.

## En résumé ...

- **Méta-modélisation** : étape préliminaire de **sélection des coefficients** ajoutée à l'ACPF.
  - Sur une base orthonormée : **décomposition de l'énergie** (*méthode choisie pour les modèles d'aléa*).
  - Sur une base quelconque : **régression Lasso**.
  - Développement d'un **package R « GpOutput2D »**
- **Analyse de sensibilité** :
  - **AS avec les composantes principales** : Influence des entrées sur la structure spatiale de la sortie.
  - **Indice généralisé** : extension de [Lamboni et al., 2011] pour toute base de fonctions.
- Premiers résultats proposant une perspective sur la **combinaison des modèles BRGM et CCR** [Perrin, 2021].

## Perspectives

- Difficulté à prédire les **cartes non inondées** (absence d'inondation).  
**Proposition** : méthodes de classification [Rohmer et al., 2018], modèle de krigeage sous contraintes [Lopera, 2019]...  
*Intérêt* : *prédiction d'absence de risques (système d'alerte)*
- **Autre mesure d'incertitude** que la variance.  
**Exemple** : mesures de dépendance [Da Veiga, 2015, De Lozzo and Marrel, 2017].
- Pour l'AS, **prise en compte de l'erreur de prédiction dans les composantes principales** de l'ACPF.  
**Proposition** : formulation probabiliste de l'ACP [Tipping and Bishop, 1999, Bishop, 1999].
- Résultats prometteurs : **la combinaison des modèles**.

The background of the slide features several overlapping, wavy, light purple lines that create a sense of movement and depth. The lines vary in thickness and opacity, with some appearing as thin, delicate strokes and others as thicker, more prominent bands. The overall effect is a soft, abstract, and modern aesthetic.

**Fin de la présentation**

**Merci pour votre attention**

# Annexes

---

## Loi de prédiction

En suivant les étapes de

l'algorithme précédent, on a :

$$Y_{\mathbf{x}^*}(\mathbf{z}) \sim \mathcal{N}(\hat{y}_{\mathbf{x}^*}(\mathbf{z}), \sigma_{y_{\mathbf{x}^*}}^2(\mathbf{z})),$$

avec :

- $\hat{y}_{\mathbf{x}^*}(\mathbf{z}) = \phi_{\tilde{K}}(\mathbf{z})^\top \mathbf{W} \hat{\mathbf{t}}(\mathbf{x}^*) + \bar{\alpha}_{-\tilde{K}}^\top \phi_{-\tilde{K}}(\mathbf{z})$
- $\sigma_{y_{\mathbf{x}^*}}^2(\mathbf{z}) = \phi_{\tilde{K}}(\mathbf{z})^\top \mathbf{W} \Sigma(\mathbf{x}^*) \mathbf{W}^\top \phi_{\tilde{K}}(\mathbf{z})$

## Notations

- On note  
 $\bar{\alpha}_{(k)} = \mathbb{E}[\alpha_{(k)}(\mathbf{X})] \approx \frac{1}{n} \sum_{i=1}^n \alpha_{(k)}(\mathbf{x}_i)$ .
- $\bar{\alpha}_{-\tilde{K}} = (\bar{\alpha}_{(\tilde{K}+1)}, \dots, \bar{\alpha}_{(K)})^\top$ ,  
 $\phi_{\tilde{K}} = (\phi_{(1)}, \dots, \phi_{(\tilde{K})})^\top$  et  
 $\phi_{-\tilde{K}}(\mathbf{z}) = (\phi_{(\tilde{K}+1)}(\mathbf{z}), \dots, \phi_{(K)}(\mathbf{z}))^\top$ .
- Matrice des vecteurs propres :  
 $\mathbf{W} = (\omega_1, \dots, \omega_{n_{PC}})$ .
- Vecteur des fonctions moyenne de RPG :  
 $\hat{\mathbf{t}}(\mathbf{x}^*) = (\hat{t}_1(\mathbf{x}^*), \dots, \hat{t}_{n_{PC}}(\mathbf{x}^*))^\top$ .
- $\sigma_l^2(\mathbf{x}^*)$  est la variance prédictive du score  
 $t_l(\mathbf{x}^*)$ ,  $l = 1, \dots, n_{PC}$ .
- Matrice de covariance  
 $\Sigma(\mathbf{x}^*) = \text{diag}(\sigma_1^2(\mathbf{x}^*), \dots, \sigma_{n_{PC}}^2(\mathbf{x}^*))$ .

