



CERAMATHS



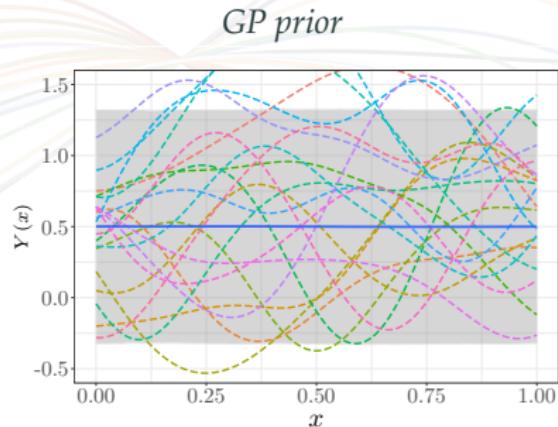
Université
Polytechnique
HAUTS-DE-FRANCE

Gaussian processes with applications in statistical learning and machine learning

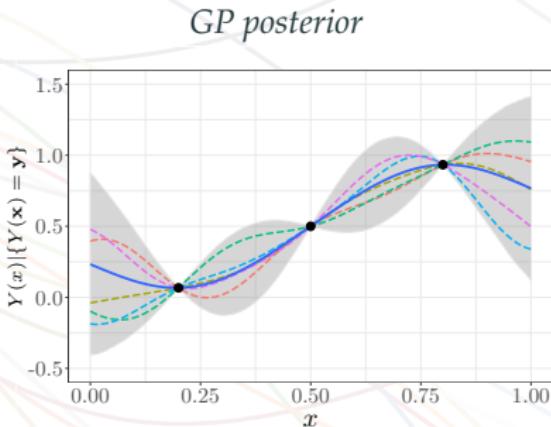
Andrés F. López-Lopera
CERAMATHS, Probability & Statistiques
Université Polytechnique Hauts-de-France
April 25, 2024

Why Gaussian processes (GPs)?

- They provide a well-founded Bayesian (non-parametric) framework



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

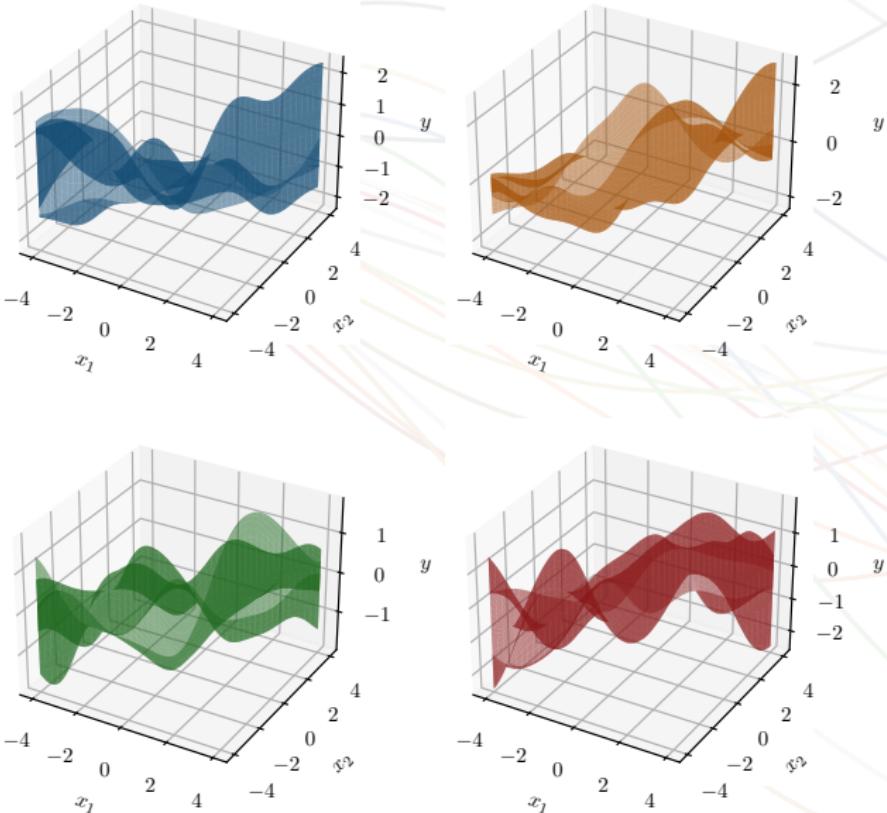


$$Y| \{Y(x) = y\} \sim \mathcal{GP}(m_{\text{cond}}, c_{\theta})$$

■ mean function ■ uncertainty ■ ■ ... ■ GP samples

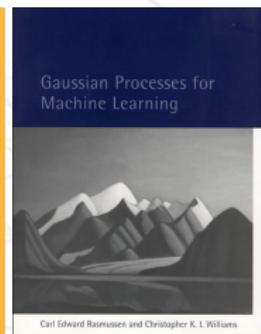
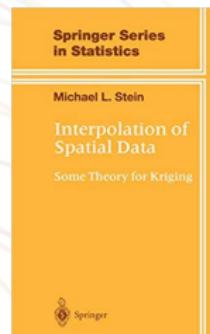
- interpolation points: $(x, y) = (x_i, y_i)_{i=1}^n$

Why Gaussian processes (GPs)?



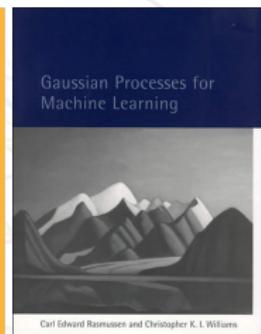
Why Gaussian processes (GPs)?

- GPs have been successfully applied in statistical learning, machine learning and beyond:
 - Regression and classification tasks
 - Spatial modeling
 - Transfer learning
 - Active learning
 - Reinforcement learning
 - Bayesian optimization



Why Gaussian processes (GPs)?

- GPs have been successfully applied in statistical learning, machine learning and beyond:
 - Regression and classification tasks
 - Spatial modeling
 - Transfer learning
 - Active learning
 - Reinforcement learning
 - Bayesian optimization



Why Gaussian processes (GPs)?

NeurIPS Proceedings 40 0

gaussian process

Search

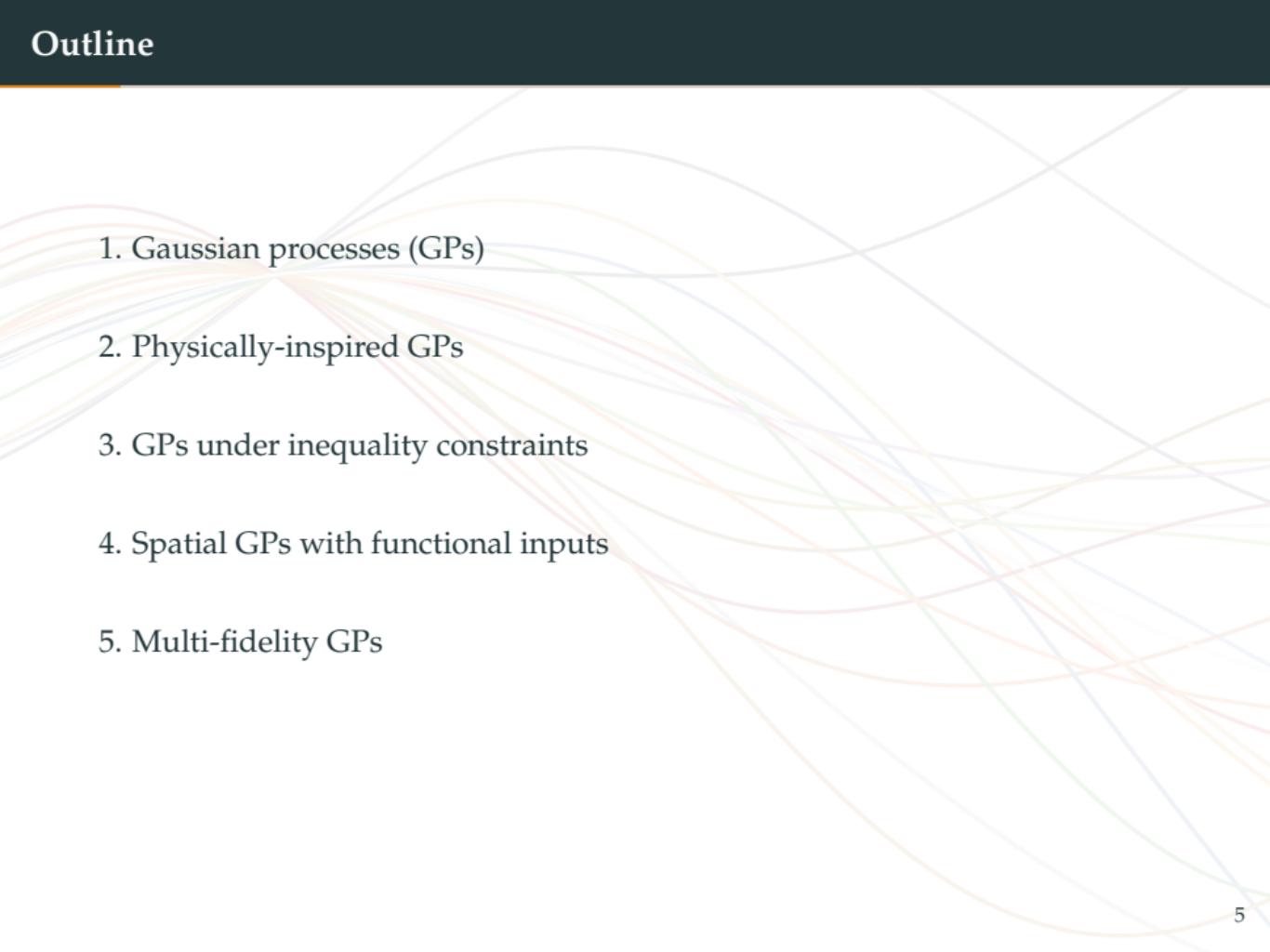
Search Results

Main Conference Track

- (2023) [Variational Gaussian Processes with Decoupled Conditionals](#) Xinran Zhu, Kaiwen Wu, Natalie Maus, Jacob Gardner, David Bindel
- (2023) [Pointwise uncertainty quantification for sparse variational Gaussian process regression with a Brownian motion prior](#) Luke Travis, Kolyan Ray
- (2023) [Sharp Calibrated Gaussian Processes](#) Alexandre Capone, Sandra Hirche, Geoff Pleiss
- (2023) [Explaining the Uncertain: Stochastic Shapley Values for Gaussian Process Models](#) Siu Lun Chau, Krikamol Muandet, Dino Sejdinovic
- (2023) [Wide Neural Networks as Gaussian Processes: Lessons from Deep Equilibrium Models](#) Tianxiang Gao, Xiaokai Huo, Hailiang Liu, Hongyang Gao
- (2023) [MMGP: a Mesh Morphing Gaussian Process-based machine learning method for regression of physical problems under nonparametrized geometrical variability](#) Fabien Casenave, Brian Staber, Xavier Roynard
- (2023) [Failure-Aware Gaussian Process Optimization with Regret Bounds](#) Shogo Iwazaki, Shion Takeno, Tomohiko Tanabe, Mitsuji Irie
- (2023) [A Bayesian Take on Gaussian Process Networks](#) Enrico Giudice, Jack Kuipers, Gijsje Moffa
- (2023) [On the Identifiability and Interpretability of Gaussian Process Models](#) Jiawen Chen, Wancenc Mu, Yun Li, Didong Li
- (2023) [Leveraging Locality and Robustness to Achieve Massively Scalable Gaussian Process Regression](#) Robert Allison, Anthony Stephenson, Samuel F, Edward O Pyzer-Knapp
- (2023) [Sampling from Gaussian Process Posteriors using Stochastic Gradient Descent](#) Jihao Andreas Lin, Javier Antorán, Shreyas Padhy, David Janz, José Miguel Hernández-Lobato, Alexander Terenin
- (2023) [Thin and deep Gaussian processes](#) Daniel Augusto de Souza, Alexander Nikitin, St John, Magnus Ross, Mauricio A Álvarez, Marc Deisenroth, João Paulo Gomes, Diego Mesquita, César Lincoln Mattos
- (2023) [Posterior Contraction Rates for Matérn Gaussian Processes on Riemannian Manifolds](#) Paul Ross, Slava Borovitskiy, Alexander Terenin, Judith Rousseau
- (2023) [Contextual Gaussian Process Bandits with Neural Networks](#) Haoteng Zhang, Jinghai He, Rhonda Righter, Zuo-Jun Shen, Zeyu Zheng
- (2023) [GAUCHE: A Library for Gaussian Processes in Chemistry](#) Ryan-Rhys Griffiths, Leo Klemmer, Henry Moss, Aditya Ravuri, Sung Truong, Yuqiu Du, Samuel Stanton, Gary Tom, Bojana Rankovic, Arjan Jamasb, Aryan Deshwal, Julius Schwartz, Austin Tripp, Gregory Kell, Simon Frieder, Anthony Bourbache, Alex Chan, Jacob Moss, Chengzhi Guo, Johannes Peter Döhrhoff, Soudamini Chaurasia, Ji Won Park, Felix Strieth-Kalthoff, Alpha Lee, Bingqing Cheng, Alan Aspuru-Guzik, Philippe Schwaller, Jian Tang
- (2023) [Gaussian Process Probes \(GPP\) for Uncertainty-Aware Probing](#) Zi Wang, Alexander Xu, Jason Baldridge, Tom Griffiths, Been Kim
- (2023) [Implicit Manifold Gaussian Process Regression](#) Bernardo Fichera, Slava Borovitskiy, Andreas Krause, Aude G Billard
- (2023) [Graph-structured Gaussian Processes for transferable graph Learning](#) Jun Wu, Lisa Ainsworth, Andrew Leakey, Haizun Wang, Jingui He
- (2023) [Variational Gaussian processes for linear inverse problems](#) Thibault RANDEIANARANIA, Botond Szabó
- (2022) [High-dimensional Additive Gaussian Processes under Monotonicity Constraints](#) Andrés López-Lopera, François Bachoc, Olivier Roustant
- (2022) [Posterior and Computational Uncertainty in Gaussian Processes under Monotonicity Constraints](#) Jonathan Wehenk, Geert Peeters, Kravinn Pilonier, Philipp Herding, Julian P. Cunningham
- (2022) [Symplectic Spectrum Gaussian Processes: Learning Hamiltonians from Noisy and Sparse Data](#) Yusuke Tanaka, Tomoharu Iwata, naonori ueda
- (2022) [On the inability of Gaussian process regression to optimally learn compositional functions](#) Matteo Giordano, Kolyan Ray, Johannes Schmidt-Hieber
- (2022) [Bezier Gaussian Processes for tall and Wide Data](#) Martin Jørgensen, Michael A Osborne

· For further papers in ML: <https://proceedings.mlr.press/>

Outline

The background of the slide features a complex, abstract design composed of numerous thin, curved lines in various colors, including light blue, orange, green, and red. These lines overlap and intersect in a way that suggests a network or a family of functions.

1. Gaussian processes (GPs)
2. Physically-inspired GPs
3. GPs under inequality constraints
4. Spatial GPs with functional inputs
5. Multi-fidelity GPs

Gaussian processes (GPs)

Gaussian processes (GPs)

- Let $\{Y(x); x \in \mathcal{D}\}$ be a stochastic process defined on a compact input space $\mathcal{D} \subseteq \mathbb{R}^d$ (e.g. $\mathcal{D} = [0, 1]^d$).
- Y is GP-distributed if, for all $x_1, \dots, x_n \in \mathcal{D}$,

$$\mathbf{Y}_n := \begin{bmatrix} Y(x_1), \dots, Y(x_n) \end{bmatrix}^\top \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}).$$

with mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$.

- By convention, we denote the GP Y as

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

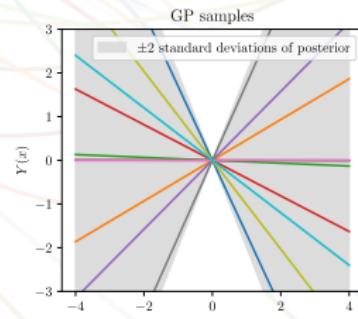
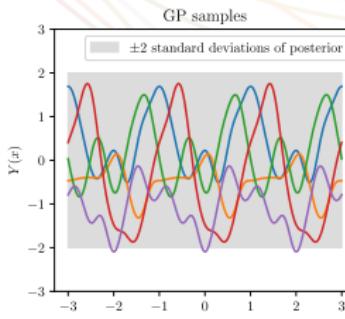
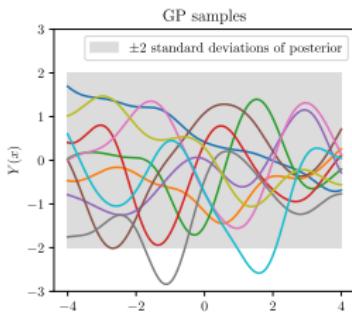
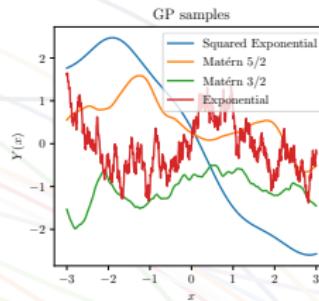
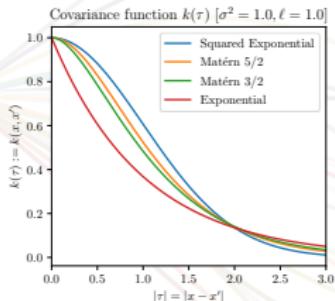
with mean function $\mu(x) = \mathbb{E}(Y(x))$ and covariance function (or kernel)
 $k(x, x') := \text{cov}(Y(x), Y(x')) = \mathbb{E}([Y(x) - \mu(x)][Y(x') - \mu(x')])$, for $x, x' \in \mathcal{D}$.

