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

José Betancourt

Advisor: Prof. Thierry Klein
Co-advisor: Prof. François Bachoc

PhD thesis defense - Applied Mathematics
June 08, 2020

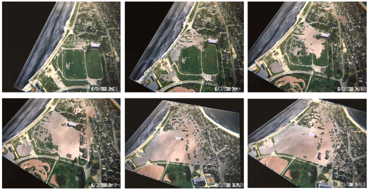
Outline

- Problem description
- Chapter 2. First metamodeling experiences
- Chapter 3. ACO based structural optimization
- Chapter 4. Consolidation of funGp R package
- Chapter 5. Structural optimization in RISCOPE
- Chapter 6. Asymptotics of trans-Gaussian Processes
- Perspectives and contributions

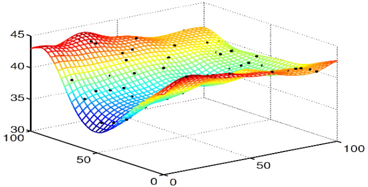
Problem description

RISCOPE project framework: Coastal flooding early warning

- **Objective:** provide fast-running statistical models to replace the slow-running computer code → **Metamodeling**



hours to days



few seconds

Research problem

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

WW3 → SWASH

Code A3

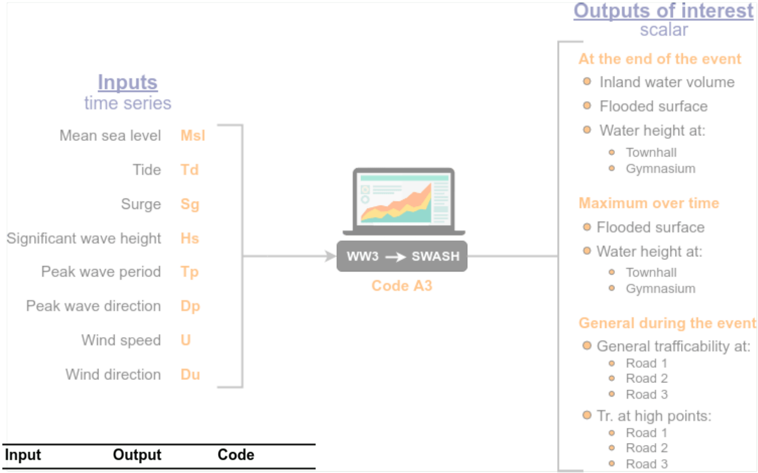
Outputs of interest
scalar

At the end of the event
- Inland water volume
- Flooded surface
- Water height at:
  - Townhall
  - Gymnasium

Maximum over time
- Flooded surface
- Water height at:
  - Townhall
  - Gymnasium

General during the event
- General trafficability at:
  - Road 1
  - Road 2
  - Road 3
- Tr. at high points:
  - Road 1
  - Road 2
  - Road 3

Research problem



Input	Output	Code	
functional	scalar	deterministic	← **This thesis**
scalar	scalar	stochastic	← J. Rohmer (BRGM)
functional	functional	deterministic	← A. Lopez-Lopera (IMT-BRGM)

Questions

Functional inputs

- what functional inputs are worth keeping active?
- what dimension reduction method is ideal?
- what is a suitable projection dimension?

Gaussian process metamodel

- what is a suitable kernel?
- what is a convenient distance to measure similarities between functional input points within the kernel function?

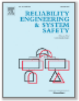
→ We call these, **structural parameters** of the metamodel

Chapter 2. First metamodeling experiences

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]

- First functional-input metamodeling attempts
- Proof of concept for goal-oriented structural optimization
- Development of first R routines

Ch.2 - Simplified numerical code

Inputs
time series

Tide	**Td**
Surge	**Sg**
Significant wave height	**Hs**
Peak wave period	**Tp**

SWAN → EurOtop

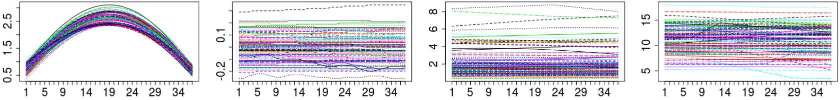
Code A1

Output of interest
scalar

At the end of the event

- Inland water volume

Inputs



37 time instants

- We split each functional input into a scalar and a functional representation, x and f, respectively.

- We observe a deterministic code of the form

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

  - $\boldsymbol{x}$ the vector of d_s scalar inputs $\left(x^{(i)} \in \mathbb{R}, \forall i = 1, \ldots, d_s\right)$
  - $\boldsymbol{f}$ the vector of d_f functional inputs $\left(f^{(i)} : T_i \subset \mathbb{R} \to \mathbb{R}, \forall i = 1, \ldots, d_f\right)$

- We see f_{code} as a realization of a Gaussian process Y

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

  - m the mean function
  - k the covariance function.

1. **Dimension reduction:** project $f^{(i)} : T_i \subset \mathbb{R} \rightarrow \mathbb{R}$ onto a space of lower dimension p_i.

$$f^{(i)}(t) \approx \Pi\left(f^{(i)}\right)(t) = \sum_{r=1}^{p} \alpha_r B_r(t)$$

- **which projection method?**
    - B-splines (Muehlenstaedt, Fruth, & Roustant, 2017)
    - PCA (Lataniotis, Marelli, & Sudret, 2018)
    - Wavelets (Cohen, Daubechies, & Feauveau, 1992)
    - PLS (Papaioannou, Ehre, & Straub, 2019)
- **which dimension?**

2. Covariance function

$$k(Y(\boldsymbol{x},\boldsymbol{f}),Y(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{f}})) = \sigma^2 \; R\left(\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\|_{L^2,\boldsymbol{\theta_s}}\right) R\left(\|\boldsymbol{f}-\tilde{\boldsymbol{f}}\|_{L^2,\boldsymbol{\theta_f}}\right)$$

Distance: measure the similarity between points

- **For scalars:** $\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\|_{L^2,\boldsymbol{\theta_s}} = \sqrt{\sum_{i=1}^{ds} \frac{\left\|x^{(i)}-\tilde{x}^{(i)}\right\|^2}{\left(\theta_s^{(i)}\right)^2}}$ classical

- **For functions:** $\left\|\Pi(\boldsymbol{f})-\Pi(\tilde{\boldsymbol{f}})\right\|_{S,\dot{\boldsymbol{\theta}}_{\boldsymbol{f}}} := \sqrt{\sum_{i=1}^{df}\sum_{r=1}^{p_i} \frac{\left(\alpha_r^{(i)}-\tilde{\alpha}_r^{(i)}\right)^2}{\left(\dot{\theta}_{f,r}^{(i)}\right)^2}}$ Nanty et al. (2016)

$$\left\|\Pi(\boldsymbol{f})-\Pi(\tilde{\boldsymbol{f}})\right\|_{D,\boldsymbol{\theta_f}} := \sqrt{\sum_{i=1}^{df} \frac{\int_{T_i}\left(\sum_{r=1}^{p_i}\left(\alpha_r^{(i)}-\tilde{\alpha}_r^{(i)}\right) B_r^{(i)}(t)\right)^2 dt}{\left(\theta_f^{(i)}\right)^2}}$$