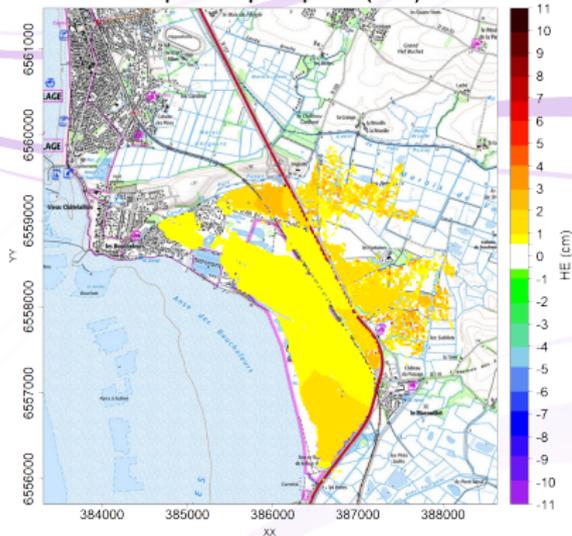
## Campbell2D : influence de la taille de l'échantillon

		<i>n</i>				
		50	100	150	200	500
Méthodes	$GP^{PCA}$	69.4	85.9	93	96.6	97.6
	$GP^{FPCA}_{wavelet}$	79	93.3	96.1	96.6	97.6
	$GP^{FPCA}_{B-splines}$	87.9	93.6	96.1	96.6	97.6

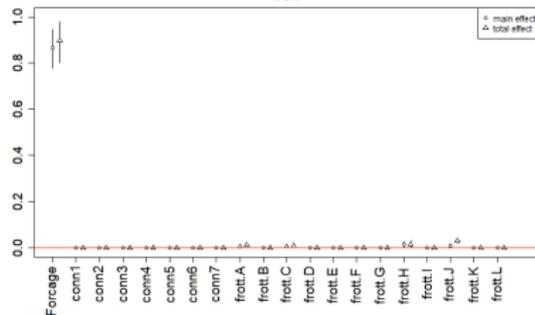
**Table 2:**  $Q^2$  global (en %) de  $GP^{PCA}$ ,  $GP^{FPCA}_{wavelet}$  et  $GP^{FPCA}_{B-splines}$ , pour différentes tailles d'échantillons d'apprentissage ( $n$ )

# Analyse de sensibilité du modèle CCR (Xynthia, 2010)

Composante principale 1 (88%)

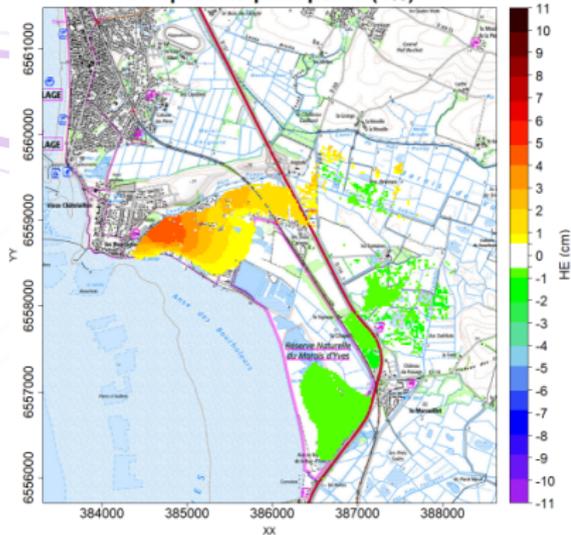


Composante principale 1 (88%)  
CCR

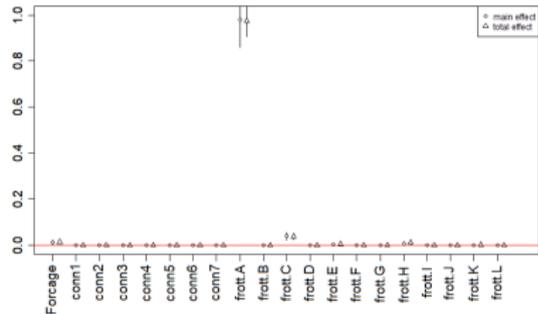


# Analyse de sensibilité du modèle CCR (Xynthia, 2010)

Composante principale 2 (4%)