- Without loss of generality, in the following we will assume $\mu(\cdot) = 0$. Then,

$$k(x, x') := \text{cov}(Y(x), Y(x')) = \mathbb{E}(Y(x)Y(x')).$$

Gaussian processes (GPs)

- In GPs, regularity assumptions are commonly encoded in kernels



[link]

GP regression

- Let $\{Y(x), x \in \mathbb{R}^d\}$ be a zero-mean GP with covariance function k
- In regression tasks, we aim at computing the conditional distribution:

$$Y | \{Y(x_1) = y_1, \dots, Y(x_n) = y_n\},$$

for a set of observations $(x_i, y_i)_{1 \leq i \leq n}$ for $n \in \mathbb{N}$

- This conditional process (namely, the *posterior*) is also GP-distributed with (conditional) mean and covariance functions given by

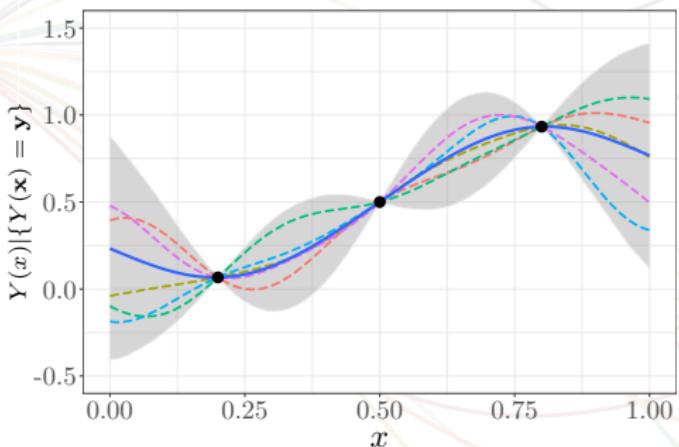
$$\mu(x) = k^\top(x) \mathbf{K}_n^{-1} y,$$

$$c(x, x') = k(x, x') - k^\top(x) \mathbf{K}_n^{-1} k(x'),$$

with $k(x) = (\text{cov}\{Y(x), Y(x_i)\})_{1 \leq i \leq n}$ and $\mathbf{K}_n = (\text{cov}\{Y(x_i), Y(x_j)\})_{1 \leq i, j \leq n}$

GP regression

GP regression



- conditional mean ■ confidence intervals ■ ... ■ GP samples
- training data: $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}_i, y_i)_{i=1}^n$

[link]

GP regression with noisy observations

- For noisy observations, we have the conditional process:

$$Y | \{Y(x_1) + \varepsilon_1 = y_1, \dots, Y(x_n) + \varepsilon_n = y_n\},$$

where $\varepsilon_i \sim \mathcal{N}(0, \tau^2)$ are (i.i.d) additive noises with noise variance $\tau^2 \in \mathbb{R}^+$

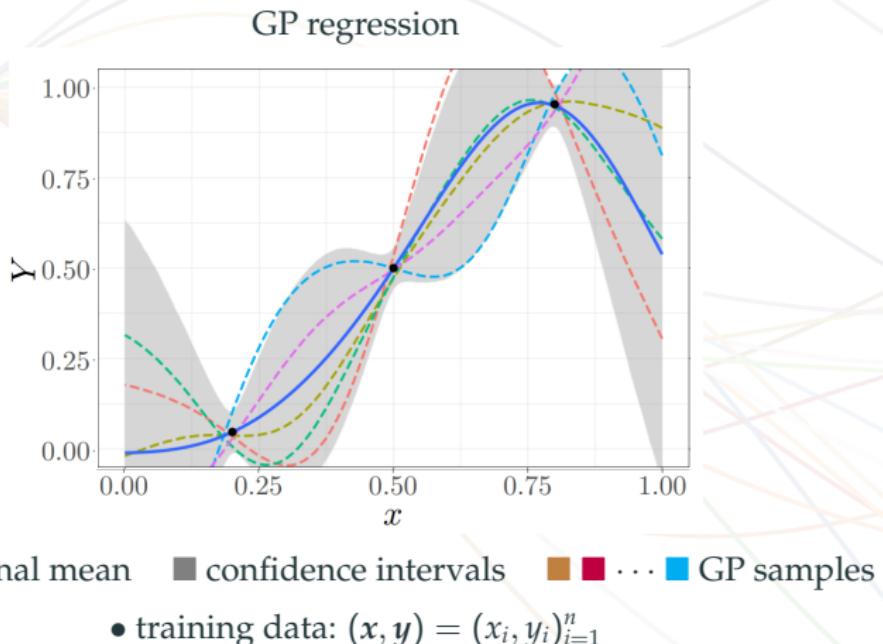
- This conditional process (namely, the *posterior*) is also GP-distributed with conditional mean and covariance functions given by

$$\tilde{\mu}(x) = k^\top(x) [\mathbf{K}_n + \tau^2 \mathbf{I}]^{-1} \mathbf{y},$$

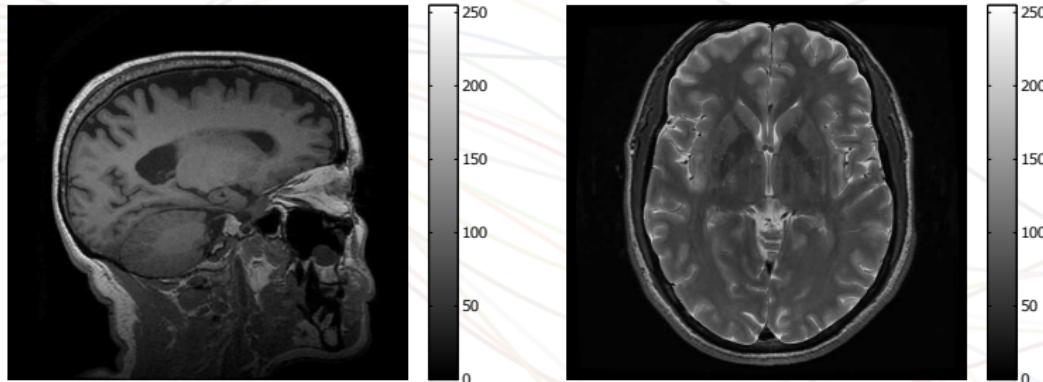
$$\tilde{c}(x, x') = k(x, x') - [\mathbf{K}_n + \tau^2 \mathbf{I}]^{-1} k(x'),$$

with $k(x) = (\text{cov}\{Y(x), Y(x_i)\})_{1 \leq i \leq n}$ and $\mathbf{K}_n = (\text{cov}\{Y(x_i), Y(x_j)\})_{1 \leq i, j \leq n}$

GP regression with noisy observations

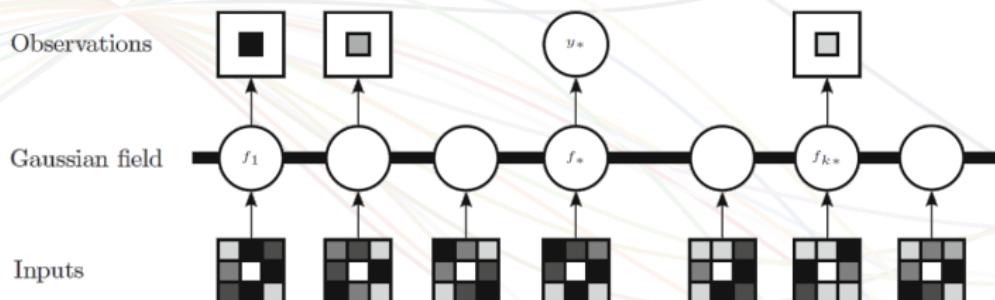


Application in neuroscience: Super-resolution of MR images



 H. Vargas, A. F. López-Lopera, M. Álvarez, A. Orozco, J. Hernández and N. Malpica (2015). Gaussian processes for slice-based super-resolution MR images. International Symposium on Advances in Visual Computing (ISVC), Lecture Notes in Computer Science.

Application in neuroscience: Super-resolution of MR images



H. Vargas, A. F. López-Lopera, M. Álvarez, A. Orozco, J. Hernández and N. Malpica (2015). Gaussian processes for slice-based super-resolution MR images. International Symposium on Advances in Visual Computing (ISVC), Lecture Notes in Computer Science.

Physically-inspired GPs

Physically-inspired GPs: First-order ODE

- Consider the first-order ODE:

$$\frac{d\mathbf{Y}(t)}{dt} + \gamma \mathbf{Y}(t) = S\mathbf{U}(t), \quad (1)$$

with $\gamma \in \mathbb{R}^+, S \in \mathbb{R}^+$

- Without loss of generality, assume that $\mathbf{Y}(0) = 0$. Then we have

$$\mathbf{Y}(t) = Sc(t) \int_0^t \mathbf{U}(\tau) \exp(\gamma\tau) d\tau, \quad \text{with } c(t) = \exp(-\gamma t)$$

- If $\mathbf{U} \sim \mathcal{GP}(0, k_{\mathbf{u}, \mathbf{u}})$, we can show that \mathbf{Y} is also a GP with covariance function:

$$\begin{aligned} k_{\mathbf{y}, \mathbf{y}}(t, t') &:= \text{cov}\{\mathbf{Y}(t), \mathbf{Y}(t')\} (= \mathbb{E}\{\mathbf{Y}(t)\mathbf{Y}(t')\}) \\ &= S^2 c(t)c(t') \int_0^t \exp(\gamma\tau) \int_0^{t'} \exp(\gamma\tau') \underbrace{k_{\mathbf{u}, \mathbf{u}}(\tau, \tau')}_{\mathbb{E}\{\mathbf{U}(\tau), \mathbf{U}(\tau')\}} d\tau' d\tau \end{aligned}$$

- In addition,

$$k_{\mathbf{y}, \mathbf{u}}(t, t') := \text{cov}\{\mathbf{Y}(t), \mathbf{U}(t')\} = Sc(t) \int_0^t \exp(\gamma\tau) k_{\mathbf{u}, \mathbf{u}}(\tau, t') d\tau$$

Physically-inspired GPs: First-order ODE

- Since the joint process (\mathbf{U}, \mathbf{Y}) is GP-distributed, then the joint distribution of $\mathbf{U} = (\mathbf{U}(t_1), \dots, \mathbf{U}(t_n))$ and $\mathbf{Y} = (\mathbf{Y}(t_1), \dots, \mathbf{Y}(t_n))$ is Gaussian:

$$\begin{bmatrix} \mathbf{U} \\ \mathbf{Y} \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_{\mathbf{u}, \mathbf{u}} & \mathbf{K}_{\mathbf{y}, \mathbf{u}}^\top \\ \mathbf{K}_{\mathbf{y}, \mathbf{u}} & \mathbf{K}_{\mathbf{y}, \mathbf{y}} \end{bmatrix} \right)$$

- GP properties can be used, e.g., for inferring \mathbf{U} given observations of \mathbf{Y} (**inverse problem**) by establishing the conditional distribution:

$$\mathbf{U}(t) | \{\mathbf{Y}(t_1) = y_1, \dots, \mathbf{Y}(t_n) = y_n\} \sim \mathcal{GP}(\mu_{\mathbf{u}|\mathbf{y}}, k_{\mathbf{u}|\mathbf{y}}),$$

where

$$\begin{aligned} \mu_{\mathbf{u}|\mathbf{y}}(t) &= k_{\mathbf{y}, \mathbf{u}}^\top(t) \mathbf{K}_{\mathbf{y}, \mathbf{y}}^{-1} \mathbf{y}, \\ k_{\mathbf{u}|\mathbf{y}}(t, t') &= k_{\mathbf{u}, \mathbf{u}}(t, t') - k_{\mathbf{y}, \mathbf{u}}^\top(t) \mathbf{K}_{\mathbf{y}, \mathbf{y}}^{-1} \mathbf{K}_{\mathbf{y}, \mathbf{u}}(t), \end{aligned}$$

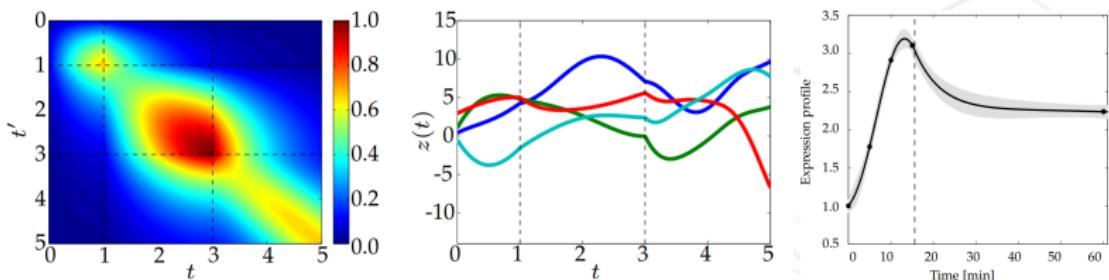
with $k_{\mathbf{y}, \mathbf{u}}(t) = (k_{\mathbf{y}, \mathbf{u}}(t, t_j))_{1 \leq j \leq n}$

Applications in biology: Transcriptional regulation in genes



M. Álvarez

Univ. of Manchester, UK
GPs & Machine Learning



We considered the coupled system of ODEs:

$$\frac{dY_d(t)}{dt} + \gamma_d Y_d(t) = B_d + \sum_{r=1}^R S_{r,d} U_r(t),$$

with U_1, \dots, U_R independent GPs.

■ A. F. López-Lopera and M. Álvarez (2019). Switched latent force models for reverse-engineering transcriptional regulation in genes. IEEE/ACM Transaction on Computational Biology and Bioinformatics, 16(1).



M. Alvarez



Univ. of Manchester, UK
GPs & Machine Learning



N. Durrande



Shift Lab, UK
GPs & Machine Learning

- We can consider the (linear) reaction-diffusion equation:

$$\frac{\partial Y(x, t)}{\partial t} = SU(x, t) - \lambda Y(x, t) + D \frac{\partial^2 Y(x, t)}{\partial x^2},$$

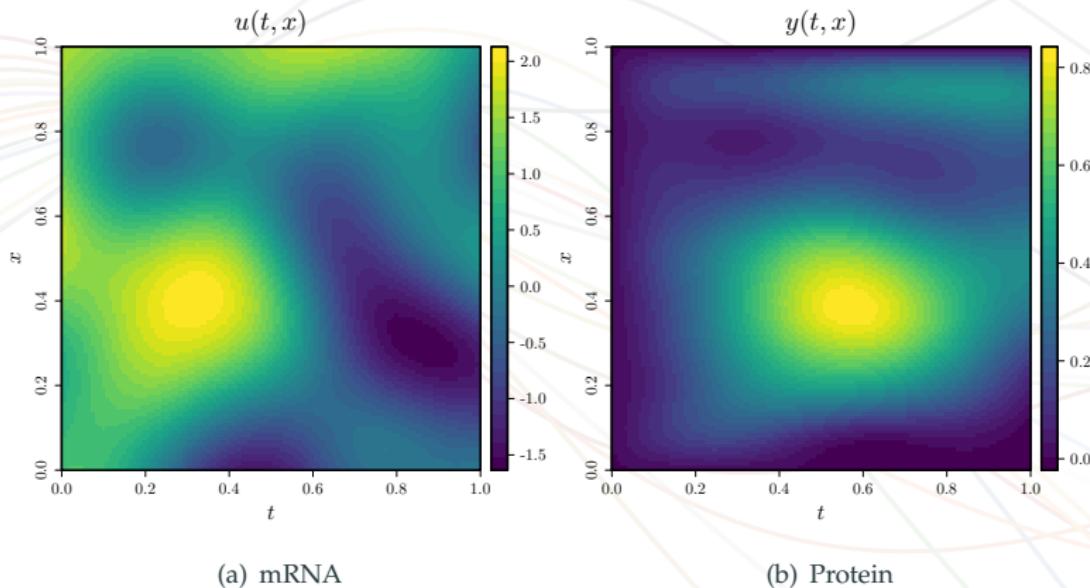
where

- Y : relative gap protein concentration;
 - U : messenger RNA (mRNA);
 - S, D, λ : translation, decay and diffusion rate constants, respectively.
- For simplicity, we assume homogeneous conditions:

$$Y(x, t = 0) = 0, \quad Y(x = 0, t) = Y(x = l, t) = 0, \quad \text{for } x \in [0, l], l \in \mathbb{R}^+.$$

■ A. F. López-Lopera, N. Durrande and M. Álvarez (2021). Physically-inspired Gaussian process models for post-transcriptional regulation in Drosophila. IEEE/ACM Transaction on Computational Biology and Bioinformatics, 18(2).