Muehlenstaedt et al. (2017)

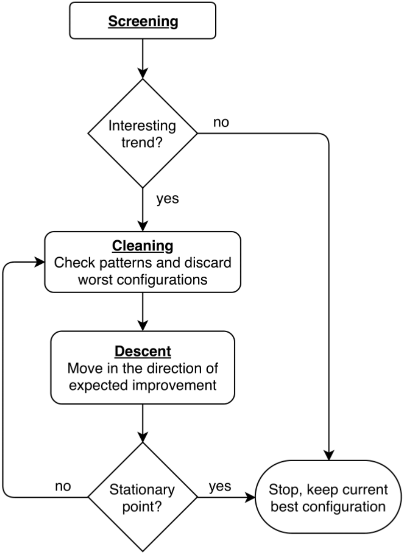
Ch.2 - Ingredients of our Gp metamodel

Features to define:

- State of functional inputs
  - active
  - inactive
- Projection dimension
  - $\{1, \ldots, 37\}$
- Projection method
  - B-splines
  - PCA
- Distance for functions
  - $\|\cdot\|_{S,\dot{\boldsymbol{\theta}}_{\boldsymbol{f}}}$
  - $\|\cdot\|_{D,\boldsymbol{\theta}_{\boldsymbol{f}}}$

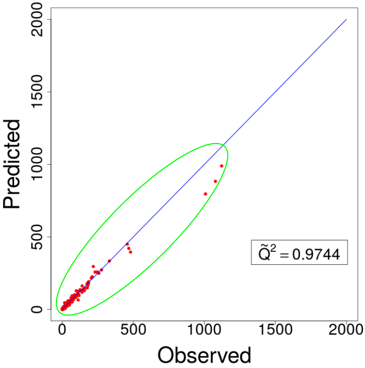
Features fixed:

- State of scalar inputs: active
- Kernel function: Matérn $5/2$

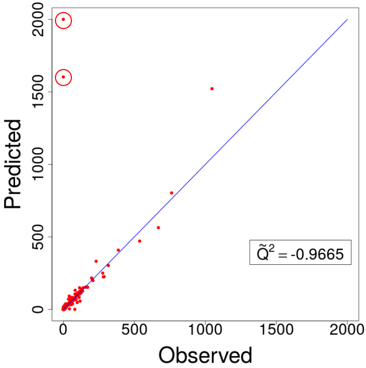


Based on Response Surface Methodology (Montgomery, 1997; Pulido et al., 2012)

Best structural configuration found

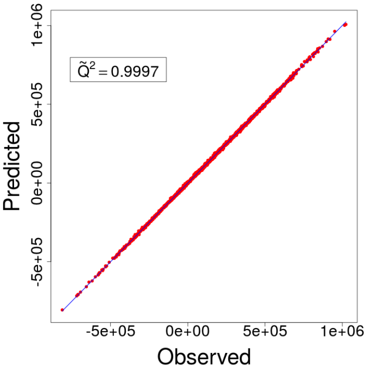


Best sample

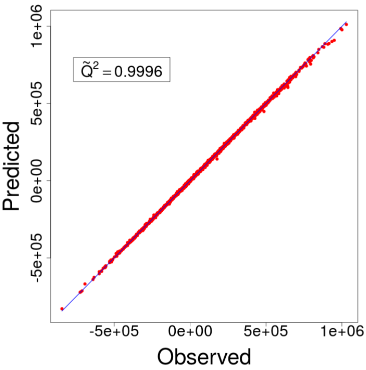


Worst sample

Best structural configuration found

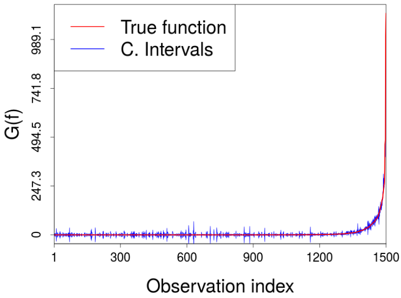


Best sample

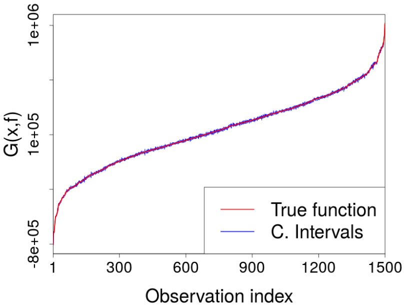


Worst sample

Fit of sorted output



RISCOPE



Analytic

Results

- Outstanding metamodels in all the experiments
- Ideal projection method and dimension varied with the application
- $\|\cdot\|_{D,\boldsymbol{\theta_f}}$ consistently outperformed $\|\cdot\|_{S,\dot{\boldsymbol{\theta}}_{\boldsymbol{f}}}$

Conclusions

- Effective tool for structural optimization
- The optimization must be performed each time
- Intensive exploration of more features and levels might lead to policies
- A more efficient exploration approach would be required

Chapter 3. ACO based structural optimization

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

Technical Report - April 2020

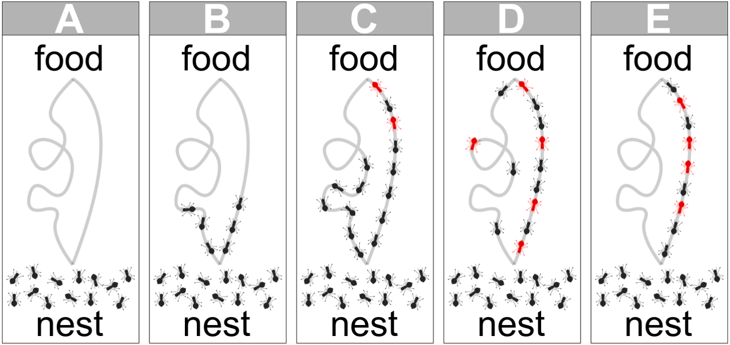
Deliverable: D3.b
Reference: WP3.2

INSTITUT de MATHEMATIQUES de TOULOUSE

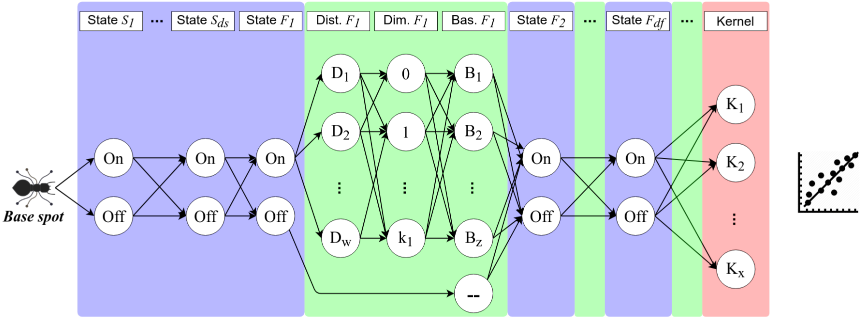
José Betancourt, François Bachoc, Thierry Klein, Fabrice Gamboa