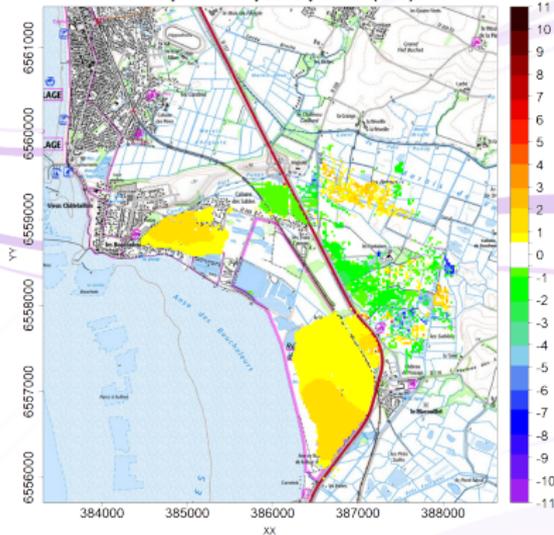


Composante principale 2 (4%)  
CCR

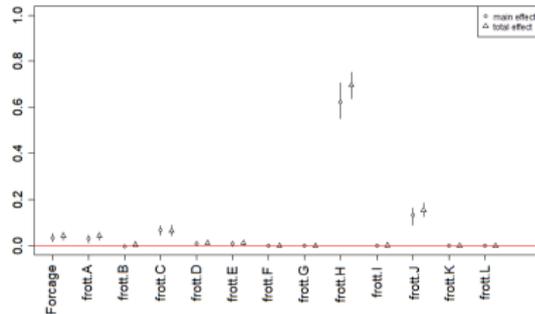


# Analyse de sensibilité du modèle CCR (Xynthia, 2010)

Composante principale 4 (2%)



Composante principale 4 (2%)  
CCR



L'influence de *frott.H* et *frott.J* n'est pas interprétée avec l'indice généralisé.

*frott.H* : Marais intérieurs | *frott.J* : Eaux estuariennes

## Combinaison des modèles : Méthode testée

- **Concaténer les simulations des deux modèles** et ajouter une variable catégorielle en entrée, qui indique quel modèle est utilisé :  $u \in \{BRGM, CCR\}$ .
- On considère le modèle suivant :

$$\begin{aligned} f &: \{BRGM, CCR\} \times \mathcal{X} &\rightarrow \mathbb{L}^2(\mathcal{Z}) \\ \omega = (u, \mathbf{x}) &&\mapsto y_\omega(\mathbf{z}) \end{aligned}$$

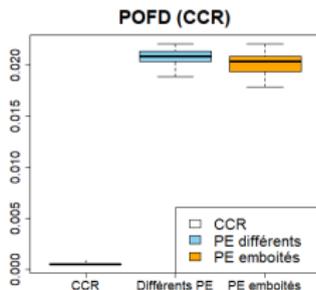
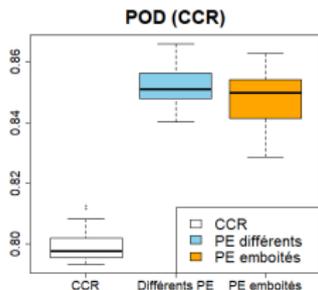
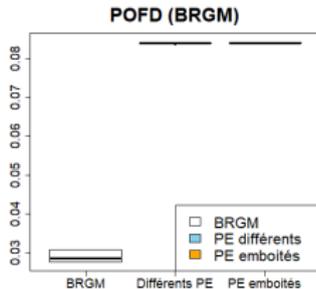
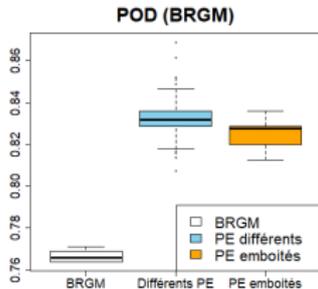
avec  $\mathcal{X}$  l'espace des entrées des deux modèles.

- Pour  $\mathbf{x}$ , **les même entrées que pour l'analyse de sensibilité** sont considérées.
- **Application de**  $GP_{wavelet}^{FPCA}$ .

## Combinaison des modèles : Plans d'expériences (PE)

- Moins de simulations à disposition pour le BRGM que la CCR (temps de calcul plus long).
- **Nombre de simulations :**
  - BRGM : 256.
  - CCR : 1280.
- **Deux configurations de PE** testées :
  - **PE emboîtés** : PE du modèle BRGM est inclus dans celui de la CCR.
  - **PE différents** : les PE des variables continues sont différents.

# Combinaison des modèles : Scores des estimations



Plusieurs indices existent  
[Schaefer, 1990].

- Scores entre 0 et 1.
- POD : probabilité de détection.
- POFD : probabilité de fausse détection.



Bates, P. D., Dawson, R. J., Hall, J. W., Horritt, M. S., Nicholls, R. J., Wicks, J., and Hassan, M. A. A. M. (2005).

**Simplified two-dimensional numerical modelling of coastal flooding and example applications.**

*Coastal Engineering*, 52(9) :793–810.



Bates, P. D., Horritt, M. S., and Fewtrell, T. J. (2010).

**A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling.**

*Journal of Hydrology*, 387(1-2) :33–45.



Bishop, C. M. (1999).

**Bayesian PCA.**

In *Advances in neural information processing systems*, pages 382–388.