Physically-inspired GPs: Spatio-temporal PDE in biology

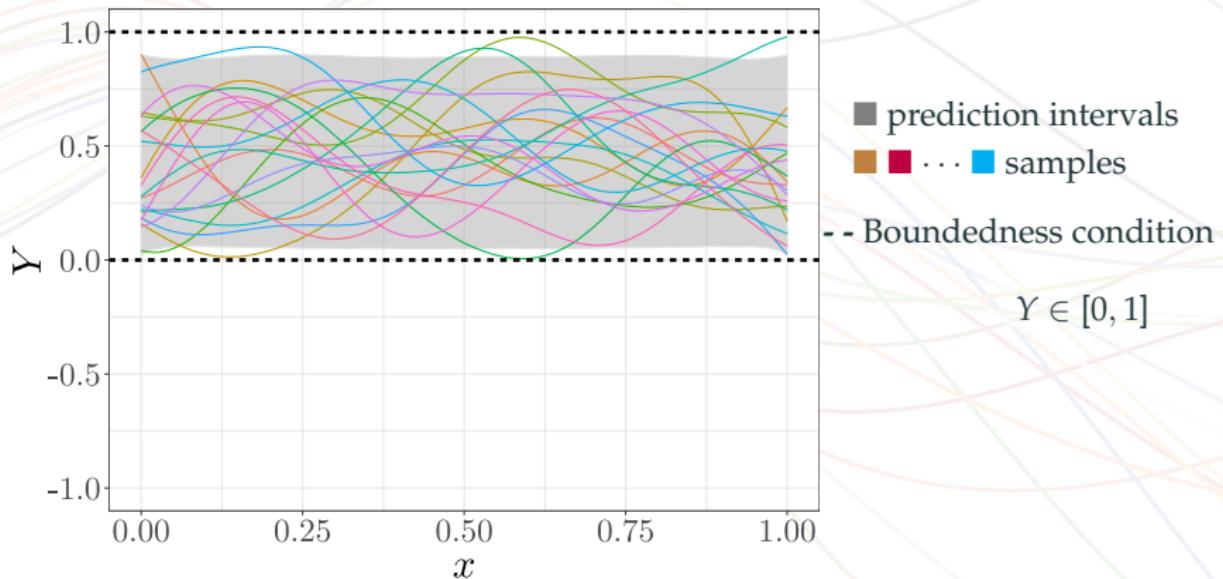


Sample from the GP-mRNA model

GPs under inequality constraints

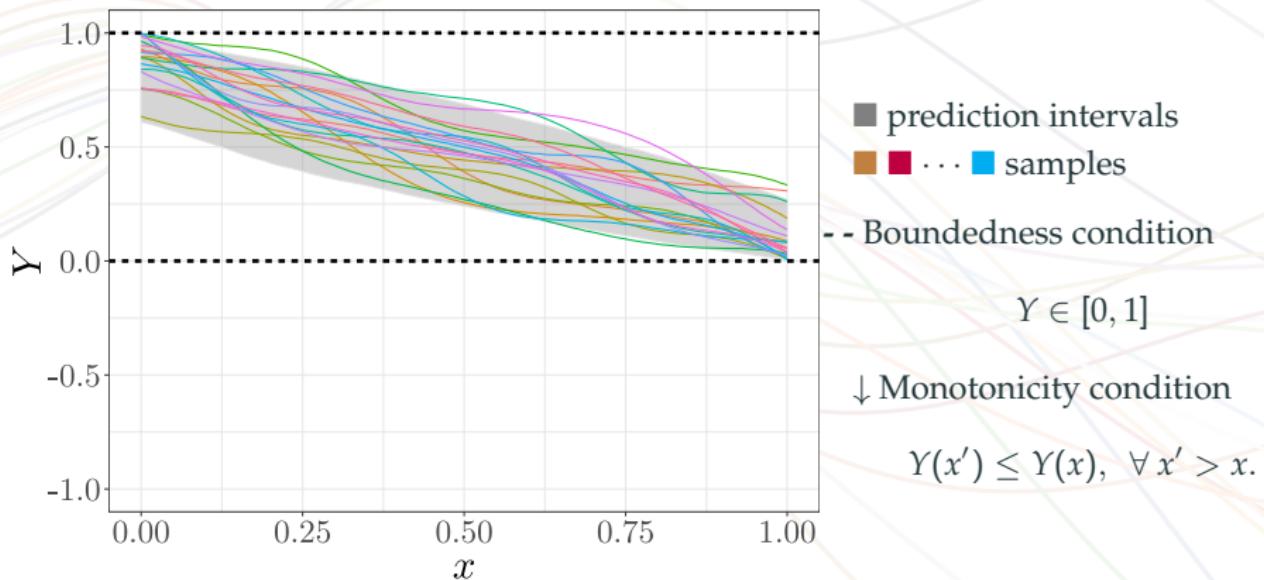
GPs under inequality constraints

Our interest: GP-based priors satisfying some inequality constraints...

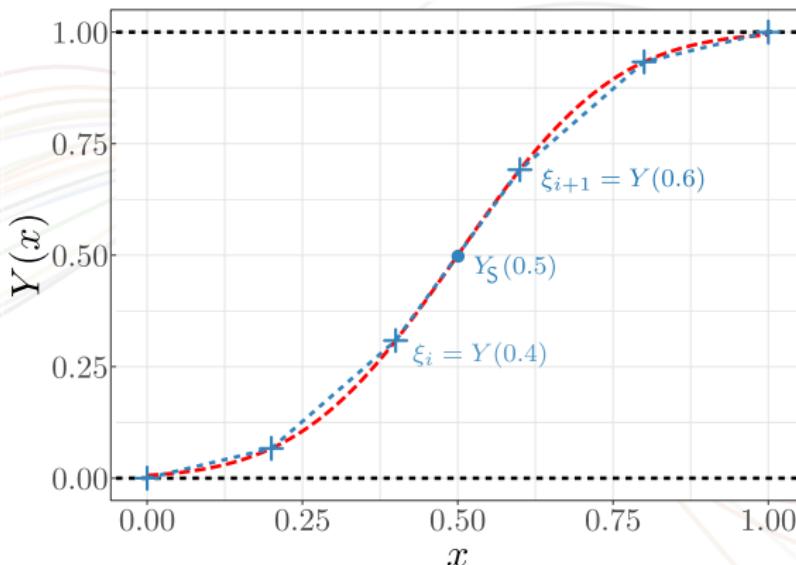


GPs under inequality constraints

Our interest: GP-based priors satisfying some inequality constraints...



Finite-dimensional approximation of GPs



smooth function Y

piecewise approximation Y_S

Note that:

- If $\xi_j \in [0, 1]$ for $j = 1, \dots, m$,

$$Y_S(0.5) \in [0, 1].$$

- Or if $\xi_j < \xi_{j+1}$ for $j = 1, \dots, m - 1$,

$$\xi_j < Y_S(0.5) < \xi_{j+1}.$$

Pro: imposing constraints over knots is enough [Maatouk and Bay, 2017]:

$$Y_S \in \mathcal{E} \Leftrightarrow \xi \in \mathcal{C}.$$

Finite-dimensional approximation of GPs

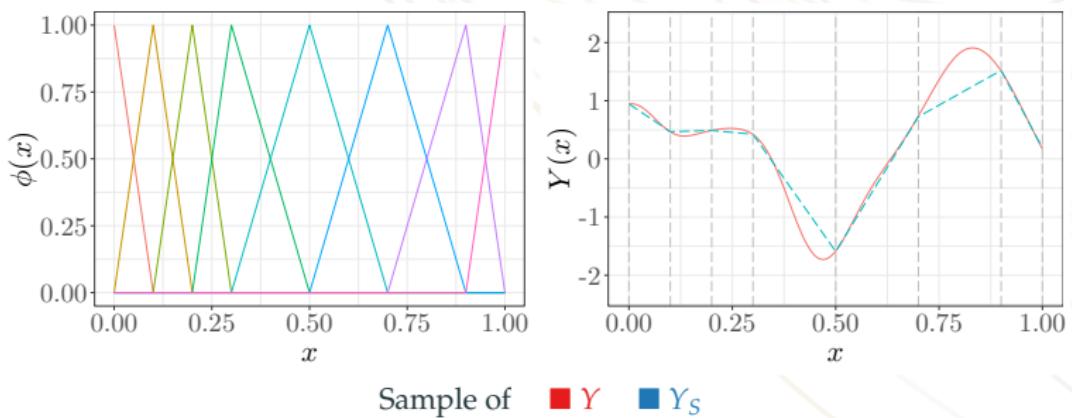
- Let \mathcal{Y}_S be the finite-dimensional GP with an ordered set of knots:

$$\mathcal{S} = \{t_0, \dots, t_m\}, \quad \text{with } 0 = t_0 < \dots < t_m = 1,$$

such that

$$Y_S(x) = \sum_{j=1}^m Y(t_j) \phi_j(x), \quad (2)$$

where $x \in [0, 1]$, $Y \sim \mathcal{GP}(0, k_\theta)$, and $\phi_j : [0, 1] \mapsto \mathbb{R}$ are (asymmetric) hat basis functions.



Finite-dimensional approximation of GPs

- For regression tasks under inequality constraints, we have

$$Y_S(x) = \sum_{j=1}^m \xi_j \phi_j(x), \text{ s.t. } \begin{cases} Y_S(x_i) + \varepsilon_i = y_i & (\text{regression conditions}), \\ \mathbf{l} \leq \boldsymbol{\Lambda} \boldsymbol{\xi} \leq \mathbf{u} & (\text{linear inequality conditions}), \end{cases} \quad (3)$$

where $x_i \in [0, 1]$, $y_i \in \mathbb{R}$ for $i = 1, \dots, n$, and

- $\xi_j := Y(t_j)$ for $j = 1, \dots, m$, i.e. $\boldsymbol{\xi} = [\xi_1, \dots, \xi_m]^\top \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\theta)$ with covariance matrix $\boldsymbol{\Sigma}_\theta = (k_\theta(t_j, t_{j'}))_{1 \leq j, j' \leq m}$
- $\varepsilon_i \sim \mathcal{N}(0, \tau^2)$, for $i = 1, \dots, n$, with noise variance τ^2
- $(\boldsymbol{\Lambda}, \mathbf{l}, \mathbf{u})$ define the constraints. For monotonicity constraints:

$$\underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{l}} \leq \underbrace{\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_m \end{bmatrix}}_{\boldsymbol{\xi}} \leq \underbrace{\begin{bmatrix} \infty \\ \infty \\ \vdots \\ \infty \end{bmatrix}}_{\mathbf{u}}$$

Finite-dimensional approximation of GPs

- Since $Y_S \in \mathcal{E} \Leftrightarrow \xi \in \mathcal{C}$, then *uncertainty quantification* relies on simulating the **truncated vector ξ** :

$$\Lambda \xi | \{\Phi \xi + \varepsilon = y, l \leq \Lambda \xi \leq u\} \sim \mathcal{T}\mathcal{N}(\Lambda \mu_c, \Lambda \Sigma_c \Lambda^\top, l, u), \quad (4)$$

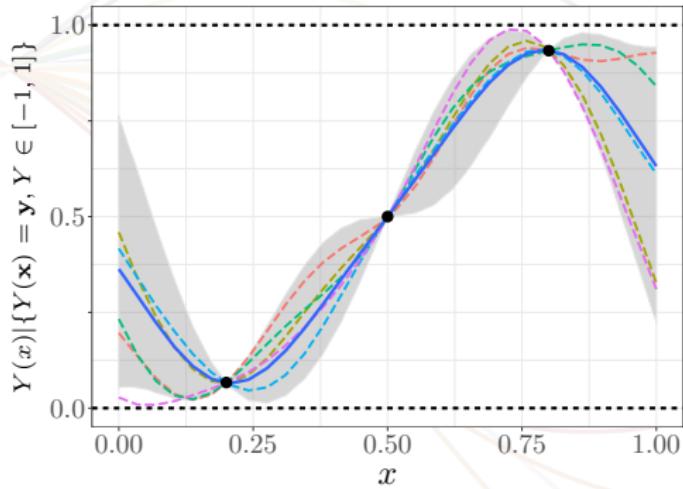
with conditional parameters μ_c and Σ_c given by

$$\mathbf{K} = \Phi \Sigma \Phi^\top + \tau^2 \mathbf{I}, \quad \mu_c = \Sigma \Phi^\top \mathbf{K}^{-1} y, \quad \Sigma_c = \Sigma - \Sigma \Phi^\top \mathbf{K}^{-1} \Phi \Sigma. \quad (5)$$

- (4) can be approximated via *Monte Carlo* (MC) or *Markov chain MC* (MCMC):
 - e.g. *Hamiltonian Monte Carlo* (HMC) [Pakman and Paninski, 2014]

GPs under inequality constraints: Numerical illustration

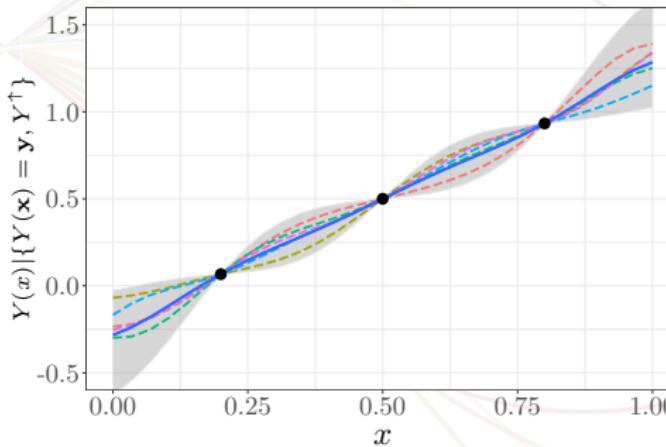
1D example with **boundedness** constraints via HMC



$$\underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}}_l \leq \underbrace{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \vdots \\ \xi_{m-1} \\ \xi_m \end{bmatrix}}_{\xi} \leq \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}}_u$$

GPs under inequality constraints: Numerical illustration

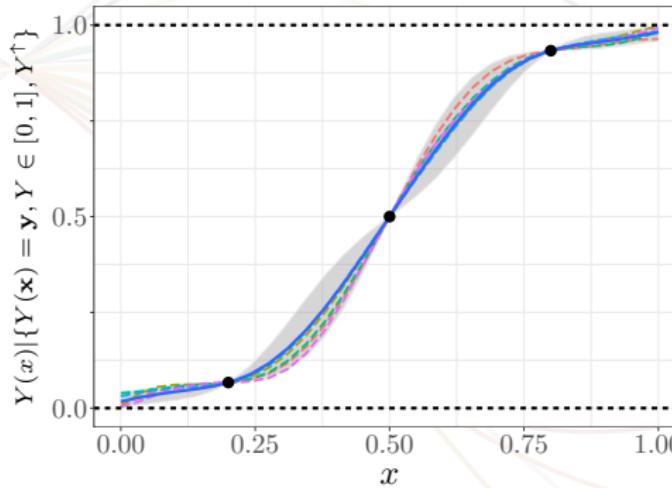
1D example with **boundedness & monotonicity** constraints via HMC



$$\underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_l \leq \underbrace{\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \vdots \\ \xi_{m-1} \\ \xi_m \end{bmatrix}}_{\xi} \leq \underbrace{\begin{bmatrix} \infty \\ \infty \\ \vdots \\ \infty \end{bmatrix}}_u$$

GPs under inequality constraints: Numerical illustration

1D example with **boundedness & monotonicity** constraints via HMC



$$\underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}}_l \leq \underbrace{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \vdots \\ \xi_{m-1} \\ \xi_m \end{bmatrix}}_{\xi} \leq \underbrace{\begin{bmatrix} 1 \\ \infty \\ \infty \\ \vdots \\ \infty \\ 1 \end{bmatrix}}_u$$

The maximum a posteriori (mode) function in 1D

- Let $\hat{\xi}$ be the mode that maximizes the pdf of $\xi | \{\Phi\xi + \varepsilon = y, l \leq \Lambda\xi \leq u\}$:

$$\hat{\xi} = \underset{\xi \text{ s.t. } l \leq \Lambda\xi \leq u}{\arg \max} \{-[\xi - \mu_c]^\top \Sigma_c^{-1} [\xi - \mu_c]\}, \quad (6)$$

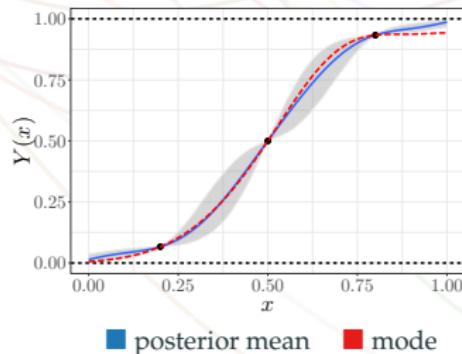
with $\hat{\xi} = [\hat{\xi}_1, \dots, \hat{\xi}_m]^\top$.