- More efficient exploration of the solution space

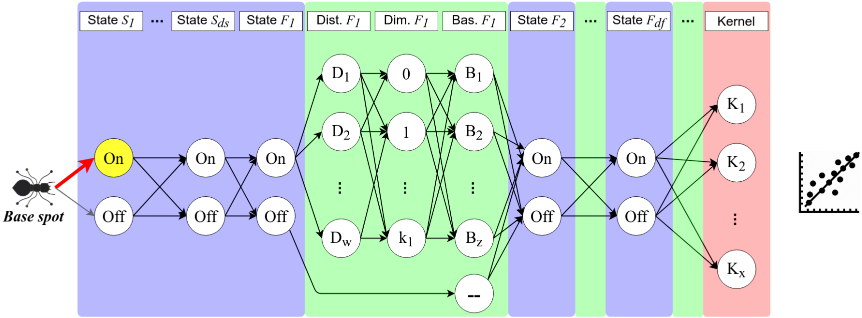
The double bridge experiment (Goss et al., 1989)

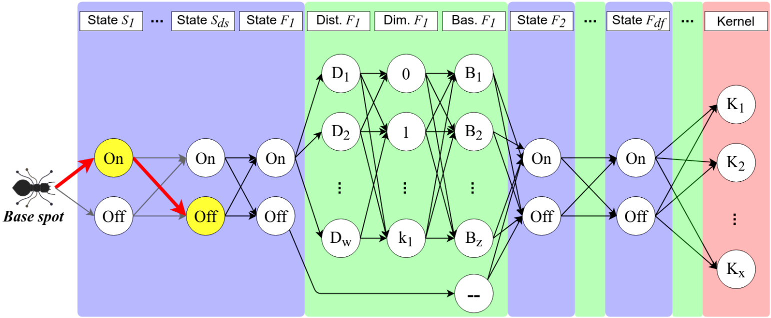


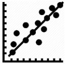
Ch.3 - Decision network for structural optimization

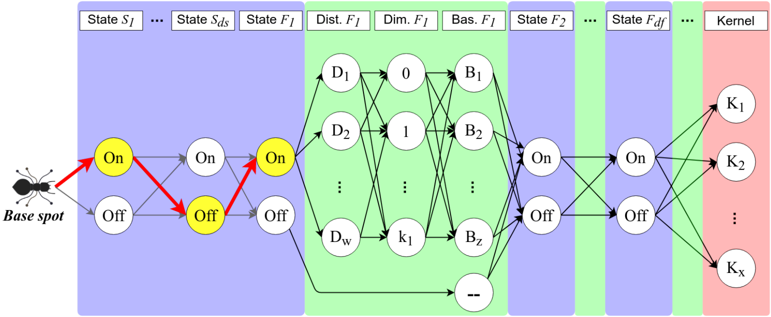


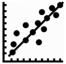
Ch.3 - Decision network for structural optimization

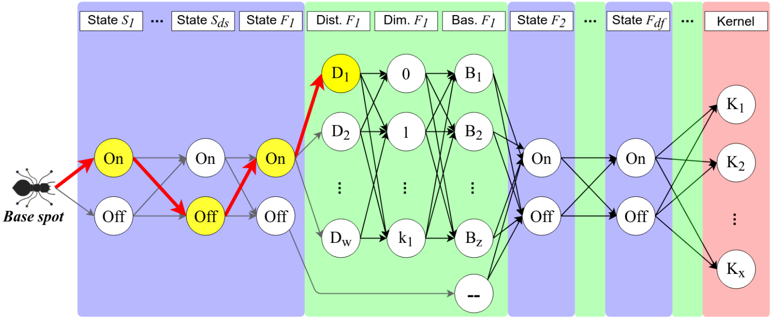


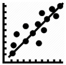

State S_1
...
State S_{ds}
State F_1
Dist. F_1
Dim. F_1
Bas. F_1
State F_2
...
State F_{df}
...
Kernel
Base spot
On
Off
On
Off
On
Off
D_1
D_2
D_w
0
1
k_1
B_1
B_2
B_z
--
On
Off
On
Off
K_1
K_2
K_x

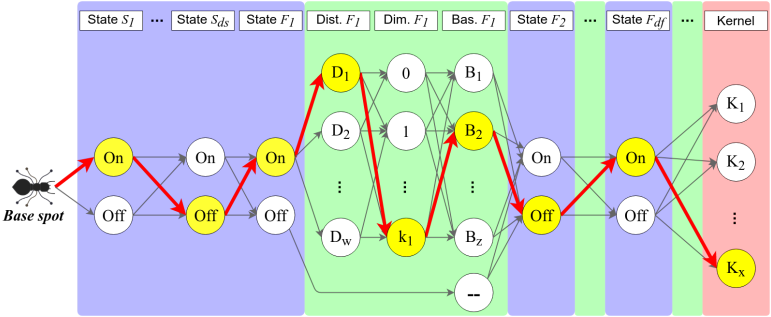


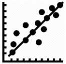

State S_1
...
State S_{ds}
State F_1
Dist. F_1
Dim. F_1
Bas. F_1
State F_2
...
State F_{df}
...
Kernel
Base spot
On
Off
On
Off
On
Off
D_1
D_2
D_w
0
1
k_1
B_1
B_2
B_z
--
On
Off
On
Off
K_1
K_2
K_x

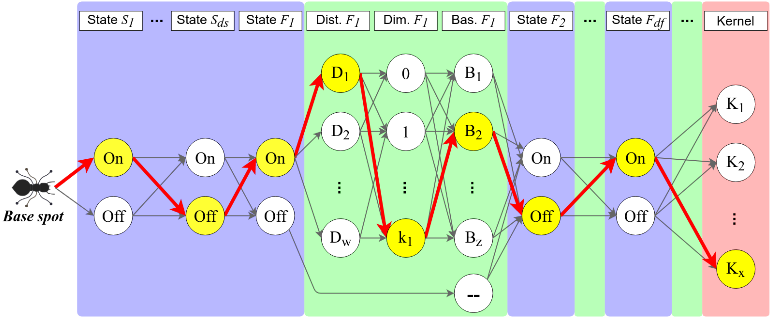


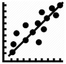

State S_1
...
State S_{ds}
State F_1
Dist. F_1
Dim. F_1
Bas. F_1
State F_2
...
State F_{df}
...
Kernel
On
Off
On
Off
On
Off
D_1
D_2
D_w
0
1
k_1
B_1
B_2
B_z
--
On
Off
On
Off
K_1
K_2
K_x
Base spot




