Gaussian Process Modelling under Inequality Constraints

Andrés F. López-Lopera

PhD Defense on Applied Mathematics
Mines Saint-Étienne, France

Supervisor:	Olivier Roustant	Mines Saint-Étienne (France)
Co-supervisors:	François Bachoc	Univ. Paul Sabatier (France)
	Nicolas Durrande	PROWLER.io (UK)
Jury:	Sonja Kuhnt	FH Dortmund (Germany)
	Anthony Nouy	École Centrale de Nantes (France)
	Clémentine Prieur	Univ. Grenoble Alpes (France)
	Maurizio Filippone	EURECOM (France)

September 19, 2019

The chair of applied mathematics OQUAIDO

Academic partners:

Industrial and research partners:

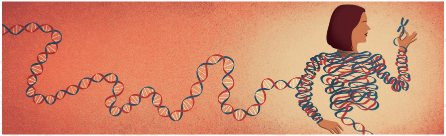
Table of Contents

MINES Saint-Étienne

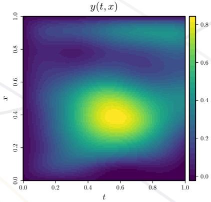
INSTITUT DE MATHÉMATIQUES DE TOULOUSE

Some real-world applications satisfying inequality constraints

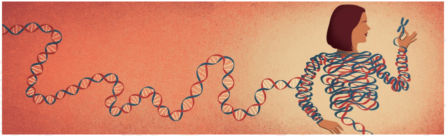
Regulation of gene expressions – positivity constraints (Lawrence et al., 2007).



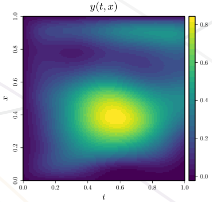
https://elifesciences.org/



Regulation of gene expressions – positivity constraints (Lawrence et al., 2007).



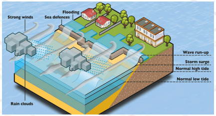
https://elifesciences.org/



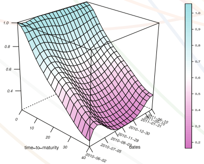
Other applications

- Coastal flooding – positivity & monotonicity (Rohmer and Idier, 2012).
- Econometrics – positivity or monotonicity (Cousin et al., 2016).
- Nuclear physics – monotonicity & convexity (Zhou et al., 2019).

⋮



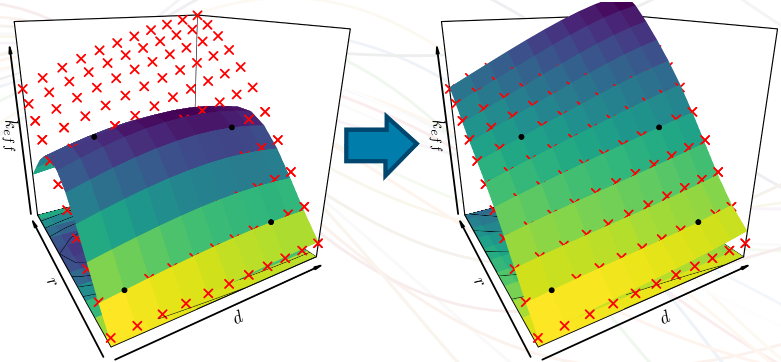
https://nerc.ukri.org/



Cousin et al. (2016)

2D models for interpolating the IRSN's dataset (López-Lopera et al., 2018).

k_{eff}: effective neutron multiplication factor

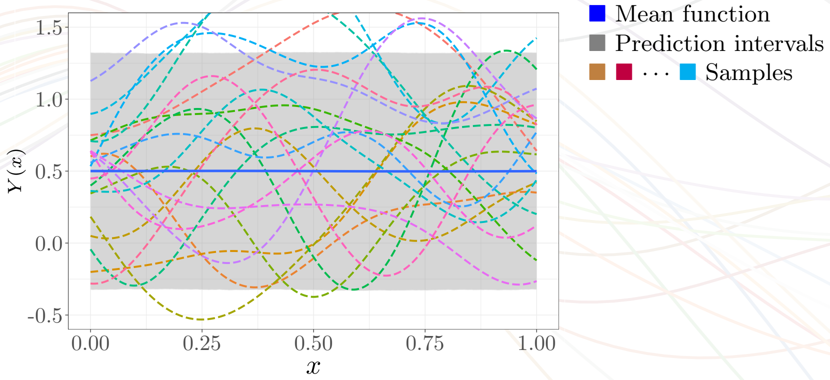


(a) unconstrained model (b) model with positivity & monotonicity

• interpolation points ($n = 4$) × test data

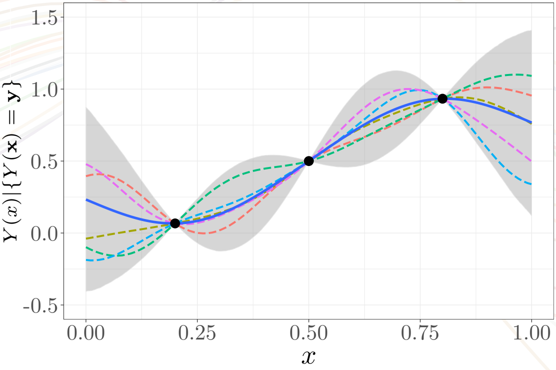
Gaussian process (GP) regression models: motivation

GPs form a flexible **prior over functions** (Rasmussen and Williams, 2005):



$$Y \sim \mathcal{GP}\left(\underbrace{\mu(x)}_{\text{mean}} = 0.5,\ \underbrace{k_\theta(x, x')}_{\text{covariance}} = \sigma^2 \exp\left\{-\frac{(x-x')^2}{2\ell^2}\right\}\right)$$

GPs form a flexible **prior over functions** (Rasmussen and Williams, 2005):