- The MAP estimate of Y_S is given by

$$\hat{Y}_S(x) = \sum_{j=1}^m \hat{\xi}_j \phi_j(x). \quad (7)$$

Pro:

- \hat{Y}_S can be used as a point estimate
- Fast to compute
- Convergence of \hat{Y}_S to the spline solution as $m \rightarrow \infty$ [Bay et al., 2016]
- Starting point for MCMC methods



■ posterior mean ■ mode

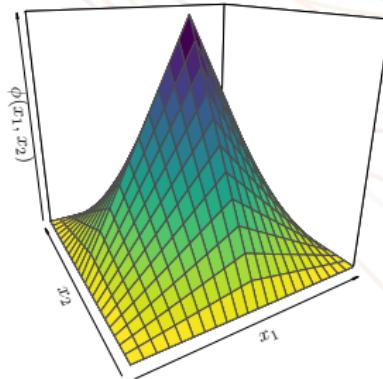
GPs under inequality constraints: Extension to d dimensions

- The extension to d dimensions is obtained by **tensorization**:

$$Y_S(x) = \sum_{j_1, \dots, j_d=1}^{m_1, \dots, m_d} \left[\prod_{p=1, \dots, d} \phi_{j_p}^{(p)}(x_p) \right]_{\xi_{j_1, \dots, j_d}}, \text{ s.t. } \begin{cases} Y_m(x_i) + \varepsilon_i = y_i, \\ \xi \in \mathcal{C}, \end{cases} \quad (8)$$

where $x_i \in [0, 1]^d$, $y_i \in \mathbb{R}$, $\varepsilon_i \sim \mathcal{N}(0, \tau^2)$, for $i = 1, \dots, n$; and

- $\xi = [\xi_{1, \dots, 1}, \dots, \xi_{m_1, \dots, m_d}]^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_\theta)$,
- \mathcal{C} is a convex set of linear inequality constraints, and
- $\phi_{j_i}^{(i)} : [0, 1] \mapsto \mathbb{R}$ are hat basis functions.





O. Roustant

INSA, Toulouse

GPs & Applied Maths



F. Bachoc

IMT, Toulouse

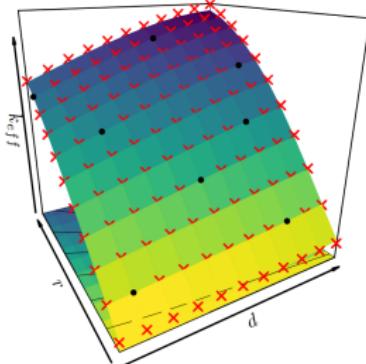
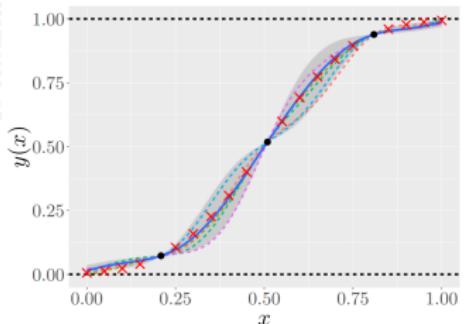
GPs & Applied Maths



N. Durrande

Shift Lab, UK

GPs & Machine Learning



■ A. F. López-Lopera, N. Durrande, F. Bachoc and O. Roustant (2018). Finite-dimensional Gaussian approximation with linear inequality constraints. SIAM/ASA Journal on Uncertainty Quantification, 6(3).

■ A. F. López-Lopera, F. Bachoc, N. Durrande, J. Rohmer, D. Idier, and O. Roustant (2019). Approximating Gaussian process emulators with linear inequality constraints and noisy observations via MC and MCMC. In International Conference in Monte Carlo & Quasi-Monte Carlo Methods, Springer Proceedings in Mathematics & Statistics.



N. Durrande



Shift Lab, UK

GPs & Machine Learning

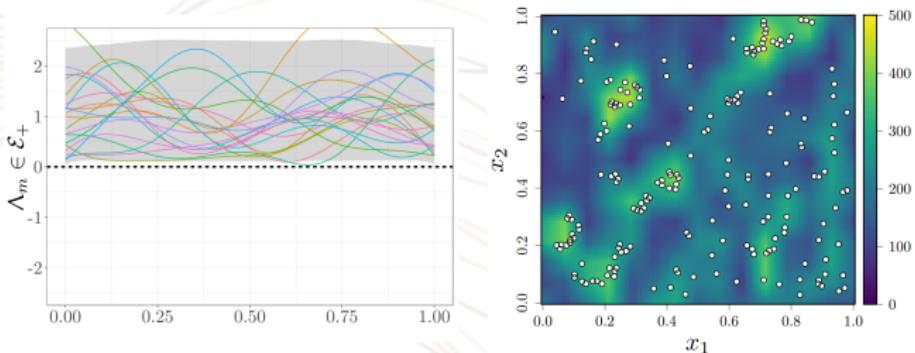
Ti John



Aalto University

Machine Learning

Geostatistics: Spatial location of redwood trees [Ripley, 1977]



- We considered Cox processes with a (non-negative) GP-distributed stochastic intensity function:

$$\Lambda_m(x) = \sum_{j=1}^m \phi_j(x) \xi_j \quad \text{s.t.} \quad \Lambda_m \in \mathcal{E}_+$$

- A. F. López-Lopera, S. John and N. Durrande (2019). Gaussian process modulated Cox processes under linear inequality constraints. International Conference on Artificial Intelligence and Statistics (AISTATS).

GPs under inequality constraints



O. Roustant

INSA, Toulouse

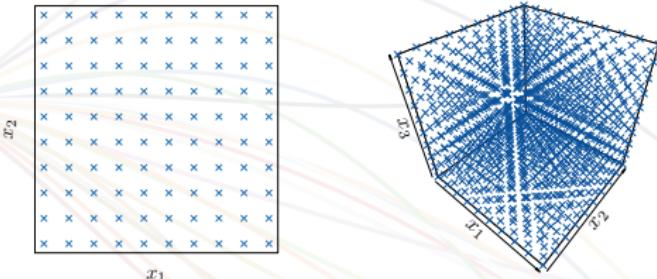
GPs & Applied Maths



F. Bachoc

IMT, Toulouse

GPs & Applied Maths



- Con: the cost of Y_S increases as d increases.
- This drawback can be mitigated by considering:
 - a "*smarter*" construction of rectangular grids of knots thanks to the asymmetric construction of the hat basis functions
 - and/or *further assumptions for complexity simplification*
→ e.g. *inactive variables, additive structures*

■ F. Bachoc, A. F. López-Lopera, and O. Roustant (2022). Sequential construction and dimension reduction of Gaussian processes under inequality constraints. SIAM Journal on Mathematics of Data Science, 4(2).

■ A. F. López-Lopera, F. Bachoc, and O. Roustant (2022). High-dimensional additive Gaussian processes under monotonicity constraints. In Advances in Neural Information Processing Systems (NeurIPS), volume 35.

The MaxMod algorithm in 1D

- Let \widehat{Y}_S be the MAP function with an ordered set of knots:

$$S = \{t_0, \dots, t_m\}, \quad \text{with } 0 = t_0 < \dots < t_m = 1.$$

- Here, we aim at adding a new knot t in S (where?)
- To do so, we aim at *maximizing the total modification of the MAP*:

$$I_S(t) = \int_{[0,1]} \left(\widehat{Y}_{S \cup t}(x) - \widehat{Y}_S(x) \right)^2 dx. \quad (9)$$

- The integral in (18) has a closed-form expression.

Algorithm MaxMod (maximum modification of the MAP) in 1D

Input parameters: the initial subdivision $S^{(0)} \in \mathcal{S}$.

Sequential procedure: for $\kappa \in \mathbb{N}$, do:

- Set $t_{\kappa+1}^* \in [0, 1]$ such that

$$I_{S^{(\kappa)}}(t_{\kappa+1}^*) \geq \sup_{t \in [0,1]} I_{S^{(\kappa)}}(t)$$

- $S^{(\kappa+1)} = S^{(\kappa)} \cup t_{\kappa+1}^*$.
-

The MaxMod algorithm in 1D

1D example under boundedness and monotonicity constraints

MAP estimate

conditional sample-path

- training points
- predictive mean
- + knots
- MAP estimate
- 90% confidence intervals

The MaxMod algorithm in higher dimensions

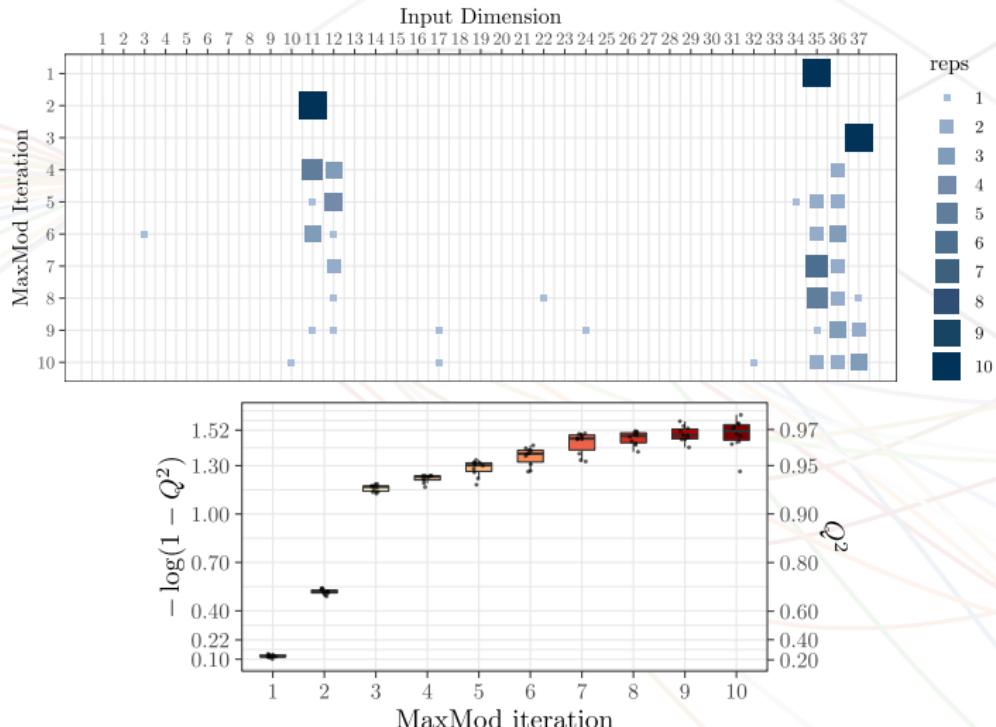
- Let $\widehat{Y}_{\mathcal{J}, \mathbf{s}}$ be the MAP function with $|\mathcal{J}|$ active variables and ordered sets of knots $\mathbf{s}_{\mathcal{J}}$ for $\mathcal{J} \subseteq \{1, \dots, D\}$.
- Then, the criterion to maximise is given by

$$I_{\mathcal{J}, \mathbf{s}}(\mathbf{i}, \mathbf{t}) = \begin{cases} \frac{1}{N_{\mathbf{s}, \mathcal{J}, i}} \int_{[0,1]^d} (\widehat{Y}_{\mathcal{J}, \mathbf{s} \cup_i \mathbf{t}}(\mathbf{x}) - \widehat{Y}_{\mathcal{J}, \mathbf{s}}(\mathbf{x}))^2 d\mathbf{x} & \text{if } i \in \mathcal{J}, \\ \frac{1}{N_{\mathbf{s}, \mathcal{J}, i}} \int_{[0,1]^{d+1}} (\widehat{Y}_{\mathcal{J} \cup \{i\}, \mathbf{s} + i}(\mathbf{x}) - \widehat{Y}_{\mathcal{J}, \mathbf{s}}(\mathbf{x}))^2 d\mathbf{x} & \text{if } i \notin \mathcal{J}, \end{cases} \quad (10)$$

where $N_{\mathbf{s}, \mathcal{J}, i}$ is the increase of the number of basis functions.

 F. Bachoc, A. F. López-Lopera, and O. Roustant (2022). Sequential construction and dimension reduction of Gaussian processes under inequality constraints. SIAM Journal on Mathematics of Data Science, 4(2).

GPs under inequality constraints: Flood study of the Vienne river



The choice made by MaxMod per iteration (top) and Q^2 boxplots (bottom). Results are computed over 10 replicates. For the first panel, a bigger and darker square implies a more repeated choice.

- ★ The extension to block-additivity is being studied by Mathis Deronzier (PhD student at the IMT), for instance for disjoint blocks:

$$\sum_{I \in \mathcal{I}} Z_I(x_I),$$

where \mathcal{I} contains groups of input variables. We seek to study:

- Variable selection (i.e. inference of the partition \mathcal{I})
 - Further real-world applications
- ★ The extension to Student- t processes.

■ M. Deronzier, A. F. López-Lopera, F. Bachoc, and O. Roustant (2024). Block additive Gaussian processes under monotonicity constraints: Covariance structure and model selection. In preparation.

■ A. F. López-Lopera and Ari Pakman (2024). t -processes under inequality constraints via exact Hamiltonian Monte Carlo. Work in progress.

Maximum likelihood estimation under constraints



F. Bachoc

IMT, Toulouse

GPs & Applied Maths



A. Lagnoux

IMT, Toulouse

Applied Maths

- Consider $\{k_\theta; \theta \in \Theta\}$, with $\Theta \subset \mathbb{R}^p$, a parametric family of covariance functions where θ defines the covariance parameters

- The maximum likelihood estimator, with log-likelihood function

$$\mathcal{L}_n(\theta) := \log p_\theta(\mathbf{Y}_n),$$
 is given by

$$\hat{\theta}_{\text{MLE}} \in \arg \max_{\theta \in \Theta} \mathcal{L}_n(\theta).$$

- We have studied the estimation of θ accounting for inequality constraints:

$$\hat{\theta}_{\text{cMLE}} \in \arg \max_{\theta \in \Theta} \mathcal{L}_{\mathcal{C},m}(\theta),$$

with the conditional log-likelihood function $\mathcal{L}_{\mathcal{C},m}(\theta) := \log p_\theta(\mathbf{Y}_n | \boldsymbol{\xi} \in \mathcal{C})$:

$$\mathcal{L}_{\mathcal{C},m}(\theta) = \log p_\theta(\mathbf{Y}_n) + \log P_\theta(\boldsymbol{\xi} \in \mathcal{C} | \Phi \boldsymbol{\xi} = \mathbf{Y}_m) - \log P_\theta(\boldsymbol{\xi} \in \mathcal{C})$$

■ A. F. López-Lopera, N. Durrande, F. Bachoc and O. Roustant (2018). Finite-dimensional Gaussian approximation with linear inequality constraints. SIAM/ASA Journal on Uncertainty Quantification, 6(3).

■ F. Bachoc, A. Lagnoux and A. F. López-Lopera (2019). Maximum likelihood estimation for Gaussian processes under inequality constraints. Electronic Journal of Statistics, 13(2).

Asymptotic consistency of the MLE & cMLE

- Let \mathcal{E}_κ be one of the following convex set of functions (mild conditions)

$$\mathcal{E}_\kappa = \begin{cases} f : \mathbb{X} \rightarrow \mathbb{R}, f \text{ is } C^0 \text{ and } \forall x \in \mathbb{X}, \ell \leq f(x) \leq u & \text{if } \kappa = 0, \\ f : \mathbb{X} \rightarrow \mathbb{R}, f \text{ is } C^1 \text{ and } \forall x \in \mathbb{X}, \forall i = 1, \dots, d, \frac{\partial}{\partial x_i} f(x) \geq 0 & \text{if } \kappa = 1, \\ f : \mathbb{X} \rightarrow \mathbb{R}, f \text{ is } C^2 \text{ and } \forall x \in \mathbb{X}, \frac{\partial^2}{\partial x^2} f(x) \text{ is a p.s.d. matrix} & \text{if } \kappa = 2. \end{cases}$$

- Denote: $\boldsymbol{\theta}_0$ (true covariance parameters), $\hat{\boldsymbol{\theta}}_n$ (MLE), $\hat{\boldsymbol{\theta}}_{n,c}$ (cMLE).

Proposition (Consistency of the MLE and cMLE)

Assume $\forall \varepsilon > 0$ and $\forall M < \infty$,

$$P(\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \varepsilon} (\mathcal{L}_n(\boldsymbol{\theta}) - \mathcal{L}_n(\boldsymbol{\theta}_0)) \geq -M) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Then,

$$P(\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \varepsilon} (\mathcal{L}_{n,c}(\boldsymbol{\theta}) - \mathcal{L}_{n,c}(\boldsymbol{\theta}_0)) \geq -M \mid Y \in \mathcal{E}_\kappa) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Consequently, both the MLE and cMLE are consistent estimators:

$$\hat{\boldsymbol{\theta}}_n \in \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \mathcal{L}_n(\boldsymbol{\theta}) \xrightarrow[n \rightarrow +\infty]{P} \boldsymbol{\theta}_0, \quad \hat{\boldsymbol{\theta}}_{n,c} \in \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \mathcal{L}_{n,c}(\boldsymbol{\theta}) \xrightarrow[n \rightarrow +\infty]{P|Y \in \mathcal{E}_\kappa} \boldsymbol{\theta}_0.$$

Asymptotic normality of the MLE & cMLE

- For instance, we focus on estimating a single variance parameter σ_0^2 , i.e.

$$k_{\sigma_0^2}(x, x') = \sigma_0^2 k_1(x, x'),$$

with fixed known correlation function k_1 .