State S_1
...
State S_{ds}
State F_1
Dist. F_1
Dim. F_1
Bas. F_1
State F_2
...
State F_{df}
...
Kernel
Base spot
On
Off
On
Off
On
Off
D_1
D_2
D_w
0
1
k_1
B_1
B_2
B_z
--
On
Off
On
Off
K_1
K_2
K_x

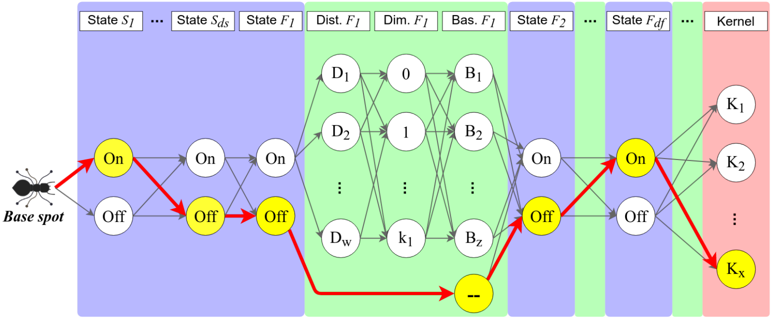


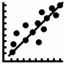




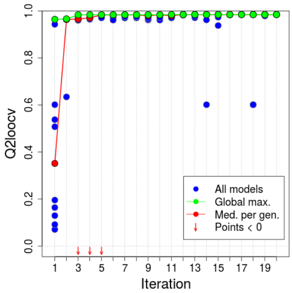
$local$ *pheromone update* $\longrightarrow$

$\longleftarrow$ $global$ *pheromone update*

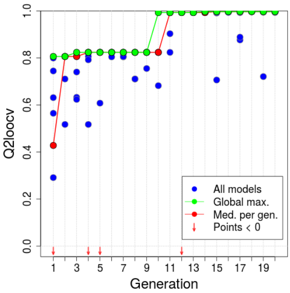




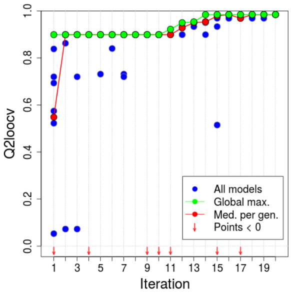
case of ***inactive*** *functional input*



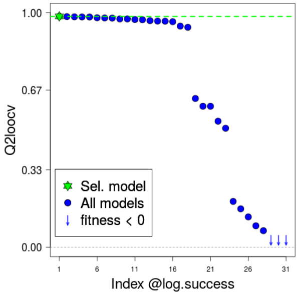
(a) Black-box 1.



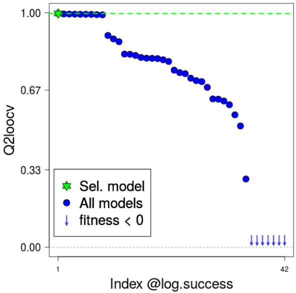
(b) Black-box 2.



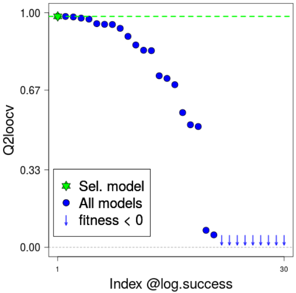
(c) Black-box 3.



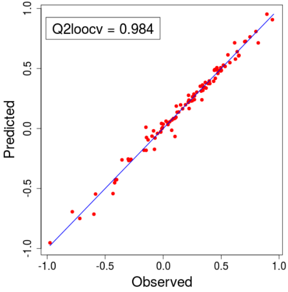
(a) Black-box 1.



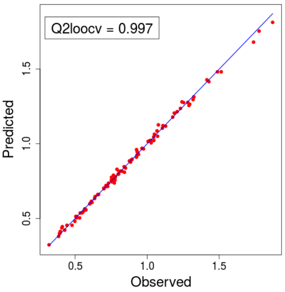
(b) Black-box 2.



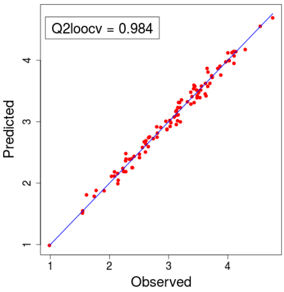
(c) Black-box 3.



(a) Black-box 1.



(b) Black-box 2.



(c) Black-box 3.

- **What our algorithm does:** smart selection of configurations to test
  What is doesn't do: faster evaluation of one configuration

Conclusions

- Better use the time based stopping condition
- The new algorithm allows for a wider exploration of features and levels
- Desirable to make this development publicly accessible (R package?)

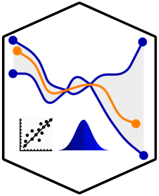
funGp

Gaussian Process Regression for Scalar and Functional Inputs with funGp
The in-depth tour

Authors: José Betancourt, François Bachoc, Thierry Klein.
Contributors: Déborah Idier, Jérémy Rhomer.

Official manual for `funGp` 0.1.0 (2020), downloadable from **CRAN** and **GitHub**.

- **Scalar-, Functional-, and Hybrid-input Gaussian Process Regression**

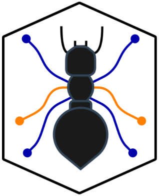


`Construction`

`Prediction`

`Simulation`

`Plotting`

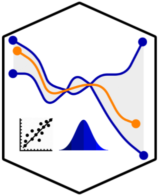


`Structural optimization`

`LOOCV or Hold-out V`

Ch.4 - funGp R package

- **Scalar-, Functional-, and Hybrid-input Gaussian Process Regression**

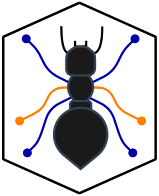


Construction

Prediction

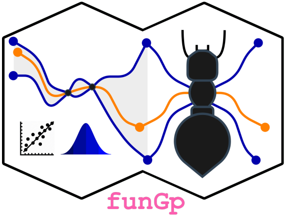
Simulation

Plotting



Structural optimization

LOOCV or Hold-out V



Efficient exploration of metamodel configurations

Selection of models with superior predictability

Ch.4 - funGp R package

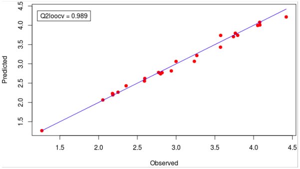
Metamodeling

- `fgpm(sIn, fIn, sOut)`
- `fgpm(sIn, fIn, sOut, basis)`
- `fgpm(sIn, fIn, sOut, kernel)`
- `fgpm(sIn, fIn, sOut, basis, dimen, dist, kernel, n.strt, cl)`