Mean function

Prediction intervals

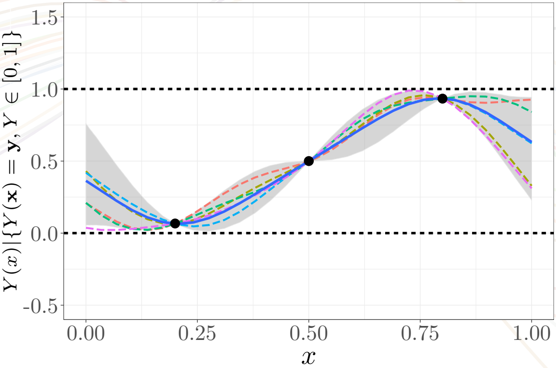
$\cdots$ Samples

- Interpolation conditions

$$\mathcal{D} = (x_i, y_i)_{i=1}^{n} = (\boldsymbol{x}, \mathbf{y})$$

$$Y|\{Y(\mathbf{x}) = \mathbf{y}\} \sim \mathcal{GP}(\underbrace{m(x)}_{\text{cond. mean}}, \underbrace{c_\theta(x, x')}_{\text{cond. covariance}})$$

GPs form a flexible **prior over functions** (Rasmussen and Williams, 2005):



- Mean function
- Prediction intervals
- ... Samples

- Interpolation conditions

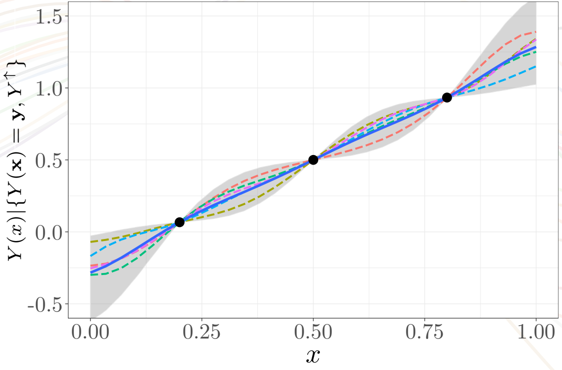
$$\mathcal{D} = (x_i, y_i)_{i=1}^{n} = (\boldsymbol{x}, \mathbf{y})$$

- - Boundedness condition

$$Y \in [0,1]$$

$$Y|\{Y(\mathbf{x}) = \mathbf{y}, Y \in [0,1]\} \sim \textbf{?}$$

GPs form a flexible **prior over functions** (Rasmussen and Williams, 2005):



- Mean function
- Prediction intervals
- $\cdots$ Samples
- $\bullet$ Interpolation conditions

$$\mathcal{D} = (x_i, y_i)_{i=1}^{n} = (\boldsymbol{x}, \mathbf{y})$$

- - Boundedness condition

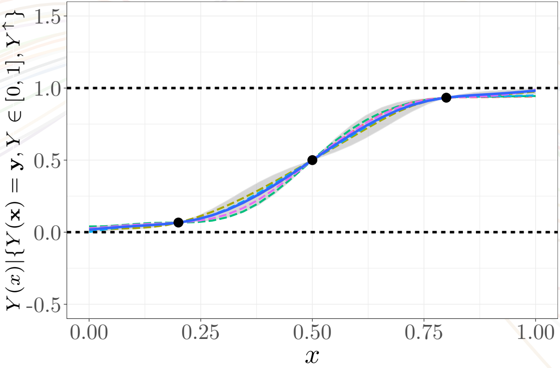
$$Y \in [0, 1]$$

$\uparrow$ Monotonicity condition

$$Y(x) \leq Y(x'), \quad \forall\, x \leq x'.$$

$$Y|\{Y(\mathbf{x}) = \mathbf{y}, Y^{\uparrow}\} \sim \ ?$$

GPs form a flexible **prior over functions** (Rasmussen and Williams, 2005):



Mean function

Prediction intervals

$\cdots$ Samples

• Interpolation conditions

$$\mathcal{D} = (x_i, y_i)_{i=1}^{n} = (\boldsymbol{x}, \mathbf{y})$$

- - Boundedness condition

$$Y \in [0, 1]$$

↑ Monotonicity condition

$$Y(x) \leq Y(x'), \quad \forall\, x \leq x'.$$

$$Y | \{Y(\mathbf{x}) = \mathbf{y}, Y \in [0,1], Y^{\uparrow}\} \sim \textbf{?}$$

The contributions in this thesis are threefold:

1. The improvement of the **applicability of GPs under constraints**.
   - Previous implementations: expensive and applicable up to 2D.
2. The **scalability** of constrained GPs **to high dimensions**:
   - i.e. involving hundreds of input variables.
3. **Parameter estimation under inequality constraints**.
   - Previously estimated by costly-to-evaluate cross-validation methods.