Theorem (Asymptotic normality of the MLE and cMLE)

- Assume *mild conditions*. Then, the MLE $\hat{\sigma}_n^2$ of σ_0^2 conditioned on $\{Y \in \mathcal{E}_\kappa\}$ is asymptotically Gaussian distributed. More precisely, for $\kappa = 0, 1, 2$,

$$\sqrt{n} \left(\hat{\sigma}_n^2 - \sigma_0^2 \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}|Y \in \mathcal{E}_\kappa} \mathcal{N}(0, 2\sigma_0^4).$$

- Furthermore, the cMLE $\hat{\sigma}_{n,c}^2$ of σ_0^2 conditioned on $\{Y \in \mathcal{E}_\kappa\}$ is also asymptotically Gaussian distributed:

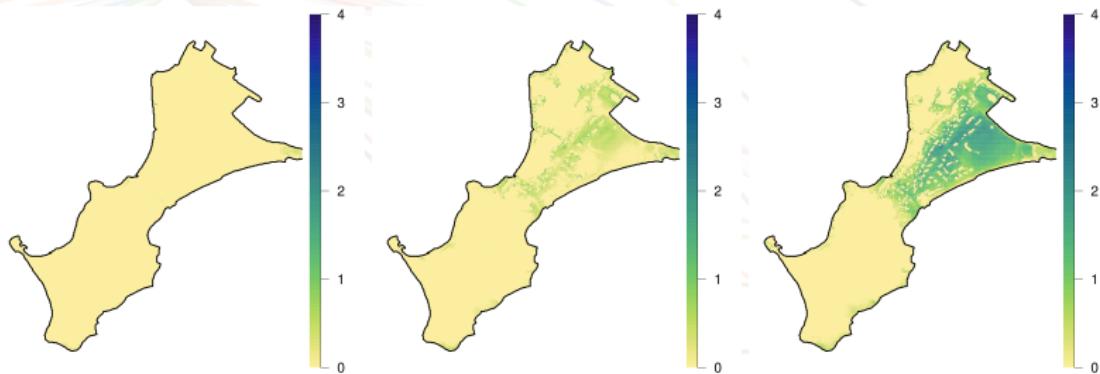
$$\sqrt{n} \left(\hat{\sigma}_{n,c}^2 - \sigma_0^2 \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}|Y \in \mathcal{E}_\kappa} \mathcal{N}(0, 2\sigma_0^4).$$

- The results can be extended for Matérn models

Spatial GPs with functional inputs

Motivation: Coastal flooding assessment

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

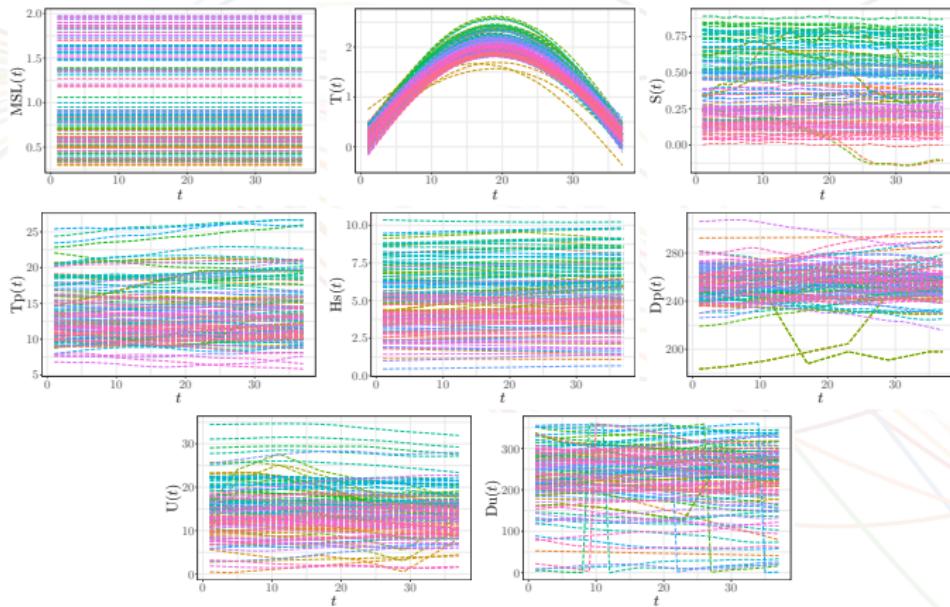


Challenge 1:

- Each flood event takes ~ 3 days of simulation.

Motivation: Coastal flooding assessment

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



Challenge 2:

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

Motivation: Coastal flooding assessment



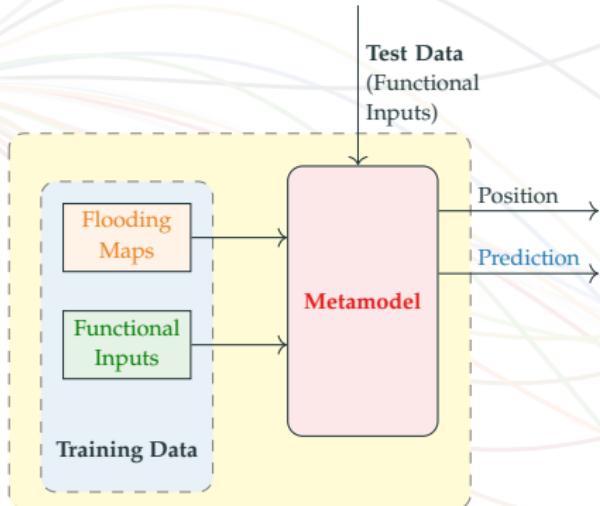
D. Idier BRGM, Orléans
Coastal Engineering



J. Rohmer BRGM, Orléans
Statistical Learning



F. Bachoc IMT, Toulouse
GPs & Applied Maths



- This will lead to faster (approximate) predictions.

Application: forecasting and early warning systems.

A. F. López-Lopera, D. Idier, J. Rohmer and F. Bachoc (2022). Multioutput Gaussian processes with functional data: A study on coastal flood hazard assessment, Reliability Engineering and System Safety, 218.

Spatial Gaussian processes with functional inputs

- According to the coastal flooding application, we need to consider functions

$$y(\mathbf{x}, \mathcal{F}) = \begin{cases} \mathbb{R} \times \mathcal{F}(\mathcal{T}, \mathbb{R})^Q & \rightarrow \mathbb{R}, \\ (\mathbf{x}, \mathcal{F}) & \mapsto y(\mathbf{x}, \mathcal{F}), \end{cases}$$

where

- $\mathcal{F}(\mathcal{T}, \mathbb{R})$ is the set of functions from $\mathcal{T} \subseteq \mathbb{R}$ to \mathbb{R} ,
 - $\mathcal{F} = (f_1, \dots, f_Q)$ are the functional inputs,
 - $\mathbf{x} = (x_1, x_2)$ is the spatial location.
- We can assume the GP $\{Y(\mathbf{x}, \mathcal{F}); \mathbf{x} \in \mathbb{R}^2, \mathcal{F} \in \mathcal{F}(\mathcal{T}, \mathbb{R})^Q\}$ with kernel k :

$$k((\mathbf{x}, \mathcal{F}), (\mathbf{x}', \mathcal{F}')) = \text{cov} \{ Y(\mathbf{x}, \mathcal{F}), Y(\mathbf{x}', \mathcal{F}') \} = k_s(\mathbf{x}, \mathbf{x}') k_f(\mathcal{F}, \mathcal{F}'), \quad (11)$$

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

Spatial Gaussian processes with functional inputs

- For k_f , we need a *measure of “dissimilarity”*, e.g.:

$$d(\mathcal{F}, \mathcal{F}') = \|\mathcal{F} - \mathcal{F}'\|_{\ell} = \sqrt{\sum_{i=1}^Q \|\mathbf{f}_i - \mathbf{f}'_i\|_{\ell_i}^2}, \quad (12)$$

with the L^2 -norm given by

$$\|\mathbf{f}_i - \mathbf{f}'_i\|_{\ell_i}^2 = \frac{\int_{\mathcal{T}} (\mathbf{f}_i(t) - \mathbf{f}'_i(t))^2 dt}{\ell_i^2}. \quad (13)$$

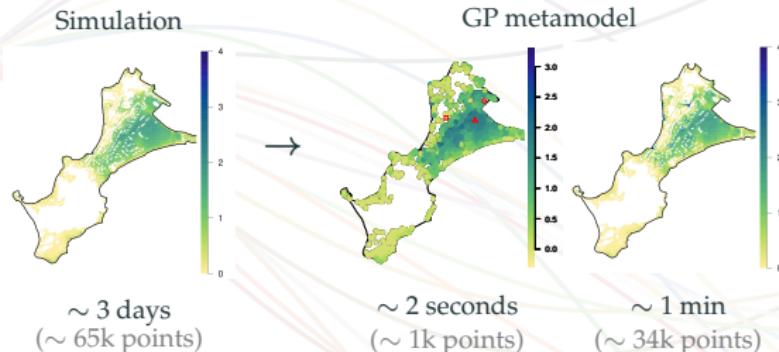
- Then, an example of a valid kernel is then given by

(Squared Exponential) $k_{f, \sigma^2, \ell}(\mathcal{F}, \mathcal{F}') = \sigma^2 \exp \left\{ -\frac{\|\mathcal{F} - \mathcal{F}'\|_{\ell}^2}{2} \right\}. \quad (14)$

Spatial GPs with functional inputs

Advantages

- Approximate predictions under **smoothness assumptions**.



- The proposed GP framework can be scaled to large datasets via:
 - *Kronecker-products*
 - *Sparse-variational inference*

Challenges:

- Non-stationary flooding maps (i.e. non-negativity, discontinuities due to presence of buildings)

Perspectives

- ★ Within a mechanical context, the extension to multivariate functional inputs is being studied in the ANR JCJC GAME project

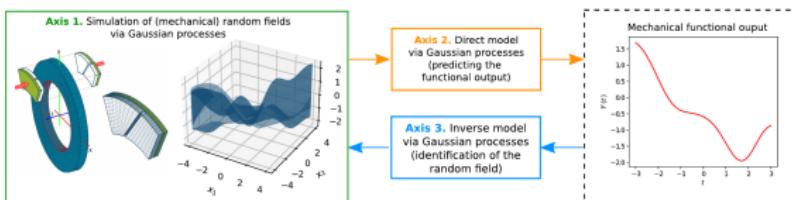


Figure 1: General framework proposed in the ANR JCJC GAME project for the automotive application in Section 1.1.3.

- ★ The interest also relies on theoretical guarantees (asymptotic consistency and normality of the ML estimator) of GPs with functional inputs

■ ANR JCJC GAME – GAussian process modeling of transient MEchanical random fields: a complete study from simulation to identification (2023 - 2027). PI: A. F. López-Lopera.
URL: <https://anfelopera.github.io/funding/GAME/>

■ A. F. López-Lopera, F. Massa, I. Turpin, and N. Leconte (2022). Modeling complex mechanical computer codes with functional input via Gaussian processes. In The XLIII Ibero-Latin American Congress on Computational Methods in Engineering (CILAMCE).

■ L. Reding, A. F. López-Lopera, and F. Bachoc (2024). Asymptotic analysis for covariance parameter estimation of Gaussian processes with functional inputs. Submitted in Electronic Journal of Statistics.



L. Reding

CERAMATHS
GPs & Applied Maths



I. Turpin

CERAMATHS
Applied Maths



R. Le Riche

EMSE
GPs & Optimization



F. Massa

LAMIH
Mechanics



J. Bruchon

EMSE
Mechanics

Multi-fidelity GPs

Motivation: Aerodynamic data-fusion



- Data fusion (DF)-based frameworks aim at jointly treating data acquisition schemes while accounting for their corresponding levels of fidelity

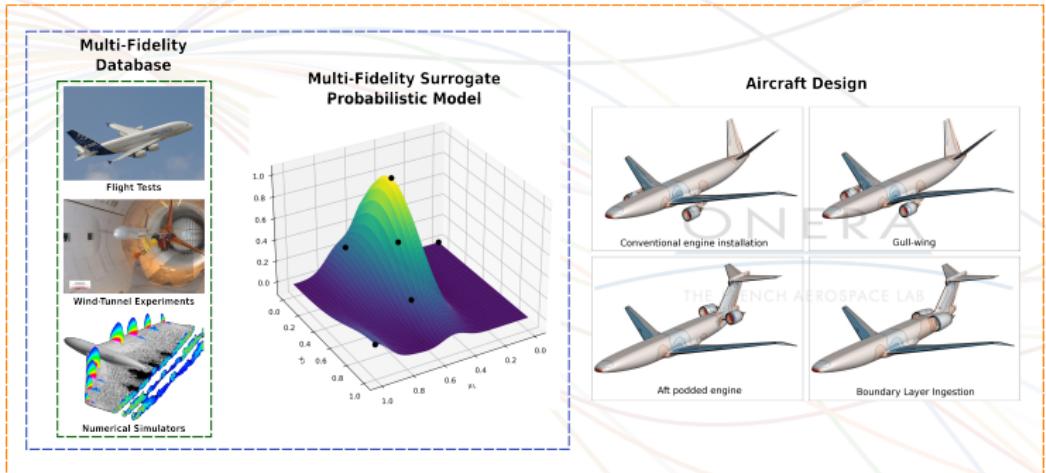
N. Bartoli
ONERA
GPs & Optimization



S. Mouton
ONERA
Aerodynamics



T. Lefebvre
ONERA
Aerodynamics



Multi-fidelity DF-based architecture

Acquisition Scheme	Level of Fidelity	Data Availability	Cost
flight tests	high	very low	expensive
wind-tunnel tests	upper-intermediate	intermediate	moderate
simulators	low or intermediate	high	cheap or moderate

Multi-fidelity model based on GPs

- In aerodynamics, for two levels of fidelity, it is often considered the autoregressive model:

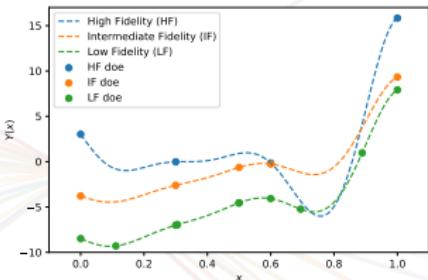
$$Y_1(\boldsymbol{x}) = \rho(\boldsymbol{x})Y_0(\boldsymbol{x}) + \delta(\boldsymbol{x}), \quad (15)$$

where

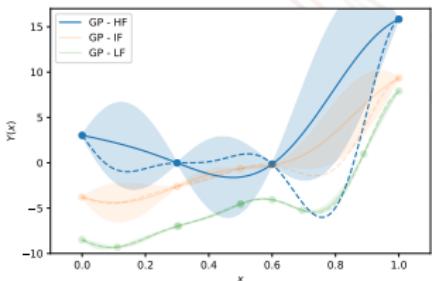
- $Y_0 \sim \mathcal{GP}(m_0, k_0)$ and $\delta(\boldsymbol{x}) \sim \mathcal{GP}(m_\delta, k_\delta)$
- $\rho : \mathcal{D} \rightarrow \mathbb{R}$ is a scale factor between Y_1 and Y_0
- $\delta : \mathcal{D} \rightarrow \mathbb{R}$ is the discrepancy function tasked with capturing the differences between Y_1 and Y_0 beyond scaling
- If $Y_0 \perp \delta$, then Y_1 is also GP-distributed
- Note that knowledge from Y_0 is transferred to learn Y_1 (**transfer learning**)
- (15) can be generalized (recursively) to $\ell > 2$ levels of fidelity

■ R. Conde-Arenzana, A. F. López-Lopera, S. Mouton, N. Bartoli and T. Lefebvre (2021). Multi-fidelity Gaussian process model for CFD and wind tunnel data fusion. Proceedings of the AeroBest conference, 2021.

Multi-fidelity model based on GPs



(a) True functions and design points

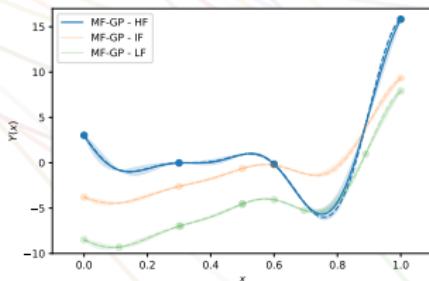


(b) Independent GP models

- The nested design of experiment (DoE) contains:

- 4 high-fidelity (HF) design points
- 5 intermediate-fidelity (IF) design points
- 8 low-fidelity (LF) design points

- Nested DoEs are required for efficient implementations [Le Gratiet, 2013]



(c) Multi-fidelity GP model

1D multi-fidelity regression example with 3 levels of fidelity

- ★ The extension to large datasets is being investigated by Mauricio Castaño-Aguirre (PhD student at the UPHF-ONERA). We seek to study:
 - Sparse-variational inference without requiring nested DoEs
 - Multi-fidelity active learning (global metamodeling) for noisy data

 H. Valayer, N. Bartoli, M. Castaño-Aguirre, R. Lafage, T. Lefebvre, A. F. López-Lopera, and S. Mouton (2024). A Python toolbox for data-driven aerodynamic modeling using sparse Gaussian processes. *Aerospace*, 11(4).