Campbell, K., McKay, M. D., and Williams, B. J. (2006).

**Sensitivity analysis when model outputs are functions.**

*Reliability Engineering & System Safety*,  
91(10-11) :1468–1472.



Chen, T., Hadinoto, K., Yan, W., and Ma, Y. (2011).

**Efficient meta-modelling of complex process simulations with time–space-dependent outputs.**

*Computers & chemical engineering*, 35(3) :502–509.



Da Veiga, S. (2015).

**Global sensitivity analysis with dependence measures.**

*Journal of Statistical Computation and Simulation*,  
85(7) :1283–1305.



De Lozzo, M. and Marrel, A. (2017).

**Sensitivity analysis with dependence and variance-based measures for spatio-temporal numerical simulators.**

*Stochastic environmental research and risk assessment*,  
31(6) :1437–1453.



FFSA, G. (2011).

**La tempête Xynthia du 28 février 2010. Bilan chiffré au 31 décembre 2010.**

Technical report, Association française de l'assurance.



Lamboni, M., Monod, H., and Makowski, D. (2011).

**Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models.**

*Reliability Engineering & System Safety*, 96(4) :450–459.



Lazure, P. and Dumas, F. (2008).

**An external–internal mode coupling for a 3d hydrodynamical model for applications at regional scale (mars).**

*Advances in water resources*, 31(2) :233–250.



Lopera, A. F. L. (2019).

***Gaussian Process Modelling under Inequality Constraints.***

PhD thesis, Université de Lyon.

-  Marrel, A., Iooss, B., Jullien, M., Laurent, B., and Volkova, E. (2010).

**Global sensitivity analysis for models with spatially dependent outputs.**

*Environmetrics*, 22(3) :383–397.

-  Marrel, A., Perot, N., and Mottet, C. (2015).

**Development of a surrogate model and sensitivity analysis for spatio-temporal numerical simulators.**

*Stochastic environmental research and risk assessment*, 29(3) :959–974.



Naulin, J., Moncoulon, D., Le Roy, S., Pedreros, R., Idier, D., and Oliveros, C. (2015).

**Estimation of insurance related losses resulting from coastal flooding in france.**

*Natural Hazards and Earth System Sciences Discussions*, 3(4) :2811–2846.



Perrin, T., Roustant, O., Rohmer, J., Alata, O., Naulin, J., Idier, D., Pedreros, R., Moncoulon, D., and Tinard, P. (2021).

**Functional principal component analysis for global sensitivity analysis of model with spatial output.**

*Reliability Engineering & System Safety*, 211 :107522.



Perrin, T. V.-V. E. (2021).

***Méta-modélisation et analyse de sensibilité pour les modèles avec sortie spatiale. Application aux modèles de submersion marine.***

PhD thesis, Université de Lyon.



Ramsay, J. O. (2006).

***Functional data analysis.***

Wiley Online Library.



Rasmussen, C. E. and Williams, C. K. (2006).

***Gaussian processes for machine learning, volume 2.***

MIT press Cambridge, MA.

-  Rohmer, J., Idier, D., Paris, F., Pedreros, R., and Louisor, J. (2018).

**Casting light on forcing and breaching scenarios that lead to marine inundation : Combining numerical simulations with a random-forest classification approach.**

*Environmental Modelling & Software*, 104 :64–80.

-  Roustant, O., Padonou, E., Deville, Y., Clément, A., Perrin, G., Giorla, J., and Wynn, H. (2020).

**Group kernels for Gaussian process metamodels with categorical inputs.**

*SIAM/ASA Journal on Uncertainty Quantification*, 8(2) :775–806.



Saltelli, A. (2002).

**Making best use of model evaluations to compute sensitivity indices.**

*Computer physics communications*, 145(2) :280–297.



Schaefer, J. T. (1990).

**The critical success index as an indicator of warning skill.**

*Weather and forecasting*, 5(4) :570–575.



Sobol, I. M. (1993).

**Sensitivity estimates for nonlinear mathematical models.**

*Mathematical modelling and computational experiments*,  
1(4) :407–414.



Tibshirani, R. (1996).

**Regression shrinkage and selection via the lasso.**

*Journal of the Royal Statistical Society : Series B (Methodological)*, 58(1) :267–288.



Tipping, M. E. and Bishop, C. M. (1999).

**Probabilistic principal component analysis.**

*Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 61(3) :611–622.



Xiao, H. and Li, L. (2016).

**Discussion of paper by matieyendou lamboni, hervé monod, david makowski “multivariate sensitivity analysis to measure global contribution of input factors in dynamic models”, *reliab. eng. syst. saf.* 99 (2011) 450–459.**

*Reliability Engineering & System Safety*, 147 :194–195.