Table of Contents

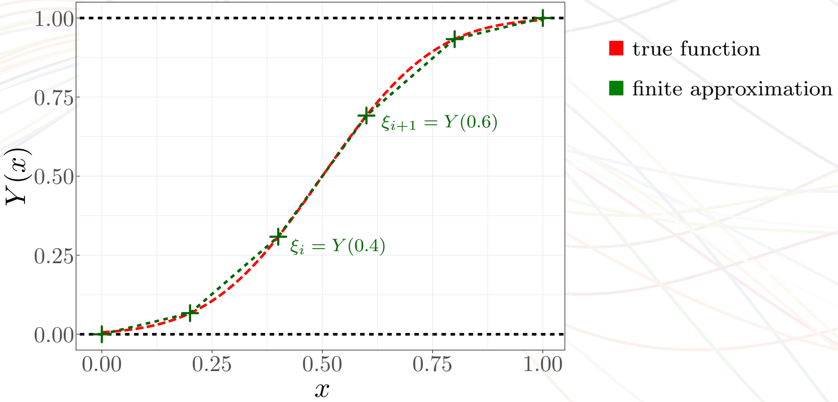
Table of Contents

MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

1D finite-dimensional Gaussian approximation

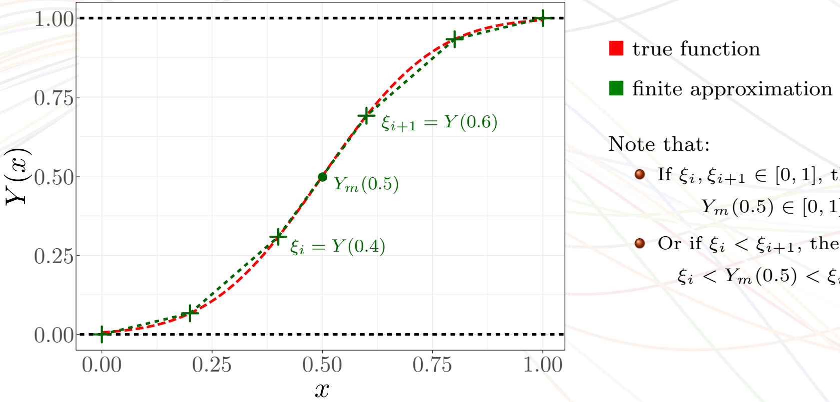
Finite-dimensional representation Y_m: also bounded & monotonic.



Imposing constraints on the knots is enough (Maatouk and Bay, 2017).

1D finite-dimensional Gaussian approximation

Finite-dimensional representation Y_m: also bounded & monotonic.



Note that:

- If $\xi_i, \xi_{i+1} \in [0, 1]$, then $Y_m(0.5) \in [0, 1]$.
- Or if $\xi_i < \xi_{i+1}$, then $\xi_i < Y_m(0.5) < \xi_{i+1}$.

Imposing constraints on the knots is enough (Maatouk and Bay, 2017).

Using the constrained GP framework in (Maatouk and Bay, 2017), allows...

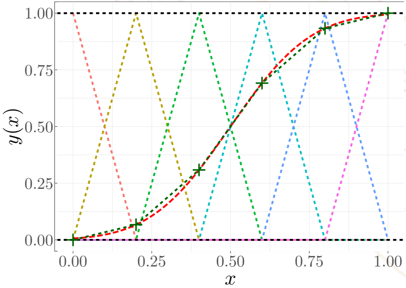
1. To impose either **boundedness**, **monotonicity** or **convexity**.
   - Other GP models deal with only one of those conditions (e.g. log GPs, Vanhatalo and Vehtari, 2007).
2. To ensure the **inequalities everywhere in the input space**.
   - Other frameworks satisfy constraints only over a finite number of evaluations (e.g. Riihimäki and Vehtari, 2010).
3. To approximate the constrained process by **sampling methods**.
   - Other approaches can lead to expensive or intractable computations.

- Let the (constrained) finite-dimensional GP Y_m be defined as

$$Y_m(x) = \sum_{j=1}^{m} \xi_j \phi_j(x), \text{ s.t. } \begin{cases} Y_m(x_i) = y_i & \text{(interpolation conditions)}, \\ \boldsymbol{l} \leq \boldsymbol{\Lambda\xi} \leq \boldsymbol{u} & \text{(linear inequality conditions)}, \end{cases} \tag{1}$$

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

- $\boldsymbol{\xi} = [\xi_1, \dots, \xi_m]^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_\theta)$ with covariance matrix $\boldsymbol{\Gamma}_\theta$,
- $(\boldsymbol{\Lambda}, \boldsymbol{l}, \boldsymbol{u})$ defines the inequality conditions, and
- $\phi_j : [0, 1] \mapsto \mathbb{R}$ are hat basis functions:



- Since the **Gaussianity** is preserved for *linear operations*:

$$\boldsymbol{\Lambda}\boldsymbol{\xi} | \{\boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{y}\} \sim \mathcal{N}\left(\boldsymbol{\Lambda}\boldsymbol{\mu}, \boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^\top\right), \quad \text{(conditional distribution)} \tag{2}$$

with conditional parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given by

$$\mathbf{K} = \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^\top, \quad \boldsymbol{\mu} = \boldsymbol{\Gamma}\boldsymbol{\Phi}^\top\mathbf{K}^{-1}\mathbf{y}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{\Phi}^\top\mathbf{K}^{-1}\boldsymbol{\Phi}\boldsymbol{\Gamma}.$$

- Since the **Gaussianity** is preserved for *linear operations*:

$$\boldsymbol{\Lambda}\boldsymbol{\xi}|\{\boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{y}\} \sim \mathcal{N}\left(\boldsymbol{\Lambda}\boldsymbol{\mu}, \boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^\top\right), \quad \text{(conditional distribution)} \qquad (2)$$

with conditional parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given by

$$\mathbf{K} = \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^\top, \quad \boldsymbol{\mu} = \boldsymbol{\Gamma}\boldsymbol{\Phi}^\top\mathbf{K}^{-1}\mathbf{y}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{\Phi}^\top\mathbf{K}^{-1}\boldsymbol{\Phi}\boldsymbol{\Gamma}.$$

- Then, *quantifying uncertainty* on $Y_m(x) = \sum_{j=1}^m \xi_j \phi_j(x)$ relies on simulating the **truncated Gaussian vector** $\boldsymbol{\xi}$:

$$\boldsymbol{\Lambda}\boldsymbol{\xi}|\{\boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{y}, \boldsymbol{l} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}\} \sim \mathcal{TN}(\boldsymbol{\Lambda}\boldsymbol{\mu}, \boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^\top, \boldsymbol{l}, \boldsymbol{u}). \qquad (3)$$