Multi-fidelity model based on GPs



N. Bartoli

ONERA



GPs & Optimization

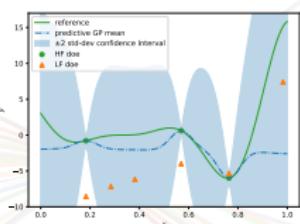


F. Massa

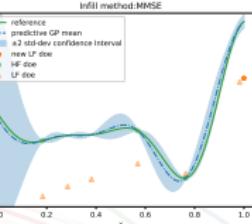
LAMIH



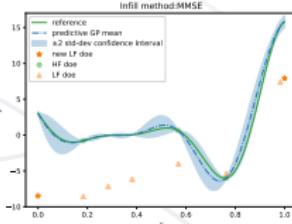
Mechanics



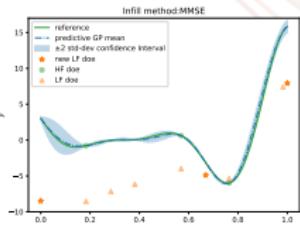
(a) Iter. 0



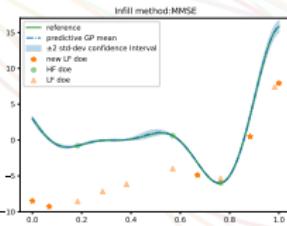
(b) Iter. 1: new LF point



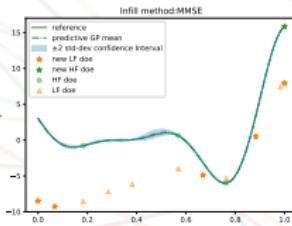
(c) Iter. 2: new LF point



(d) Iter. 3: new LF point



(e) Iter. 5: new LF point

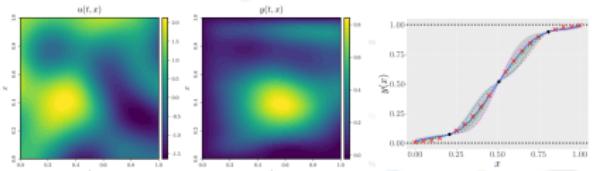


(f) Iter. 6: new HF point

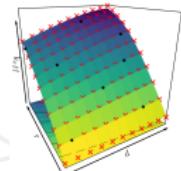
Active learning example. The panels show the resulting multi-fidelity GP model at the high-fidelity (HF) level for different iterations of the sequential adaptive algorithm.

- As an example of multi-fidelity active learning, we may consider two steps:
 - Compute $x_{n+1} = \operatorname{argmax}_{x \in \mathcal{D}} \text{MMSE}_n(x)$ for the HF level
 - Define a criterion for the selection of the level of fidelity (**open question**)

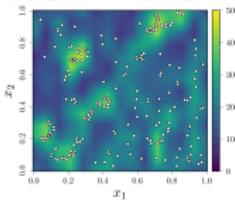
Conclusions



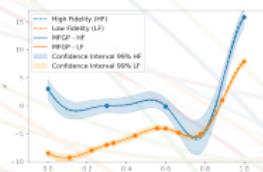
(a) Differential equations



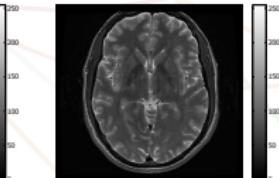
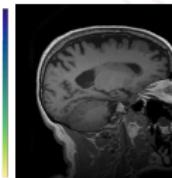
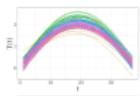
(b) Inequality constraints



(c) Point processes



(d) Multi-fidelity



(e) Functional data

(f) Super-resolution of MR images

References

- SMT: Surrogate Modeling Toolbox, 2024. Python toolbox version 2.5.0 available at:
<https://github.com/SMTorg/smt>.
- M. Álvarez, D. Luengo, and N. Lawrence. Latent force models. In *International Conference on Artificial Intelligence and Statistics*, 2009.
- F. Bachoc, A. Lagnoux, and **AFLL**. Maximum likelihood estimation for Gaussian processes under inequality constraints. *Electronic Journal of Statistics*, 13(2), 2019.
- F. Bachoc, **AFLL**, and O. Roustant. Sequential construction and dimension reduction of Gaussian processes under inequality constraints. *SIAM Journal on Mathematics of Data Science*, 4(2): 772–800, 2022.
- X. Bay, L. Grammont, and H. Maatouk. Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *Electronic Journal of Statistics*, 2016.
- R. Conde-Arenzana, **AFLL**, S. Mouton, N. Bartoli, and T. Lefebvre. Multi-fidelity Gaussian process model for CFD and wind tunnel data fusion. In *AeroBest 2021*, 2021.
- M. Deronzier, **AFLL**, F. Bachoc, and O. Roustant. Block additive Gaussian processes under monotonicity constraints: Covariance structure and model selection. In preparation for submission to a journal in Statistics (2024).

- N. Durrande, D. Ginsbourger, and O. Roustant. Additive covariance kernels for high-dimensional Gaussian process modeling. *Annales de la Faculté de Sciences de Toulouse*, 21(3):481–499, 2012.
- P. Hegde, M. Heinonen, and S. Kaski. Variational zero-inflated Gaussian processes with sparse kernels. In *Conference on Uncertainty in Artificial Intelligence*, pages 361–371, 2018.
- D. Idier, A. Aurouet, F. Bachoc, A. Baills, J. Betancourt, F. Gamboa, T. Klein, **AFLL**, R. Pedreros, J. Rohmer, and A. Thibault. A user-oriented local coastal flooding early warning system using metamodelling techniques. *Journal of Marine Science and Engineering*, 9(11), 2021.
- L. Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. Theses, Université Paris-Diderot - Paris VII, October 2013.
- H. Maatouk and X. Bay. Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences*, 2017.
- A. Pakman and L. Paninski. Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics*, 2014.
- M. Papez and A. Quinn. Bayesian transfer learning between Gaussian process regression tasks. In *IEEE International Symposium on Signal Processing and Information Technology*, pages 1–6, 2019.
- T.V.E. Perrin, O. Roustant, J. Rohmer, O. Alata, J.P. Naulin, D. Idier, R. Pedreros, D. Moncoulon, and P. Tinard. Functional principal component analysis for global sensitivity analysis of model with spatial output. *Reliability Engineering & System Safety*, 211:107522, 2021.
- S. Petit, F. Zaoui, A.-L Popelin, C. Goeury, and N. Goutal. Couplage entre indices à base de dérivées et mode adjoint pour l’analyse de sensibilité globale. Application sur le code Mascaret. HAL e-prints, September 2016.
- C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for machine learning (adaptive computation and machine learning)*. The MIT Press, Cambridge, MA, 2005.

- L. Reding, **AFLL**, and F. Bachoc. Asymptotic analysis for covariance parameter estimation of Gaussian processes with functional inputs. Submitted in *Electronic Journal of Statistics* (2024). URL <https://arxiv.org/abs/2404.17222>.
- B. D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society (Series B)*, 39(2), 1977.
- AFLL**. *Gaussian Process Modelling under Inequality Constraints*. Theses, Univ. de Lyon, 09/2019. URL <https://tel.archives-ouvertes.fr/tel-02863891>.
- AFLL**. *PhysicallyGPDrosophila: Physically-inspired Gaussian process models for post-transcriptional regulation in Drosophila*, 2018a. Matlab toolbox available at:
<https://github.com/anfelopera/PhysicallyGPDrosophila>.
- AFLL**. *SDLFM_ReverseEngineering: Switched latent force model for reverse-engineering transcriptional regulation in gene expression data*, 2018b. Matlab toolbox available at:
https://github.com/anfelopera/SDLFM_ReverseEngineering.
- AFLL**. *lineqGPR: Gaussian process regression models with linear inequality constraints*, 2019. R package version 0.3.0 available at: <https://github.com/anfelopera/lineqGPR>.
- AFLL**. *spatfGPs: Spatial Gaussian processes with functional inputs*, 2020. R package and Python library available at: <https://github.com/anfelopera/spatfGPs>.
- AFLL** and M. Álvarez. Switched latent force models for reverse-engineering transcriptional regulation in gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(1), Jan 2019.
- AFLL** and A. Pakman. *t-processes under inequality constraints via exact Hamiltonian Monte Carlo*. In preparation for submission to a (top 5) machine learning conference (2024).
- AFLL**, F. Bachoc, N. Durrande, and O. Roustant. Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3), 2018.

- AFLL**, F. Bachoc, N. Durrande, J. Rohmer, D. Idier, and O. Roustant. Approximating Gaussian process emulators with linear inequality constraints and noisy observations via MC and MCMC. In *International Conference in Monte Carlo & Quasi-Monte Carlo Methods (MCQMC)*, Jan 2019a.
- AFLL**, N. Durrande, and M. Álvarez. Physically-inspired Gaussian process models for post-transcriptional regulation in *Drosophila*. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019b.
- AFLL**, ST John, and N. Durrande. Gaussian process modulated Cox processes under linear inequality constraints. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 16–18 Apr 2019c.
- AFLL**, F. Bachoc, and O. Roustant. High-dimensional additive Gaussian processes under monotonicity constraints. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 8041–8053, 2022a.
- AFLL**, D. Idier, J. Rohmer, and F. Bachoc. Multioutput Gaussian processes with functional data: A study on coastal flood hazard assessment. *Reliability Engineering & System Safety*, 218:108139, 2022b.
- AFLL**, F. Massa, I. Turpin, and N. Leconte. Modeling complex mechanical computer codes with functional input via Gaussian processes. In *The XLIII Ibero-Latin American Congress on Computational Methods in Engineering (CILAMCE)*, 2022c.
- H. Valayer, N. Bartoli, M. Castaño Aguirre, R. Lafage, T. Lefebvre, **AFLL**, and S. Mouton. A python toolbox for data-driven aerodynamic modeling using sparse Gaussian processes. *Aerospace*, 11(4), 2024.
- H. Vargas-Cardona, **AFLL**, A. Orozco, M. Álvarez, J. Hernández-Tamames, and N. Malpica. Gaussian processes for slice-based super-resolution MR images. In *International Symposium on Advances in Visual Computing (ISVC)*, 2015.

Constrained GPs

The maximum a posteriori (mode) function in 1D

- Let $\hat{\xi}$ be the mode that maximizes the pdf of $\xi | \{\Phi \xi + \varepsilon = y, l \leq \Lambda \xi \leq u\}$:

$$\hat{\xi} = \underset{\xi \text{ s.t. } l \leq \Lambda \xi \leq u}{\arg \max} \{-[\xi - \mu_c]^\top \Sigma_c^{-1} [\xi - \mu_c]\}, \quad (16)$$

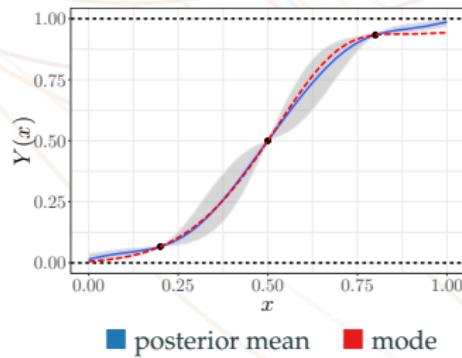
with $\hat{\xi} = [\hat{\xi}_1, \dots, \hat{\xi}_m]^\top$.

- The MAP estimate of Y_S is given by

$$\hat{Y}_S(x) = \sum_{j=1}^m \hat{\xi}_j \phi_j(x). \quad (17)$$

Pro:

- \hat{Y}_S can be used as a point estimate
- Fast to compute
- Convergence of \hat{Y}_S to the spline solution as $m \rightarrow \infty$ [Bay et al., 2016]
- Starting point for MCMC methods



Asymptotic normality of the MLE & cMLE

- For instance, we focus on estimating a single variance parameter σ_0^2 , i.e.

$$k_{\sigma_0^2}(x, x') = \sigma_0^2 k_1(x, x'),$$

with fixed known correlation function k_1 .

Theorem (Asymptotic normality of the MLE and cMLE)

- Assume *mild conditions*. Then, the MLE $\hat{\sigma}_n^2$ of σ_0^2 conditioned on $\{Y \in \mathcal{E}_\kappa\}$ is asymptotically Gaussian distributed. More precisely, for $\kappa = 0, 1, 2$,

$$\sqrt{n} \left(\hat{\sigma}_n^2 - \sigma_0^2 \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}|Y \in \mathcal{E}_\kappa} \mathcal{N}(0, 2\sigma_0^4).$$

- Furthermore, the cMLE $\hat{\sigma}_{n,c}^2$ of σ_0^2 conditioned on $\{Y \in \mathcal{E}_\kappa\}$ is also asymptotically Gaussian distributed:

$$\sqrt{n} \left(\hat{\sigma}_{n,c}^2 - \sigma_0^2 \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}|Y \in \mathcal{E}_\kappa} \mathcal{N}(0, 2\sigma_0^4).$$

- The results can be extended for Matérn models

The MaxMod algorithm in 1D

- Let $\widehat{Y}_{\mathcal{S}}$ be the MAP function with an ordered set of knots:
$$\mathcal{S} = \{t_0, \dots, t_m\}, \quad \text{with } 0 = t_0 < \dots < t_m = 1.$$

- Here, we aim at adding a new knot t in \mathcal{S} (where?)
- To do so, we aim at *maximizing the total modification of the MAP*:

$$I_{\mathcal{S}}(t) = \int_{[0,1]} \left(\widehat{Y}_{\mathcal{S} \cup \{t\}}(x) - \widehat{Y}_{\mathcal{S}}(x) \right)^2 dx. \quad (18)$$

- The integral in (18) has a closed-form expression.

Algorithm MaxMod (maximum modification of the MAP) in 1D

Input parameters: the initial subdivision $S^{(0)} \in \mathcal{S}$.

Sequential procedure: for $\kappa \in \mathbb{N}$, do:

1: Set $t_{\kappa+1}^* \in [0, 1]$ such that

$$I_{S^{(\kappa)}}(t_{\kappa+1}^*) \geq \sup_{t \in [0,1]} I_{S^{(\kappa)}}(t)$$

2: $S^{(\kappa+1)} = S^{(\kappa)} \cup t_{\kappa+1}^*$.

The MaxMod algorithm in 1D

1D example under boundedness and monotonicity constraints

MAP estimate

conditional sample-path

- training points
- predictive mean
- + knots
- MAP estimate
- 90% confidence intervals

The MaxMod algorithm in higher dimensions

- Let $\widehat{Y}_{\mathcal{J}, \mathbf{S}}$ be the MAP function with $|\mathcal{J}|$ active variables and ordered sets of knots $\mathbf{S}_{\mathcal{J}}$ for $\mathcal{J} \subseteq \{1, \dots, D\}$.
- Then, the criterion to maximise is given by

$$I_{\mathcal{J}, \mathbf{S}}(\mathbf{i}, \mathbf{t}) = \begin{cases} \frac{1}{N_{\mathbf{S}, \mathcal{J}, i}} \int_{[0,1]^d} (\widehat{Y}_{\mathcal{J}, \mathbf{S} \cup_i \mathbf{t}}(\mathbf{x}) - \widehat{Y}_{\mathcal{J}, \mathbf{S}}(\mathbf{x}))^2 d\mathbf{x} & \text{if } i \in \mathcal{J}, \\ \frac{1}{N_{\mathbf{S}, \mathcal{J}, i}} \int_{[0,1]^{d+1}} (\widehat{Y}_{\mathcal{J} \cup \{i\}, \mathbf{S} + \mathbf{i}}(\mathbf{x}) - \widehat{Y}_{\mathcal{J}, \mathbf{S}}(\mathbf{x}))^2 d\mathbf{x} & \text{if } i \notin \mathcal{J}, \end{cases} \quad (19)$$

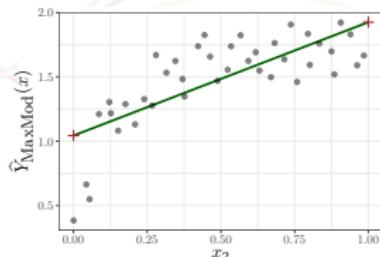
where $N_{\mathbf{S}, \mathcal{J}, i}$ is the increase of the number of basis functions.

• F. Bachoc, A. F. López-Lopera, and O. Roustant (2022). Sequential construction and dimension reduction of Gaussian processes under inequality constraints. SIAM Journal on Mathematics of Data Science, 4(2).