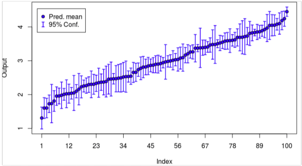
Structural optimization

- `fgpm_factory(sIn, fIn, sOut)`
- `fgpm_factory(sIn, fIn, sOut, constr, tlim, n.iter, ACOSetup, cl)`

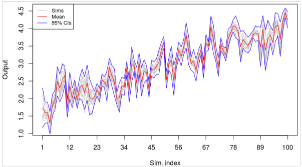
Validation



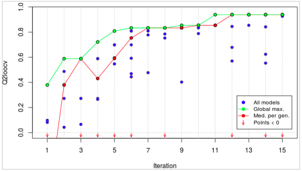
Prediction

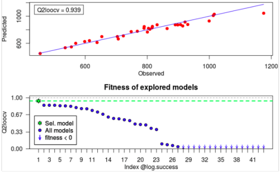


Simulation



Structural optimization





funGp

Chapter 4. Synthesis

funGp in brief

- Regression with scalar and functional inputs based on a non-parametric model
- Built-in dimension reduction
- Boosted performance through parallelization
- Modular-oriented implementation for easy extension
- Comprehensive manual available as pre-print in HAL
- Open source, available from **CRAN** and **GitHub**

Chapter 5. Structural optimization in RISCOPE - an update

RISCOPE + funGp

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**

WW3 → SWASH

Code A3

Outputs of interest
scalar

At the end of the event
- Inland water volume
- Flooded surface
- Water height at:
  - Townhall
  - Gymnasium

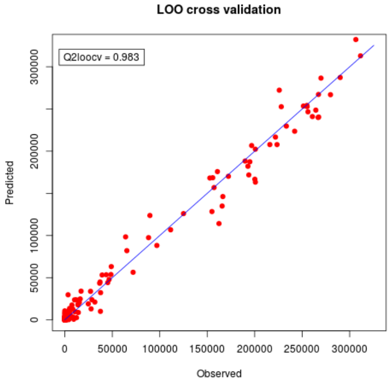
Maximum over time
- Flooded surface
- Water height at:
  - Townhall
  - Gymnasium

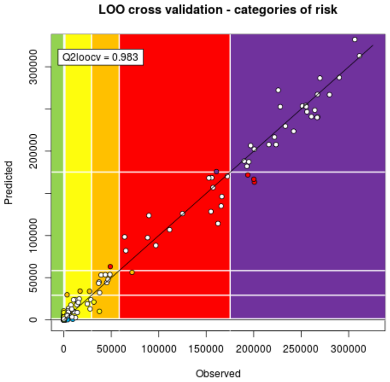
General during the event
- General trafficability at:
  - Road 1
  - Road 2
  - Road 3
- Tr. at high points:
  - Road 1
  - Road 2
  - Road 3

Q^2_{loocv} for each of the 13 output variables

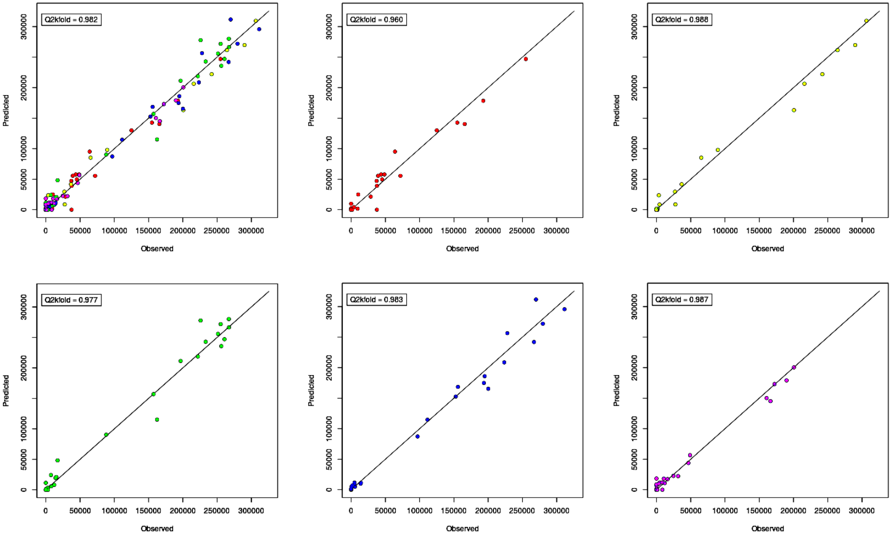
Output ID	All scalar	All scalar + all functional	All scalar + some functional	Ants selection	Ants active scalar	Ants active functional
Ysurf_max	0,959	0,957	0,962	0,980	{1,2,3,5,6,8}	{2,4}
Ysurf_fin	0,942	0,916	0,946	0,981	{1,2,3,5,6,7}	{7}
Yvol_fin	0,929	0,927	0,938	0,961	{1,2,3,5,6}	{3,4,5,7}
Ytr_ft1	0,927	0,929	0,928	0,963	{1,2,3,5,7,8}	{2}
Ytr_ft2	0,950	0,958	0,957	0,977	{1,3,5,7,8}	{2,5,6,7}
Ytr_ft3	0,533	0,655	0,469	0,764	{1,2,3,4,6,8}	{3,5,6}
Ytr_fh1	0,961	0,954	0,960	0,965	{1,2,3,5,7,8}	{2,4,5,6,7}
Ytr_fh2	0,941	0,862	0,959	0,974	{1,2,3,4,5,6,7,8}	{2}
Ytr_fh3	0,791	0,806	0,828	0,887	{1,3,4,5,7,8}	{2,4,5}
Yhcb_max1	0,576	0,499	0,443	0,900	{1,3,4,5,6}	{2,3,6,7,8}
Yhcb_max2	0,897	0,895	0,880	0,939	{1,3,4,6,7,8}	{2}
Yhcb_fin1	0,424	0,274	0,320	0,625	{1,2,4,6,7,8}	{3}
Yhcb_fin2	0,893	0,883	0,890	0,926	{1,3,5,6,7}	{2,4}