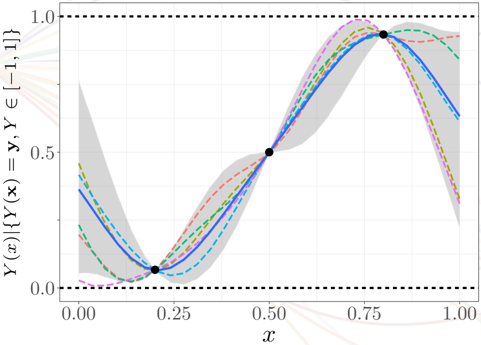
(3) is computed via *Monte Carlo* (MC) or *Markov Chain MC* (MCMC):

- **e.g. *Hamiltonian MC* (HMC) (Pakman and Paninski, 2014).**

MINES Saint-Étienne

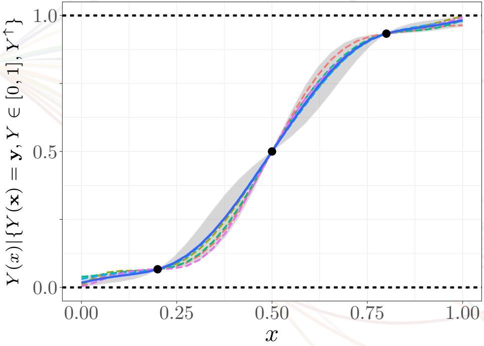
INSTITUT DE MATHÉMATIQUES DE TOULOUSE

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



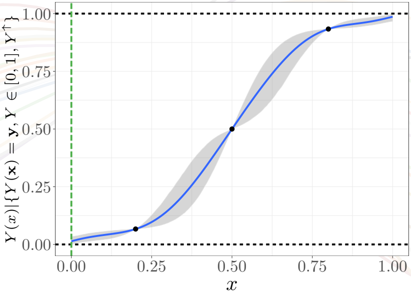
$$\underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{l}} \leq \underbrace{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\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

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

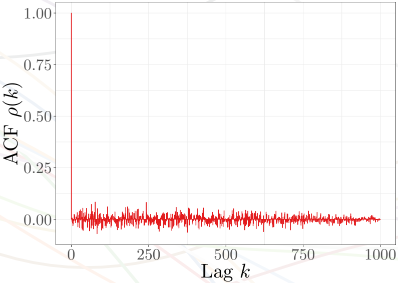


$$
\underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{l}} \leq \underbrace{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 &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

boundedness & **monotonicity** constraints



$m = 100,\ N = 10^3$



MC/MCMC Sampler	CPU Time [s]
(Maatouk and Bay, 2016)	-
(Botev, 2017)	73.7
Gibbs – (Taylor and Benjamini, 2016)	29.0
HMC – (Pakman and Paninski, 2014)	**12.6**

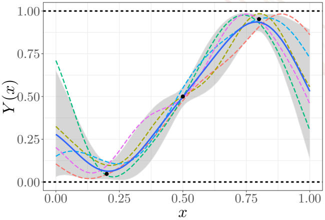
Considering noisy observations

$$Y_m(x) = \sum_{j=1}^{m} \xi_j \phi_j(x), \text{ s.t. } \begin{cases} Y_m(x_i) + \varepsilon_i = y_i & \text{(interpolation conditions)}, \\ \boldsymbol{l} \leq \boldsymbol{\Lambda\xi} \leq \boldsymbol{u} & \text{(linear inequality conditions)}, \end{cases} \quad (4)$$

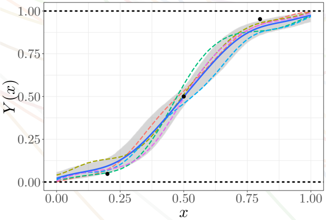
where $\varepsilon_i \sim \mathcal{N}\left(0, \tau^2\right)$ with noise variance τ^2.

Modification: $\mathbf{K} = \mathbf{\Phi\Gamma\Phi}^\top \quad \rightarrow \quad \mathbf{K} = \mathbf{\Phi\Gamma\Phi}^\top + \tau^2\mathbf{I}$.

boundedness

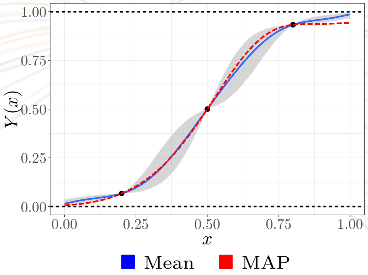


boundedness & **monotonicity**



Adding $\boldsymbol{\varepsilon}$ leads to more flexible GP models and less restrictive sample spaces for MC/MCMC algorithms (López-Lopera et al., 2019).

Maximum a posteriori (MAP) estimate



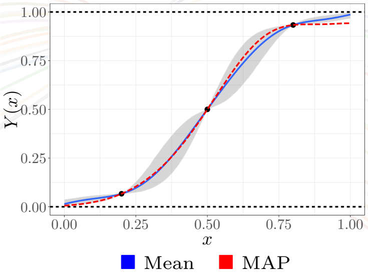
- Let $\boldsymbol{\mu}^*$ be the mode that maximises the pdf of $\boldsymbol{\xi}|\{\boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{y}, \boldsymbol{l} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}\}$:

$$\boldsymbol{\mu}^* = \underset{\boldsymbol{\xi} \text{ s.t. } \boldsymbol{l} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}}{\arg\max} \{-[\boldsymbol{\xi}-\boldsymbol{\mu}]^\top \boldsymbol{\Sigma}^{-1}[\boldsymbol{\xi}-\boldsymbol{\mu}]\}. \tag{5}$$

- The *MAP estimate* of Y_m is given by

$$Y_m^{\mathrm{MAP}}(x) = \sum_{j=1}^{m} \mu_j^* \phi_j(x), \tag{6}$$

Maximum a posteriori (MAP) estimate



- Let $\boldsymbol{\mu}^*$ be the mode that maximises the pdf of $\boldsymbol{\xi}|\{\boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{y}, \boldsymbol{l} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}\}$:

$$\boldsymbol{\mu}^* = \underset{\boldsymbol{\xi} \text{ s.t. } \boldsymbol{l} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}}{arg\,max} \{-[\boldsymbol{\xi} - \boldsymbol{\mu}]^\top \boldsymbol{\Sigma}^{-1}[\boldsymbol{\xi} - \boldsymbol{\mu}]\}. \tag{5}$$

- The *MAP estimate* of Y_m is given by

$$Y_m^{\text{MAP}}(x) = \sum_{j=1}^{m} \mu_j^* \phi_j(x), \tag{6}$$

Benefits:

- **Easy** and **fast calculations**.
- **Convergence to the spline solution** as $m \to \infty$ (Bay et al., 2016).
- **Starting point for MCMC** aiming for *fast convergence*.