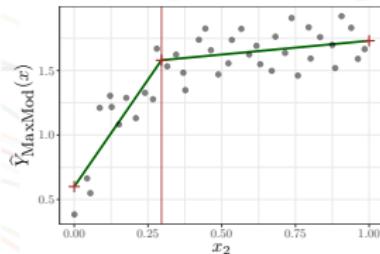
The MaxMod algorithm in higher dimensions

2D example under monotonicity constraints

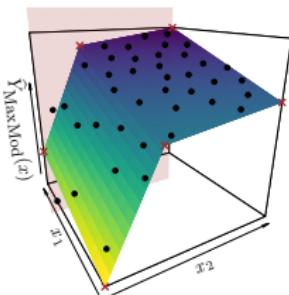
Evolution of the MaxMod algorithm using $f(x) = \frac{1}{2}x_1 + \arctan(10x_2)$



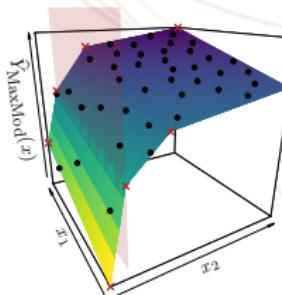
(a) iteration 0



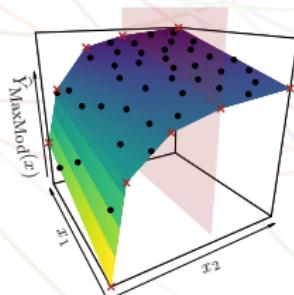
(b) iteration 1



(c) iteration 2

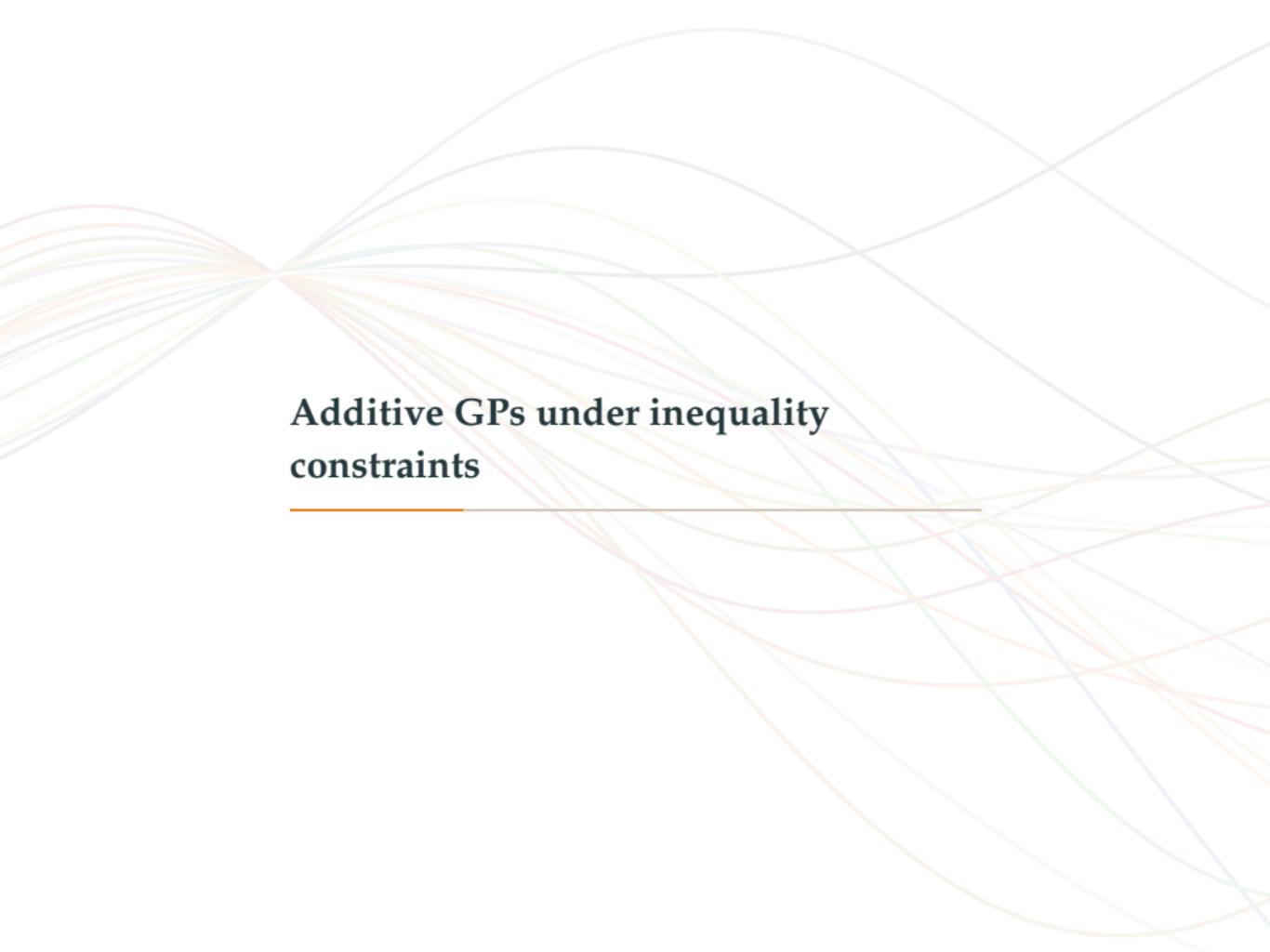


(d) iteration 3



(e) iteration 4

● training points + knots ■ MAP estimate



Additive GPs under inequality constraints

Additive GPs

- In high dimension, many statistical regression models are based on additive structures of the form:

$$y(\mathbf{x}) = y_1(x_1) + \cdots + y_d(x_d). \quad (20)$$

- Then GP priors can be placed over y_1, \dots, y_d [Durrande et al., 2012]

$$Y_i \sim \mathcal{GP}(0, k_i),$$

for $i = 1, \dots, d$. Taking Y_1, \dots, Y_d as independent GPs, the process

$$Y(\mathbf{x}) = Y_1(x_1) + \cdots + Y_d(x_d),$$

is also a GP and its kernel is given by

$$k(\mathbf{x}, \mathbf{x}') = k_1(x_1, x'_1) + \cdots + k_d(x_d, x'_d). \quad (21)$$

Finite-dimensional approximation of additive GPs

- For the constrained case, we can approximate Y_i by a finite-dimensional GP:

$$Y_{i,S_i}(x_i) = \sum_{j=1}^{m_i} \xi_{i,j} \phi_{i,j}(x_i),$$

with one-dimensional subdivision S_i , and m_i knots.

- We let $S = (S_1, \dots, S_d)$. The finite-dimensional GP is written,

$$Y_S(x) = \sum_{i=1}^d Y_{i,S_i}(x_i) = \sum_{i=1}^d \sum_{j=1}^{m_i} \xi_{i,j} \phi_{i,j}(x_i),$$

where $\xi_{i,j} = Y_i(t_{(j)}^{(S_i)})$ and $\phi_{i,j} : [0, 1] \mapsto \mathbb{R}$ are asymmetric hat basis functions.

- One can note that the total number of knots is given by $m = m_1 + \dots + m_d$

Finite-dimensional approximation of additive GPs

- We consider the componentwise constraints $Y_{i,S_i} \in \mathcal{E}_i, i = 1, \dots, d$ such that

$$Y_{i,S_i} \in \mathcal{E}_i \Leftrightarrow \boldsymbol{\xi}_i \in \mathcal{C}_i$$

where $\boldsymbol{\xi}_i = [\xi_{i,1}, \dots, \xi_{i,m_i}]^\top$ and $\mathcal{C}_i = \{\mathbf{c} \in \mathbb{R}^{m_i} : \mathbf{l}_i \leq \mathbf{\Lambda}_i \mathbf{c} \leq \mathbf{u}_i\}$.

- Examples of constraints are monotonicity and componentwise convexity
- Given the observations and the constraints, the MAP estimate is given by

$$\widehat{Y}_S(\mathbf{x}) = \sum_{i=1}^d \sum_{j=1}^{m_i} \widehat{\xi}_{i,j} \phi_{i,j}(\mathbf{x}_i).$$

Additive MaxMod algorithm

- Consider an additive cGP model that uses only a subset $\mathcal{J} \subseteq \{1, \dots, d\}$ of active variables
- Its mode function \widehat{Y}_S , from $\mathbb{R}^{|\mathcal{J}|}$ to \mathbb{R} , by, for $\mathbf{x} = (x_i; i \in \mathcal{J})$,

$$\widehat{Y}_S(\mathbf{x}) = \sum_{i \in \mathcal{J}} \sum_{j=1}^{m_i} \widehat{\xi}_{i,j} \phi_{i,j}(x_i).$$

- We measure this benefit by the squared-norm modification of the cGP mode

$$I_{S,i^*} = \int_{[0,1]^{|\mathcal{J}|+1}} \left(\widehat{Y}_{S,i^*}(\mathbf{x}) - \widehat{Y}_S(\mathbf{x}) \right)^2 d\mathbf{x} \text{ for } i^* \notin \mathcal{J}, \quad (22)$$

$$I_{S,i^*,t} = \int_{[0,1]^{|\mathcal{J}|}} \left(\widehat{Y}_{S,i^*,t}(\mathbf{x}) - \widehat{Y}_S(\mathbf{x}) \right)^2 d\mathbf{x} \text{ for } i^* \in \mathcal{J}.$$

- (22) has an analytic expression assuming $x_i \sim \text{Uniform}(0, 1)$ for $i = 1, \dots, d$ [AFLL et al., 2022a], where the computational cost is linear w.r.t.

$$m = \sum_{i \in \mathcal{J}} m_i$$

Additive GPs under inequality constraints

- For a new variable $i^* \notin \mathcal{J}$, the new mode function is

$$\widehat{Y}_{S,i^*}(\mathbf{x}) = \sum_{i \in \mathcal{J}} \sum_{j=1}^{m_i} \widetilde{\xi}_{i,j} \phi_{i,j}(x_i) + \sum_{j=1}^2 \widetilde{\xi}_{i^*,j} \phi_{i^*,j}(x_{i^*})$$

- We let $\phi_{i^*,1}(u) = 1 - u$ and $\phi_{i^*,2}(u) = u$ for $u \in [0, 1]$.

Proposition (Computation of I_{S,i^*})

We have

$$I_{S,i^*} = \sum_{i \in \mathcal{J}} \sum_{\substack{j,j'=1 \\ |j-j'| \leq 1}}^{m_i} \eta_{i,j} \eta_{i,j'} E_{j,j'}^{(S_i)} - \sum_{i \in \mathcal{J}} \left(\sum_{j=1}^{m_i} \eta_{i,j} E_j^{(S_i)} \right)^2 + \frac{\eta_{i^*}^2}{12} + \left(\sum_{i \in \mathcal{J}} \sum_{j=1}^{m_i} \eta_{i,j} E_j^{(S_i)} - \frac{\zeta_{i^*}}{2} \right)^2,$$

where $\eta_{i,j} = \widehat{\xi}_{i,j} - \widetilde{\xi}_{i,j}$, $\eta_{i^*} = \widetilde{\xi}_{i^*,2} - \widetilde{\xi}_{i^*,1}$, $\zeta_{i^*} = \widetilde{\xi}_{i^*,1} + \widetilde{\xi}_{i^*,2}$, $E_j^{(S_i)} := \int_0^1 \phi_{i,j}(t) dt$ and $E_{j,j'}^{(S_i)} := \int_0^1 \phi_{i,j}(t) \phi_{i,j'}(t) dt$ with explicit expressions in Lemma 1 [AFLL et al., 2022a, Appendix A.3]. The matrices $(E_{j,j'}^{(S_i)})_{1 \leq j,j' \leq m_i}$ are 1-band and the computational cost is linear w.r.t. $m = \sum_{i \in \mathcal{J}} m_i$.

Additive MaxMod algorithm

- For a new t added to S_{i^*} with $i^* \in \mathcal{J}$, the new mode function is

$$\widehat{Y}_{S,i^*,t}(x) = \sum_{i \in \mathcal{J}} \sum_{j=1}^{\tilde{m}_i} \tilde{\xi}_{i,j} \tilde{\phi}_{i,j}(x_i),$$

where $\tilde{m}_i = m_i$ for $i \neq i^*$, $\tilde{m}_{i^*} = m_{i^*} + 1$, $\tilde{\phi}_{i,j} = \phi_{i,j}$ for $i \neq i^*$, and $\tilde{\phi}_{i^*,j}$ is obtained from $S_{i^*} \cup \{t\}$ as in Proposition 4.

Proposition (Computation of $I_{S,i^*,t}$)

For $i \in \mathcal{J} \setminus \{i^*\}$, let $\tilde{S}_i = S_i$. Let $\tilde{S}_{i^*} = S_{i^*} \cup \{t\}$. Recall that the knots in S_{i^*} are written $0 = t_{(1)}^{(S_{i^*})} < \dots < t_{(m_{i^*})}^{(S_{i^*})} = 1$. Let $\nu \in \{1, \dots, m_{i^*} - 1\}$ be such that $t_{(\nu)}^{(S_{i^*})} < t < t_{(\nu+1)}^{(S_{i^*})}$. Then, with a linear cost w.r.t. $\tilde{m} = \sum_{i \in \mathcal{J}} \tilde{m}_i$, we have

$$I_{S,i^*,t} = \sum_{i \in \mathcal{J}} \sum_{\substack{j,j'=1 \\ |j-j'| \leq 1}}^{\tilde{m}_i} \bar{\eta}_{i,j} \bar{\eta}_{i,j'} E_{j,j'}^{(\tilde{S}_i)} - \sum_{i \in \mathcal{J}} \left(\sum_{j=1}^{\tilde{m}_i} \bar{\eta}_{i,j} E_j^{(\tilde{S}_i)} \right)^2 + \left(\sum_{i \in \mathcal{J}} \sum_{j=1}^{\tilde{m}_i} \bar{\eta}_{i,j} E_j^{(\tilde{S}_i)} \right)^2,$$

where $\bar{\eta}_{i,j} = \bar{\xi}_{i,j} - \tilde{\xi}_{i,j}$, $\bar{\xi}_{i,j} = \widehat{\xi}_{i,j}$ for $i \neq i^*$, $\bar{\xi}_{i^*,j} = \widehat{\xi}_{i^*,j}$ for $j \leq \nu$, $\bar{\xi}_{i^*,j} = \widehat{\xi}_{i^*,j-1}$ for $j \geq \nu + 2$, and

$$\bar{\xi}_{i^*,\nu+1} = \widehat{\xi}_{i^*,\nu} \frac{t_{(\nu+1)}^{(S_{i^*})} - t}{t_{(\nu+1)}^{(S_{i^*})} - t_{(\nu)}^{(S_{i^*})}} + \widehat{\xi}_{i^*,\nu+1} \frac{t - t_{(\nu)}^{(S_{i^*})}}{t_{(\nu+1)}^{(S_{i^*})} - t_{(\nu)}^{(S_{i^*})}}.$$

Numerical experiments: Monotonicity in hundreds of dimensions

- We consider the target function:

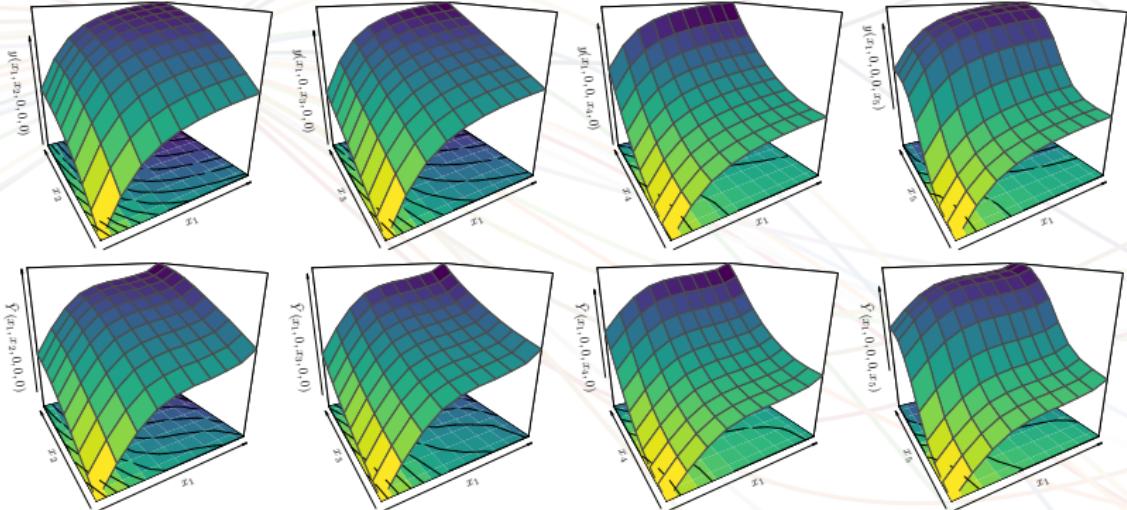
$$y(x) = \sum_{i=1}^d \arctan \left(5 \left[1 - \frac{i}{d+1} \right] x_i \right).$$

with $x \in [0, 1]^d$. y exhibits decreasing growth rates as the index i increases.