LOO cross-validation for `Ysurf_max`

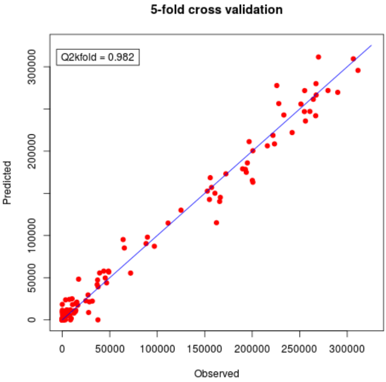


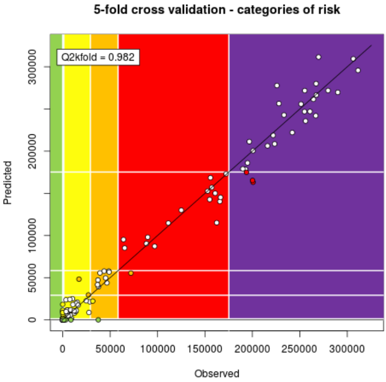


5-fold cross-validation for `Ysurf_max`



5-fold cross-validation for `Ysurf_max`





- The ACO algorithm allowed to achieve superior results the 13 metamodels
  → **robustness of the method** to variations in the input-output relationship

- Promising results for most of the outputs
  → **expected to improve** when adding `Ysurf_max` as input for the other metas

- Results achieved with `funGp` in 10 minutes
  → **expected to improve** when allowed to run for more time

- The solutions found by ACO were non-trivial

  `fgpm` **call delivered by ACO:**

```
fgpm(sIn[,c(1,2,3,4,6,8)],fIn[c(4,5)],Ysurf_max,
     basis = c("B-splines", "PCA"), dimen = c(3,3),
     dist = c("byindex", "byindex"), kernel = "gauss")
```

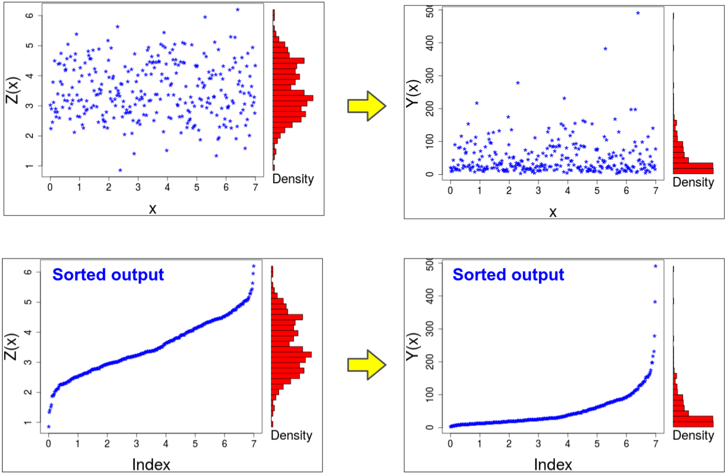
  → potential loss of predictability when we make selections by hand
  → **better to optimize**

Electronic Journal of Statistics
Vol. 14 (2020) 1962–2008
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1712

Asymptotic properties of the maximum likelihood and cross validation estimators for transformed Gaussian processes

François Bachoc[1], José Betancourt[2], Reinhard Furrer[3] and Thierry Klein[4]

Original process **Transformed process**



- Two asymptotic frameworks for covariance parameter estimation
  - fixed-domain: observation points dense in a bounded domain
  - increasing-domain: number of observation points is proportional to domain volume $\rightarrow$ unbounded observation domain
- Two estimators for covariance parameters estimation
  - Maximum likelihood (ML)
  - Cross validation (CV)
- Under increasing domain asymptotics
  - Assumption that observations come from Gaussian field: ML and CV consistent and asymptotically normal
  - Without the Gaussianity assumption: no results so far
- **Objective:** to extend the asymptotic results to observations from a non-Gaussian random field

Transformed Gaussian process

- Unobserved Gaussian process Z on $\mathbb{R}^d$, with zero mean and covariance function k_Z
- Fixed (and unknown) non-linear transformation function T
- Observed transformed Gaussian process Y defined by $Y(s) = T(Z(s))$ for any $s \in \mathbb{R}^d$. Assumed zero mean. Covariance function k_Y
- Sequence of observation locations $(s_i)_{i \in \mathbb{N}}$, $s_i \in \mathbb{R}^d$, $i \in \mathbb{N}$
- Non-Gaussian observation vector $y = (Y(s_1), ..., Y(s_n))^\top$ with covariance matrix $R = (k_Y(s_i - s_j))_{i,j=1,...,n}$, for $n \in \mathbb{N}$

Covariance parameters

- Parametric set of stationary covariance functions $\{k_{Y,\theta}; \theta \in \Theta\}$ on $\mathbb{R}^d$, with Θ a compact set of $\mathbb{R}^p$
- $\theta = (\sigma^2, \psi)$, with σ^2 and ψ the variance and correlation parameters, resp.
- $k_Y = k_{Y,\theta_0}$, with $\theta_0 \in \Theta$

- The **maximum likelihood** estimator of θ_0 is

$$\hat{\theta}_{\text{ML}} \in \underset{\theta \in \Theta}{\operatorname{argmin}}\, L_\theta,$$

with $$L_\theta = \frac{1}{n} \left(\log(\det(R_\theta)) + y^\top R_\theta^{-1} y \right)$$

- The **cross validation** estimator of ψ_0 is

$$\hat{\psi}_{\text{CV}} \in \underset{\psi \in \mathcal{S}}{\operatorname{argmin}}\, CV_\psi,$$

with $$CV_\psi = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_{i,\psi})^2$$

Bounds on the elements of inverse covariance matrices *(Theorem 1)*

- Let $(s_i)_{i\in\mathbb{N}}$ be the sequence of observation locations, with $s_i \in \mathbb{R}^d$
- Let R be the covariance matrix of the non-Gaussian observations

Then we have

$$\left|\left(R^{-1}\right)_{i,j}\right| \leq \frac{C_{\text{sup}}}{1 + |s_i - s_j|^{d+\tau}},$$

with $C_{\text{sup}} < \infty$, d the input dimension and $0 < \tau < \infty$

CLT for quadratic forms of trans-Gps *(Theorem 2)*

- Let y be the n-element vector of non-Gaussian observations.
- Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of matrices (each $n \times n$). Assume

$$|(A_n)_{i,j}| \leq \frac{C_{\text{sup}}}{1 + |s_i - s_j|^{d+C_{\text{inf}}}}, \quad \forall n\in\mathbb{N}, \quad \forall i,j=1,\dots,n$$