Table of Contents

MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

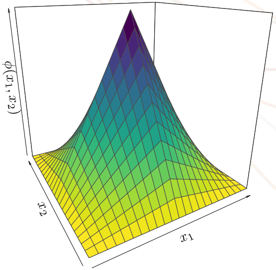
Finite-dimensional approximation in d dimensions

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

$$Y_m(\boldsymbol{x}) = \sum_{j_1,\ldots,j_d=1}^{m_1,\ldots,m_d} \left[\prod_{p=1,\ldots,d} \phi_{j_p}^{(p)}(x_p) \right] \xi_{j_1,\cdots,j_d}, \text{ s.t. } \begin{cases} Y_m(\boldsymbol{x}_i) + \varepsilon_i = y_i, \\ \boldsymbol{\xi} \in \mathcal{C}, \end{cases} \tag{7}$$

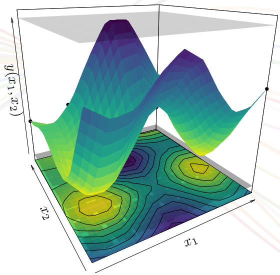
where $\boldsymbol{x}_i \in [0,1]^d$, $y_i \in \mathbb{R}$, $\varepsilon_i \sim \mathcal{N}(0, \tau^2)$, for $i = 1, \ldots, n$; and

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

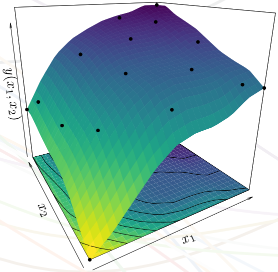


2D numerical illustration

boundedness in 2D



monotonicity in 2D.



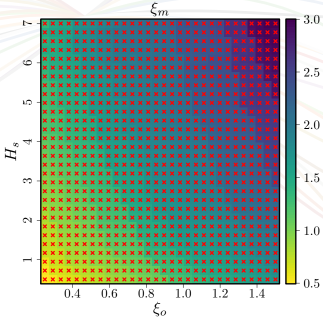
● interpolation points ■ predictive mean ■ bounds

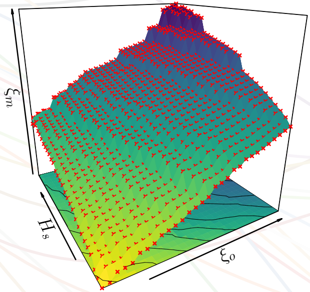
MC/MCMC Method	2D boundedness example CPU Time $[s]$	2D monotonicity example CPU Time $[s]$
(Maatouk and Bay, 2016)	-	-
(Botev, 2017)	0.9	1488.3
(Taylor and Benjamini, 2016)	9.7	-
(Pakman and Paninski, 2014)	**0.6**	**8.6**

$$m_1 = m_2 = 10, N = 10^3$$

2D coastal flooding application

ξ_m: maximum water level, ξ_0: offshore water level, H_s: wave height.

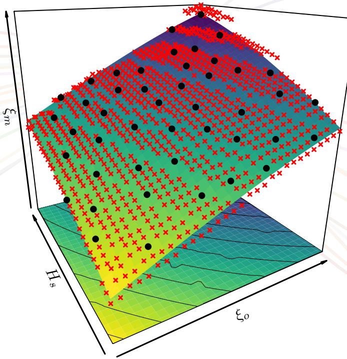




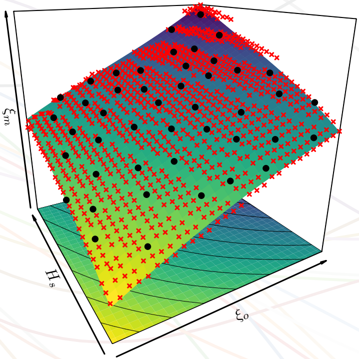
× coastal flooding data

2D coastal flooding application in (Rohmer and Idier, 2012).

Prediction using 5% of data ($n = 45$)



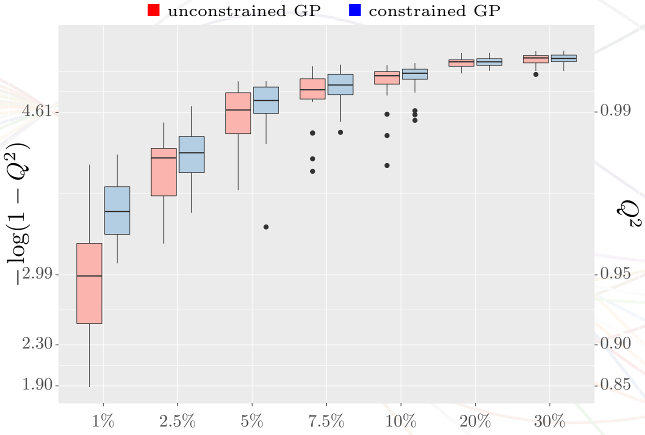
(a) Unconstr. GP: $Q^2 = 0.98$



(b) Constr. GP: $Q^2 = 0.99$

• training data × test data

Accurate predictions are obtained when $Q^2 \to 1$.