Results (mean \pm one standard deviation over 10 replicates) with $n = 2d$. For the computation of the cGP mean, 10^3 ($\dagger 50$) HMC samples are used.

d	m	CPU Time [s]		Q^2 [%]		
		cGP mode	cGP mean	GP mean	cGP mode	cGP mean
10	50	0.1 ± 0.1	0.1 ± 0.1	82.3 ± 6.2	83.8 ± 4.2	88.1 ± 1.7
100	500	0.4 ± 0.1	5.2 ± 0.5	89.8 ± 1.6	90.7 ± 1.4	91.5 ± 1.3
250	1250	4.2 ± 0.7	132.3 ± 26.3	91.7 ± 0.8	92.9 ± 0.6	93.4 ± 0.6
500	2500	37.0 ± 11.4	$\dagger 156.9 \pm 40.5$	92.5 ± 0.6	93.8 ± 0.5	$\dagger 94.3 \pm 0.5$
1000	5000	262.4 ± 35.8	$\dagger 10454.3 \pm 3399.3$	92.6 ± 0.3	94.6 ± 0.2	$\dagger 95.1 \pm 0.2$

Numerical experiments: Monotonicity in hundreds of dimensions



2D projections of the true profiles (top) and the constrained GP predictions (bottom)

Numerical experiments: Dimension reduction illustration

- We test the capability of MaxMod to account for dimension reduction considering the function in (13).
- In addition to (x_1, \dots, x_d) , we include $D - d$ virtual variables, indexed as (x_{d+1}, \dots, x_D) , which will compose the subset of inactive dimensions.
 - \hat{Y}_{MaxMod} : the mode of the additive cGP and MaxMod.
 - $\tilde{Y}_{\text{MaxMod}}$: the mode of the non-additive cGP and MaxMod.

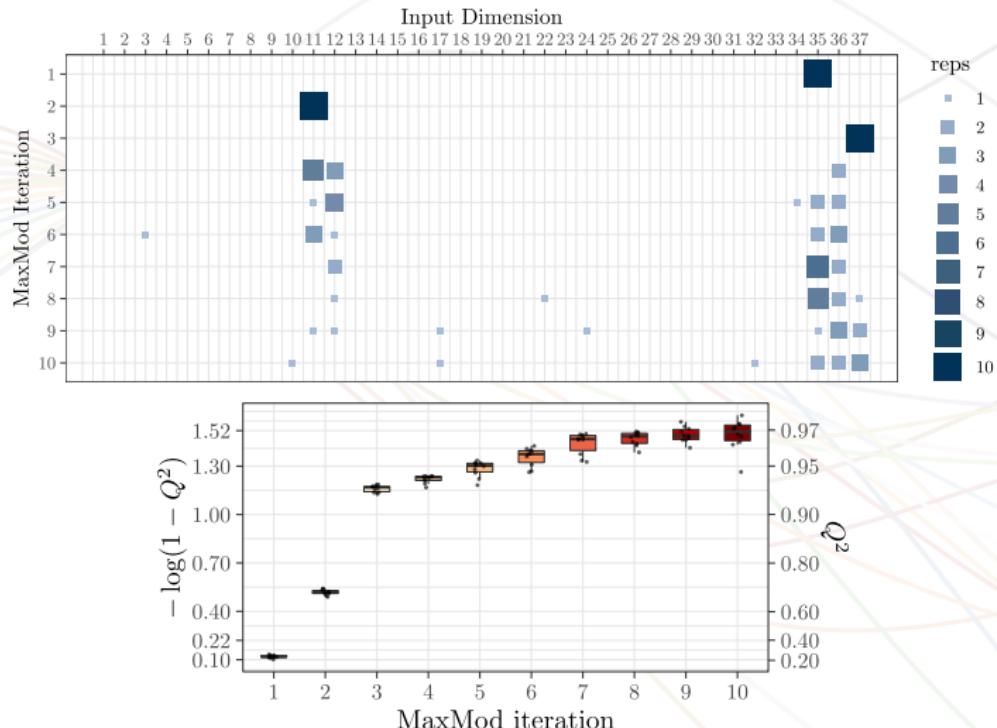
Q^2 Performance of the MaxMod algorithm with $n = 10D$.

D	d	active dimensions	knots per dimension	$Q^2(\tilde{Y}_{\text{MaxMod}}) [\%]$	$Q^2(\hat{Y}_{\text{MaxMod}}) [\%]$
10	2	(1, 2)	(4, 3)	99.5	99.8
	3	(1, 2, 3)	(5, 5, 3)	97.8	99.8
	5	(1, 2, 3, 4, 5)	(4, 4, 4, 3, 2)	91.4	99.8
20	2	(1, 2)	(5, 3)	99.7	99.8
	3	(1, 2, 3)	(4, 4, 3)	99.0	99.9
	5	(1, 2, 3, 4, 5)	(5, 4, 3, 3, 2)	96.0	99.7

Numerical experiments: Flood study of the Vienne river

- The database contains a flood study conducted by the French multinational electric utility company EDF in the Vienne river [Petit et al., 2016].
- It is composed of $N = 2 \times 10^4$ simulations.
 - 1 output: water level H
 - 37 inputs depending on: a value of flow upstream, data on the geometry of the bed, and Strickler friction coefficients
- It is possible to identify that H is decreasing along the first 24 input dimensions and increasing along dimension 37.
- According to expert knowledge, the additive assumption is realistic here, and that inputs 11, 35 and 37 explain most of the variance.
- We consider (approximated) LHD of size $n = 2d$ for training the cGP.

Additive GPs under inequality constraints: Flood study of the Vienne river



The choice made by MaxMod per iteration (top) and Q^2 boxplots (bottom). Results are computed over 10 replicates. For the first panel, a bigger and darker square implies a more repeated choice.

Spatial GPs with functional inputs

Spatial Gaussian processes with functional inputs

	$k(\mathcal{F}, \mathcal{F}')$			
\mathcal{F}_1	1.0	0.85	0.77	0.73
\mathcal{F}_2	0.85	1.0	0.98	0.96
\mathcal{F}_3	0.77	0.98	1.0	1.0
\mathcal{F}_4	0.73	0.96	1.0	1.0
\mathcal{F}_1	1.0	0.78	0.69	0.64
\mathcal{F}_2	0.78	1.0	0.97	0.94
\mathcal{F}_3	0.69	0.97	1.0	0.99
\mathcal{F}_4	0.64	0.94	0.99	1.0

(a) Squared Exponential

	$k(\mathcal{F}, \mathcal{F}')$			
\mathcal{F}_1	1.0	0.78	0.69	0.64
\mathcal{F}_2	0.78	1.0	0.97	0.94
\mathcal{F}_3	0.69	0.97	1.0	0.99
\mathcal{F}_4	0.64	0.94	0.99	1.0

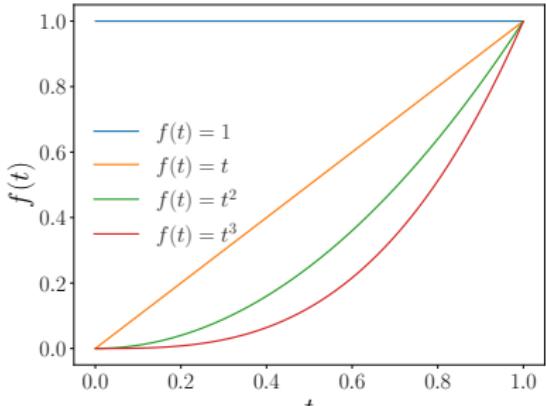
(b) Matérn 5/2

	$k(\mathcal{F}, \mathcal{F}')$			
\mathcal{F}_1	1.0	0.74	0.64	0.6
\mathcal{F}_2	0.74	1.0	0.96	0.92
\mathcal{F}_3	0.64	0.96	1.0	0.99
\mathcal{F}_4	0.6	0.92	0.99	1.0

(c) Matérn 3/2

	$k(\mathcal{F}, \mathcal{F}')$			
\mathcal{F}_1	1.0	0.47	0.43	0.41
\mathcal{F}_2	0.47	1.0	0.65	0.59
\mathcal{F}_3	0.43	0.65	1.0	0.73
\mathcal{F}_4	0.41	0.59	0.73	1.0

(d) Exponential



(e) target functions

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

Spatial Gaussian processes with functional inputs

Projection of the functional inputs onto basis functions. Consider the projection of $\mathcal{F} = (\mathbf{f}_1, \dots, \mathbf{f}_q)$ onto a set of basis:

$$\mathbf{f}_i(t) \approx \mathbf{g}_i(t) = \sum_{j=1}^p \phi_{i,j}(t) \alpha_{i,j}, \quad \text{for } i = 1, \dots, q. \quad (23)$$

· Depending on the basis $\phi_{i,\cdot}$, then $\int_{\mathcal{T}} (\mathbf{g}_i(t) - \mathbf{g}'_i(t))^2 dt$ can be computed:

$$\int_{\mathcal{T}} (\mathbf{g}_i(t) - \mathbf{g}'_i(t))^2 dt = \int_{\mathcal{T}} \left[\sum_{j=1}^p (\alpha_{i,j} - \alpha'_{i,j}) \phi_{i,j}(t) \right]^2 dt = \boldsymbol{\beta}_i^\top \boldsymbol{\Psi}_i \boldsymbol{\beta}_i, \quad (24)$$

with $\boldsymbol{\Psi}_i = \int_{\mathcal{T}} \boldsymbol{\Psi}_i(t) \boldsymbol{\Psi}_i^\top(t) dt$ and $\boldsymbol{\beta}_i = (\alpha_i - \alpha'_i)$.

· If $\boldsymbol{\Psi}_i$ form an orthogonal family of basis (e.g. Fourier, Wavelets), then:

$$\boldsymbol{\Psi}_i = \text{diag} \left(\int_{\mathcal{T}} \boldsymbol{\Psi}_i(t) \boldsymbol{\Psi}_i^\top(t) dt \right). \quad (25)$$

· For orthonormal families (e.g. PCA), then $\boldsymbol{\Psi}_i = \mathbf{I}$ (identity matrix).

Note that $\boldsymbol{\Psi}_i$ can be computed only once and stored.

Link with multi-output Gaussian processes

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

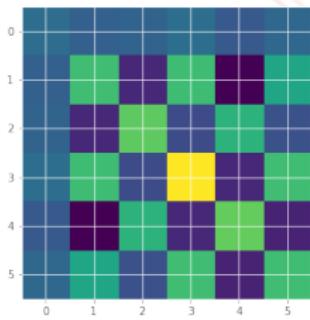
$$Z_i(\mathbf{x}) := Y(\mathcal{F}_i, \mathbf{x}), \quad \text{for } i = 1, \dots, R.$$

- In that case, \mathbf{k} can be rewritten as:

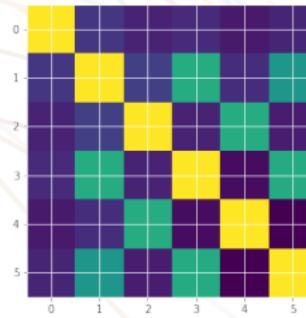
$$\mathbf{k}_{i,j}(\mathbf{x}, \mathbf{x}') = b_{i,j} \ k_f(\mathbf{x}, \mathbf{x}'), \quad (26)$$

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

- \mathbf{k} follows the structure of the *linear models of coregionalisation* (LMC):



\mathbf{B}

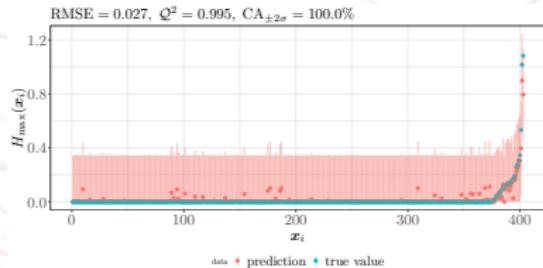
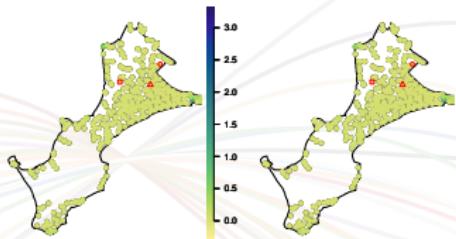


$k_f(\mathcal{F}, \mathcal{F}')$

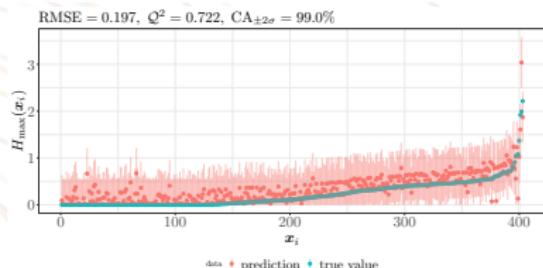
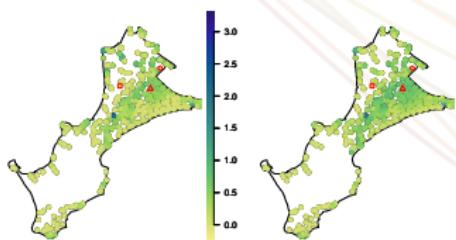
Coregionalisation matrix

Numerical illustrations

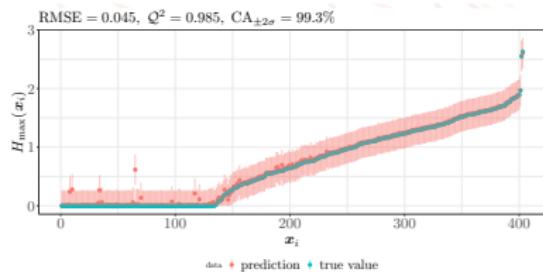
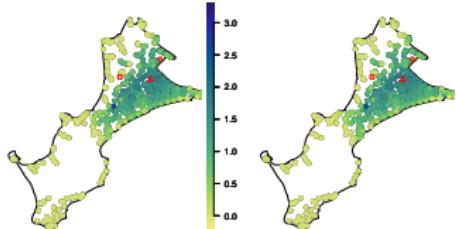
Flooding map 1



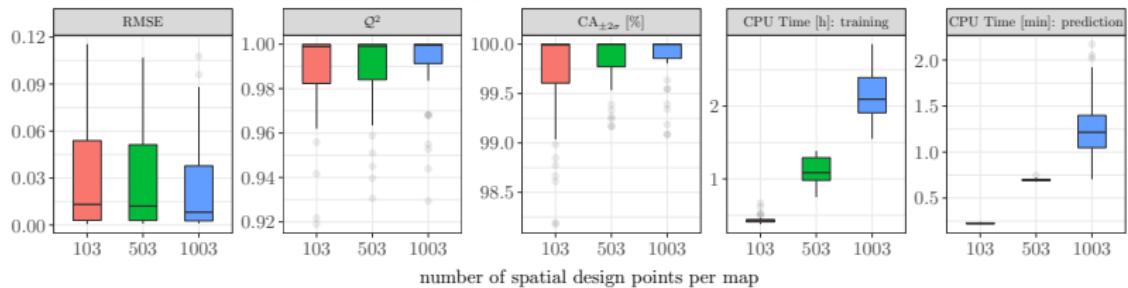
Flooding map 43



Flooding map 136



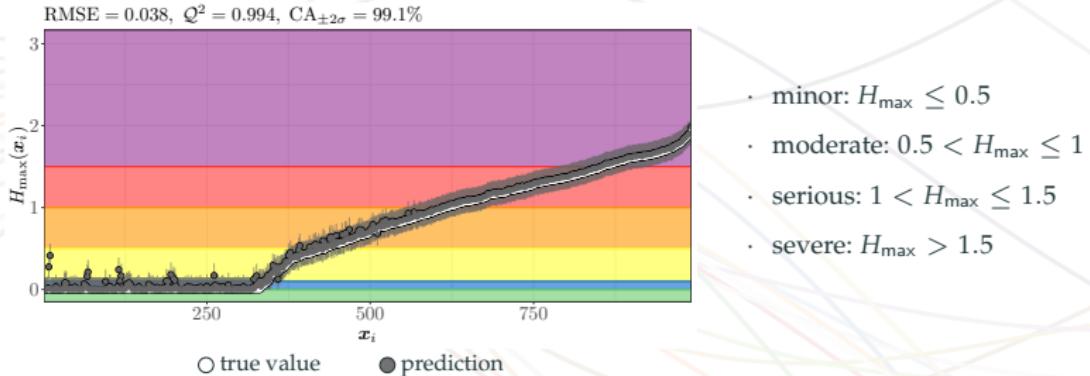
Numerical illustrations



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

Numerical illustrations

Assessment using the flood categories suggested by the French Risk Prevention Plan



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

Proportions computed on a dataset based on 131 flood events

Multi-fidelity GPs

Multi-fidelity transfer learning

- In transfer learning, data come from different types of sources, e.g. from two processes Y and Z
- Inspired by multi-fidelity models, we can consider the auto-regressive system given by

$$\begin{aligned} Y_1(x) &= \mathcal{G}_Y(Y_0(x)) + \mathcal{H}_{Y|Z}(Z_1(x)) + \delta_Y(x), \\ Z_1(x) &= \mathcal{G}_Z(Z_0(x)) + \mathcal{H}_{Z|Y}(Y_1(x)) + \delta_Z(x), \end{aligned}$$

with Y_0 and Y_1 (respectively Z_0 and Z_1) are the low and high fidelity levels of the process Y (resp. Z)

- Consider that \mathcal{G}, \mathcal{H} are linear transformations
- Assuming that $Y_0, Z_0, \delta_Y, \delta_X$ are GPs, then we can show that Y_1 and Z_1 are also GPs