- Let $V_n = \frac{1}{n} y^\top A_n y$. Let $\mathcal{L}_n$ be the distribution of $\sqrt{n}(V_n - \mathbb{E}[V_n])$

$$\text{As } n \to \infty, \quad d_w\big(\mathcal{L}_n, \mathcal{N}[0, n\,\text{Var}(V_n)]\big) \to 0$$

In addition, $(n\,\text{Var}(V_n))_{n\in\mathbb{N}}$ is bounded

Maximum likelihood *(Theorems 3 and 4, resp.)*

- **Consistency:** $\hat{\theta}_{\text{ML}} \xrightarrow{p} \theta_0$ as $n \to \infty$.
- **Asymptotic normality:**

  Let $\mathcal{L}_{\theta_0,n}$ be the distribution of $\sqrt{n}(\hat{\theta}_{\text{ML}} - \theta_0)$. Using Theorem 2,

  $$\text{as } n \to \infty, \quad d_w\left(\mathcal{L}_{\theta_0,n}, \mathcal{N}[0, M_{\theta_0}^{-1}\Sigma_{\theta_0}M_{\theta_0}^{-1}]\right) \to 0,$$

  with $$(M_{\theta_0})_{i,j} = (1/n)\,\mathrm{tr}\left(R_{\theta_0}^{-1}(\partial R_{\theta_0}/\partial\theta_i)R_{\theta_0}^{-1}(\partial R_{\theta_0}/\partial\theta_j)\right)$$
  $$(\Sigma_{\theta_0})_{i,j} = \mathrm{Cov}(\sqrt{n}\ \partial L_{\theta_0}/\partial\theta_i, \sqrt{n}\ \partial L_{\theta_0}/\partial\theta_j), \qquad i,j = 1,\dots,p.$$

Ch.6 - Main results

Maximum likelihood *(Proofs of Theorems 3 and 4, resp.)*

- **Steps for consistency:**
  1. Show that $\mathrm{Var}(L_\theta) \to 0$ as $n \to \infty$
  2. Show that $\sup_{\theta \in \theta} \left| \frac{\partial L_\theta}{\partial \theta} \right| = O_p(1)$ by finding a suitable bound for it.
  3. Proceed as in the proof of Proposition 3.1 in Bachoc (2014).

- **Steps for asymptotic normality:**
  1. Expand the first derivative of L_θ as a Taylor series around θ_0.
  2. Evaluate at $\hat{\theta}_{\mathrm{ML}}$, equal to zero, solve for $(\hat{\theta}_{\mathrm{ML}} - \theta_0)$ and multiply by $\sqrt{n}$.
  3. Apply Slutsky lemma.
     - 3.1 Check convergence in probability of $\sqrt{n} \left[-\partial^2 L_{\theta_0} / \partial \theta^2 \right]^{-1} V$.
     - 3.2 Check convergence in distribution of $\sqrt{n} \left[-\partial^2 L_{\theta 0} / \partial \theta^2 \right]^{-1} \partial L \theta_0 / \partial \theta$.
  4. Show that the elements of the asymptotic covariance matrix are bounded.
     - 4.1 Show that elements of $M_{\theta_0}^{-1}$ and Σ_{θ_0} are bounded.
     - 4.2 Conclude that $\limsup_{n \to \infty} \lambda_1 (M_{\theta_0}^{-1} \Sigma_{\theta_0} M_{\theta_0}^{-1}) < +\infty$.

Cross Validation *(Theorems 5 and 6, resp.)*

- **Consistency:** $\hat{\psi}_{\text{CV}} \xrightarrow{p} \psi_0$ as $n \to \infty$.
- **Asymptotic normality:**

  Let $\mathcal{Q}_{\psi_0,n}$ be the distribution of $\sqrt{n}(\hat{\psi}_{\text{CV}} - \psi_0)$. Using Theorem 2,

  $$\text{as } n \to \infty, \quad d_w\left(\mathcal{Q}_{\psi_0,n}, \mathcal{N}[0, N_{\psi_0}^{-1}\Gamma_{\psi_0}N_{\psi_0}^{-1}]\right) \to 0,$$

  with $(N_{\theta_0})_{i,j} =$ a long to write expression (provided in the manuscript)

  $$(\Gamma_{\psi_0})_{i,j} = \text{Cov}(\sqrt{n}\ \partial CV_{\psi_0}/\partial\psi_i, \sqrt{n}\ \partial CV_{\psi_0}/\partial\psi_j), \qquad i,j = 1,\ldots,p.$$

Joint estimator *(Theorem 7)*

- **Consistency:** follows from Theorems 3 and 5.
- **Asymptotic normality:**

Let $\mathcal{Q}_{\theta_0,n}$ be the distribution of $\sqrt{n}\begin{pmatrix}\hat{\theta}_{\text{ML}} - \theta_0 \\ \hat{\psi}_{\text{CV}} - \psi_0\end{pmatrix}$. Using Theorem 2,

$$\text{as } n \to \infty, \quad d_w\left(\mathcal{Q}_{\theta_0,n}, \mathcal{N}[0, D_{\theta_0}^{-1}\Psi_{\theta_0} D_{\theta_0}^{-1}]\right) \to 0,$$

with $D_{\theta_0} =$ a long to write expression depending on M_{θ_0} and N_{ψ_0} (provided in the manuscript)

$$\Sigma_{\theta_0} = \mathrm{Cov}(\sqrt{n}\ \partial L_{\theta_0}/\partial\theta, \partial CV_{\psi_0}/\partial\psi).$$

Conclusions

- Covariance parameters of transformed Gaussian processes can be properly estimated by ML and CV
  - Both consistent and asymptotically normal
- Same rate of convergence $\sqrt{n}$ for ML and CV
- Same rate of convergence $\sqrt{n}$ as for Gaussian processes
- Explicit expressions for the asymptotic covariance matrix

Foreseen extension

- Same questions but estimating the transform parameters
  - potential improvement in predictions
  - possibility to account for boundary constraints

Perspectives and contributions

Ongoing work and perspectives

RISCOPE till 2021! (José till August) - short term perspectives

- Metamodels for the Decision Support System
  → ACO allowed to run for a longer period
  → `Ysurf_max` as input for the other metamodels
- Analysis of patterns in the record of explored metamodels
- More levels of structural parameters → other projection methods

Beyond RISCOPE - long term perspectives

- ACO extensions
  - More levels of structural parameters → other distances, kernels
  - Other structural parameters → type of mean function
  - Functional inputs in larger dimensions → fields or images
  - Functional outputs → time as input, ensamble of metas, cokriging
- Strongly skewed data → corrective DOE, transformed Gp

Scientific production

Published papers

- **Gaussian process metamodeling of functional-input code for coastal flood hazard assessment**, *Reliability Engineering & System Safety*, 2020.
  Betancourt, Bachoc, Klein, Idier, Pedreros & Rohmer (2020)
- **Asymptotic properties of the maximum likelihood and cross validation estimators for transformed Gaussian processes**, *Electronic Journal of Statistics*.
  Bachoc, Betancourt, Furrer, & Klein (2020)