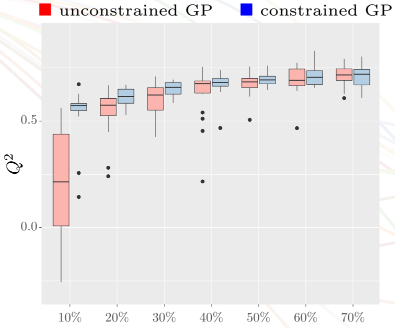


Q^2 performance using 20 replicates

Accurate predictions are obtained when $Q^2 \to 1$.

5D coastal flooding application

- Coastal flooding induced by overflow on the Atlantic coast, focusing on the **inland flooded surface** (Azzimonti et al., 2019).



Q^2 performance using 20 replicates

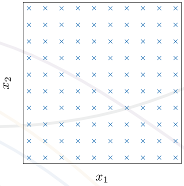
For small values of n, the constrained GP is more accurate.

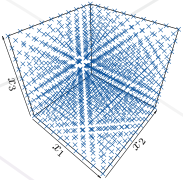
Main contributions

1. The improvement of the **applicability of GPs under constraints**: ✓
   - applicable up to 5D without further assumptions.
2. The scalability of constrained GPs to high dimensions.
3. Parameter estimation under inequality constraints.

Curse of dimensionality

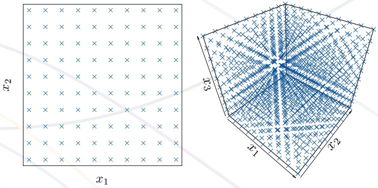
- The cost of Y_m increases as d (or $m = m_1 \times \cdots \times m_d$) increases.





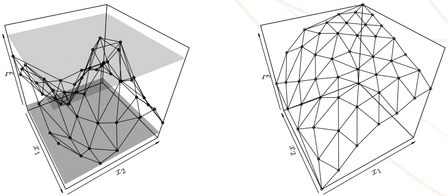
Curse of dimensionality

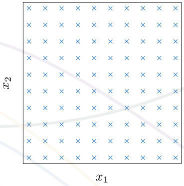
- The cost of Y_m increases as d (or $m = m_1 \times \cdots \times m_d$) increases.

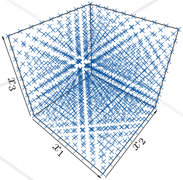


This downside can be partially mitigated in different ways...

- Using a **"smarter" construction of rectangular grids** of knots.
- Considering **further assumptions for complexity simplification**:
  - e.g. inactive dimensions, additive conditions.
- Using other types of designs of knots: e.g. **Delaunay triangulations**.





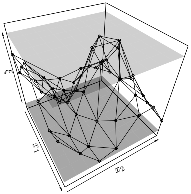


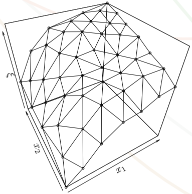
Curse of dimensionality

- The cost of Y_m increases as d (or $m = m_1 \times \cdots \times m_d$) increases.

This downside can be partially mitigated in different ways...

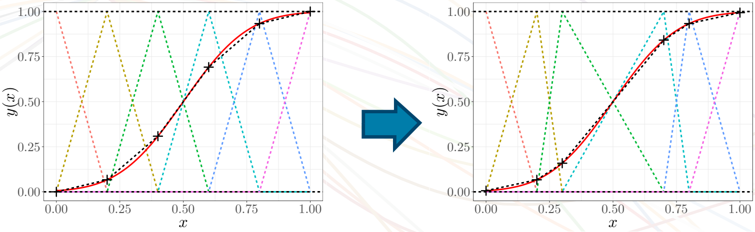
- Using a **"smarter" construction of rectangular grids** of knots. ✓
- Considering **further assumptions for complexity simplification**:
  - e.g. inactive dimensions, additive conditions. ✓
- Using other types of designs of knots: e.g. **Delaunay triangulations**.





Free-knot finite-dimensional approximation of GPs

- Consider the **asymmetric hat basis functions**:



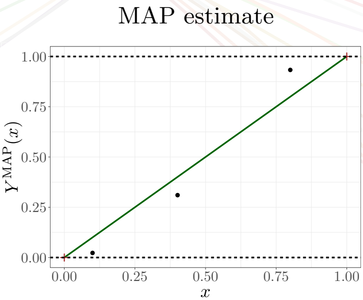
Standard hat basis functions. Asymmetric hat basis functions.

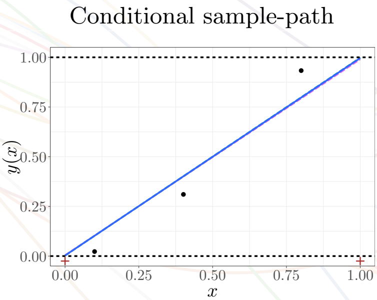
This allows sequentially refining (rectangular) grids of knots:

- **In highly variable regions.**
- **Along more active dimensions.**

- The "optimal" knot insertion is achieved by maximising the **integrated MAP squared error (iMAP-SE)** criterion:

$$t_*^{\text{opt}} = \underset{t_* \in [0,1]}{argmax} \int_0^1 [Y^{\text{MAP}}_{m_{i+1}, t_*}(x) - Y^{\text{MAP}}_{m_i}(x)]^2 dx. \quad (8)$$



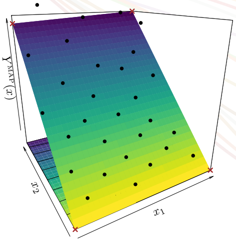


• training points + knots ■ MAP estimate
■ predictive mean ■ 90% confidence intervals