Conference paper

- **Toward a user-based, robust and fast running method for coastal flooding forecast, early warning, and risk prevention**, *Journal of coastal research, proceedings from the International Coastal Symposium (ICS) 2020*.
  Idier, Aurouet, Bachoc, Baills, Betancourt, Durand, Mouche, Rohmer, Gamboa, Klein, Lambert, Le Cozannet, Le Roy, Louisor, Pedreros & Véron (2020)

Software

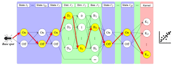
- **R package funGp**, *Available on CRAN and GitHub since 2020*.
  Authors: Betancourt, Bachoc, Klein   Contributors: Idier, Rohmer
- funGp user manual: **Gaussian process regression for scalar and functional inputs with funGp - The in-depth tour.** *Available as pre-print from HAL (2020)*.

Summary of contributions

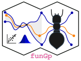
- Metamodels for Decision Support System in coastal flood early warning
  - Most of them promising

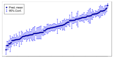
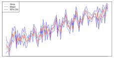
Output ID	All scalar	All scalar + all functional	All scalar + some functional	Ants selection	Ants active scalar	Ants active functional
Ysurf_max	0.959	0.957	0.962	0.980	{1,2,3,5,6,8}	{2,4}
Ysurf_fin	0.942	0.916	0.946	0.981	{1,2,3,5,6,7}	{7}
Yvol_fin	0.929	0.927	0.938	0.961	{1,2,3,5,6}	{3,4,5,7}
Ytr_ft1	0.927	0.929	0.928	0.963	{1,2,3,5,7,8}	{2}
Ytr_ft2	0.950	0.958	0.957	0.977	{1,3,5,7,8}	{2,5,6,7}
Ytr_ft3	0.533	0.655	0.469	0.764	{1,2,3,4,6,8}	{3,5,6}
Ytr_fh1	0.961	0.954	0.900	0.965	{1,2,3,5,7,8}	{2,4,5,6,7}
Ytr_fh2	0.941	0.862	0.959	0.974	{1,2,3,4,5,6,7,8}	{2}
Ytr_fh3	0.791	0.806	0.828	0.887	{1,3,4,5,7,8}	{2,4,5}
Yheb_max1	0.576	0.499	0.443	0.900	{1,3,4,5,6}	{2,3,6,7,8}
Yheb_max2	0.897	0.895	0.880	0.939	{1,3,4,6,7,8}	{2}
Yheb_fin1	0.424	0.274	0.320	0.625	{1,2,4,6,7,8}	{3}
Yheb_fin2	0.893	0.883	0.890	0.926	{1,3,5,6,7}	{2,4}

- Novel approach for optimization of structural parameters
  - Pertinency made evident for all the 13 output variables

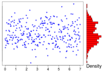


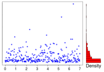
- R package integrating the metamodeling techniques used in RISCOPE
  - Extends the state of the art in various ways



- First asymptotic results for transformed Gaussian processes
  - Properties of ML and CV in a common yet unexplored setting





Maximum likelihood *(Theorems 3 and 4)*
Cross Validation *(Theorems 5 and 6)*
Joint estimator *(Theorem 7)*

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

José Betancourt

Advisor: Prof. Thierry Klein
Co-advisor: Prof. François Bachoc

PhD thesis defense - Applied Mathematics
June 08, 2020

Bibliography I

Bachoc, F. (2014). Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis*, *125*, 1-35.

Bachoc, F., Betancourt, J., Furrer, R., Klein, T., et al. (2020). Asymptotic properties of the maximum likelihood and cross validation estimators for transformed gaussian processes. *Electronic Journal of Statistics*, *14*(1), 1962–2008.

Betancourt, J., Bachoc, F., & Klein, T. (2020a). *funGp: Gaussian process models for scalar and functional inputs.* https://github.com/djbetancourt-gh/funGp. GitHub.

Betancourt, J., Bachoc, F., & Klein, T. (2020b, April). *fungp: Gaussian process models for scalar and functional inputs.* https://CRAN.R-project.org/package=funGp. (R package version 0.1.0)

Betancourt, J., Bachoc, F., & Klein, T. (2020c, April). *Gaussian process regression for scalar and functional inputs with fungp - the in-depth tour. RISCOPE project.* https://hal.archives-ouvertes.fr/hal-02536624. (hal-02536624)

Betancourt, J., Bachoc, F., Klein, T., & Gamboa. (2020, April). *Technical report: Ant colony based model selection for functional-input gaussian process regression. Ref. D3.b (WP3.2), RISCOPE project.* https://hal.archives-ouvertes.fr/hal-02532713. (hal-02532713)

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

Bibliography II

Cohen, A., Daubechies, I., & Feauveau, J.-C. (1992). Biorthogonal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, *45*(5), 485–560.

Goss, S., Aron, S., Deneubourg, J.-L., & Pasteels, J. M. (1989). Self-organized shortcuts in the argentine ant. *Naturwissenschaften*, *76*(12), 579–581.

Idier, D., Aurouet, A., Bachoc, F., Baills, A., Betancourt, J., Durand, J., . . . Vron, A. (2020). Toward a user-based, robust and fast running method for coastal flooding forecast, early warning, and risk prevention. *Journal of coastal research*, *95*, 11–15.

Lataniotis, C., Marelli, S., & Sudret, B. (2018). Extending classical surrogate modelling to ultrahigh dimensional problems through supervised dimensionality reduction: a data-driven approach. *arXiv preprint arXiv:1812.06309*.

Montgomery, D. C. (1997). Response surface methods and other approaches to process optimization. *Design and analysis of experiments*.

Muehlenstaedt, T., Fruth, J., & Roustant, O. (2017). Computer experiments with functional inputs and scalar outputs by a norm-based approach. *Statistics and Computing*, *27*(4), 1083–1097.

Nanty, S., Helbert, C., Marrel, A., Pérot, N., & Prieur, C. (2016). Sampling, metamodeling, and sensitivity analysis of numerical simulators with functional stochastic inputs. *SIAM/ASA Journal on Uncertainty Quantification*, *4*(1), 636–659.

Papaioannou, I., Ehre, M., & Straub, D. (2019). Pls-based adaptation for efficient pce representation in high dimensions. *Journal of Computational Physics*, *387*, 186–204.

Pulido, H. G., De la Vara Salazar, R., González, P. G., Martínez, C. T., & Pérez, M. d. C. T. (2012). *Análisis y diseño de experimentos*. New York: McGraw-Hill.