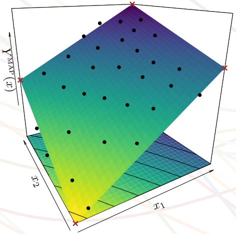
- In d dimensions, the **iMAP-SE** criterion is given by

$$t_*^{j_*,\mathrm{opt}} = \underset{j_* \in \{1,\cdots,d\},\ t_* \in [0,1]}{argmax} \int_{\boldsymbol{x} \in [0,1]^d} [Y^{\mathrm{MAP}}_{m_{i+1},j_*,t_*}(\boldsymbol{x}) - Y^{\mathrm{MAP}}_{m_i}(\boldsymbol{x})]^2 d\boldsymbol{x}. \quad (9)$$

Toy 1



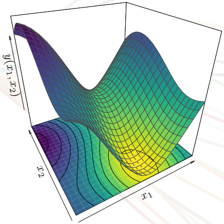
Toy 2

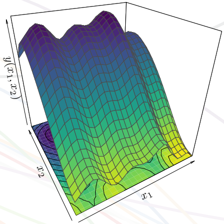


• training points + knots ■ MAP estimate

This construction has been tested up to 5D applications.

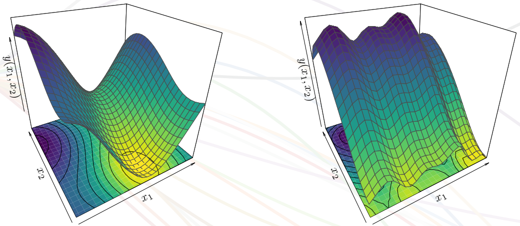
Additive GPs





2D examples of additive GPs.

Additive GPs



2D examples of additive GPs.

- Let Y be the additive process on $\mathbb{R}^d$ given by

$$Y(\boldsymbol{x}) = \sum_{p=1}^{d} Y_p(x_p), \quad \text{with} \quad Y_p \sim \mathcal{GP}(0, k_p). \tag{10}$$

- Assume that $Y_1, \ldots, Y_d$ are independent. Then, $Y \sim \mathcal{GP}(0, k)$ with:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{p=1}^{d} k_p(x_p, x_p{}').$$

Additive GPs under inequality constraints

- Assume that Y exhibits certain constraints along Y_p, then Y_m is

$$Y_p \to Y_{p,m_p}$$

$$Y_m(\boldsymbol{x}) = \sum_{p=1}^{d} Y_{p,m_p}(x_p), \quad \text{s.t.} \quad \begin{cases} Y_m(\boldsymbol{x}_i) + \varepsilon_i = y_i, \\ \boldsymbol{\xi}_\kappa \in \mathcal{C}_\kappa, \end{cases} \tag{11}$$

with $\boldsymbol{x}_i \in [0,1]^d$, $y_i \in \mathbb{R}$ and $\varepsilon_i \sim \mathcal{N}\left(0, \tau^2\right)$ for $i = 1, \ldots, n$.

Additive GPs under inequality constraints

- Assume that Y exhibits certain constraints along Y_p, then Y_m is

$$Y_p \to Y_{p,m_p}$$

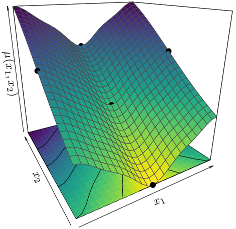
$$Y_m(\boldsymbol{x}) = \sum_{p=1}^{d} Y_{p,m_p}(x_p), \quad \text{s.t.} \quad \begin{cases} Y_m(\boldsymbol{x}_i) + \varepsilon_i = y_i, \\ \boldsymbol{\xi}_\kappa \in \mathcal{C}_\kappa, \end{cases} \tag{11}$$

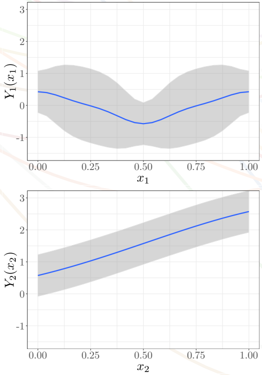
with $\boldsymbol{x}_i \in [0,1]^d$, $y_i \in \mathbb{R}$ and $\varepsilon_i \sim \mathcal{N}(0, \tau^2)$ for $i = 1, \ldots, n$.

2D toy example:

$$Y(x_1, x_2) = \underbrace{4(x_1 - 0.5)^2}_{Y_1(x_1)} + \underbrace{2x_2}_{Y_2(x_2)}$$

Predictive mean without constraints





Additive GPs under inequality constraints

- Assume that Y exhibits certain constraints along Y_p, then Y_m is

$$Y_p \to Y_{p,m_p}$$

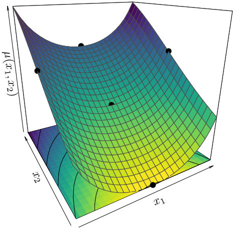
$$Y_m(\boldsymbol{x}) = \sum_{p=1}^{d} Y_{p,m_p}(x_p), \quad \text{s.t.} \quad \begin{cases} Y_m(\boldsymbol{x}_i) + \varepsilon_i = y_i, \\ \boldsymbol{\xi}_\kappa \in \mathcal{C}_\kappa, \end{cases} \tag{11}$$

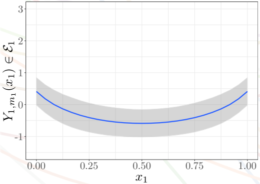
with $\boldsymbol{x}_i \in [0,1]^d$, $y_i \in \mathbb{R}$ and $\varepsilon_i \sim \mathcal{N}(0, \tau^2)$ for $i = 1, \ldots, n$.

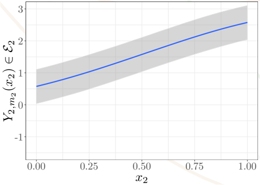
2D toy example:

$$Y(x_1, x_2) = \underbrace{4(x_1 - 0.5)^2}_{Y_1(x_1)} + \underbrace{2x_2}_{Y_2(x_2)}$$

Predictive mean with constraints



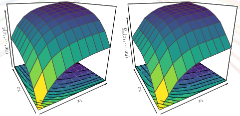




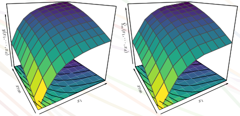
Numerical illustration in hundreds of dimensions

- Consider the (monotonic) target function:

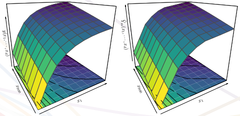
$$y(\boldsymbol{x}) = \sum_{p=1}^{d} \arctan\left(5\left[1 - \frac{p}{d}\right] x_p\right), \quad \text{with} \quad \boldsymbol{x} \in [0,1]^d.$$



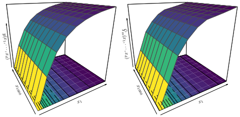
$y_1(x_1) + y_2(x_2)$



$y_1(x_1) + y_{700}(x_{700})$



$y_1(x_1) + y_{900}(x_{900})$



$y_1(x_1) + y_{1000}(x_{1000})$

	CPU Time [s]					
d	10	50	100	200	500	**1000**
Prediction	0.1	0.3	1.4	10.5	165.9	1364.5
Sampling	0.2	1.1	2.9	2.7	5.3	10.8

CPU times for sampling 10^3 trajectories via HMC.

Examples of an additive GP under monotonicity constraints in 10^3 dimensions. Each panel shows: the true (left) and predictive (right) mean profiles.

Main contributions

1. The improvement of the **applicability of GPs under constraints**: ✓
   - applicable up to 5D without further assumptions.
2. The **scalability** of constrained GPs **to high dimensions**: ✓
   - i.e. involving hundreds of input variables under additive assumptions.
3. Parameter estimation under inequality constraints.

Table of Contents

Maximum likelihood estimator (MLE)

- The covariance parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ can be estimated via ML (Stein, 1999):

$$\widehat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{argmax}\ \mathcal{L}_n(\boldsymbol{\theta}),$$

where $\mathcal{L}_n(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{Y}_n)$ is the unconstrained log-likelihood:

$$\mathcal{L}_n(\boldsymbol{\theta}) = -\frac{1}{2}\log(\det(\mathbf{R}_{\boldsymbol{\theta}})) - \frac{1}{2}\mathbf{Y}_n^\top \mathbf{R}_{\boldsymbol{\theta}}^{-1}\mathbf{Y}_n - \frac{n}{2}\log 2\pi, \tag{12}$$

with $\mathbf{R}_{\boldsymbol{\theta}} = (k_{\boldsymbol{\theta}}(x_i, x_j))_{1 \leq i,j \leq n}$.

✗ **Note that the MLE $\widehat{\boldsymbol{\theta}}_n$ does not take into account the inequality conditions.**

Constrained maximum likelihood estimator (cMLE)

- Consider a convex set with linear inequality constraints $\mathcal{E}$.
- Let $\mathcal{L}_{n,c}(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{Y}_n | Y \in \mathcal{E})$ be the constrained log-likelihood:

$$\begin{aligned}\mathcal{L}_{n,c}(\boldsymbol{\theta}) &= \log \frac{p_{\boldsymbol{\theta}}(\mathbf{Y}_n) p_{\boldsymbol{\theta}}(Y \in \mathcal{E} | \mathbf{Y}_n)}{p_{\boldsymbol{\theta}}(Y \in \mathcal{E})} \\ &= \mathcal{L}_n(\boldsymbol{\theta}) + \log p_{\boldsymbol{\theta}}(Y \in \mathcal{E} | \mathbf{Y}_n) - \log p_{\boldsymbol{\theta}}(Y \in \mathcal{E}).\end{aligned} \tag{13}$$

- Then, the cMLE is given by

$$\widehat{\boldsymbol{\theta}}_{n,c} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{argmax}\ \mathcal{L}_{n,c}(\boldsymbol{\theta}).$$

✓ **The cMLE $\widehat{\boldsymbol{\theta}}_{n,c}$ accounts for the inequality constraints.**

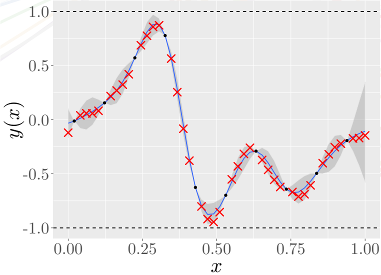
✗ **Both $p_{\boldsymbol{\theta}}(Y \in \mathcal{E} | \mathbf{Y}_n)$ and $p_{\boldsymbol{\theta}}(Y \in \mathcal{E})$ are approximated via MC.**

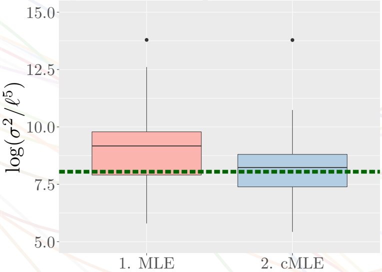
MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

Numerical illustration

Assessment of the MLE and cMLE for $\mathbf{10^2}$ **realisations** drawn from a constrained GP $Y \in [-1, 1]$ with Matérn 5/2 kernel and true parameters $\boldsymbol{\theta}_0 = (\sigma^2 = 1, \ell = 0.2)$.





(b) Consistently estimable ratio $\log(\sigma^2/\ell^5)$

Notation

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

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

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

Notation

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

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

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

Proposition (Consistency of the MLE and cMLE, López-Lopera et al. (2018))

Assume $\forall \varepsilon > 0$ and $\forall M < \infty$, ***(Consistency of the unconditional ML)***

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

Then, ***(Consistency of the conditional cML)***

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

Consequently, both the MLE and cMLE are consistent estimators:

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

Notation

- 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.

Notation

- 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, Bachoc et al. (2019))

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

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

Notation

- 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, Bachoc et al. (2019))

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

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

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

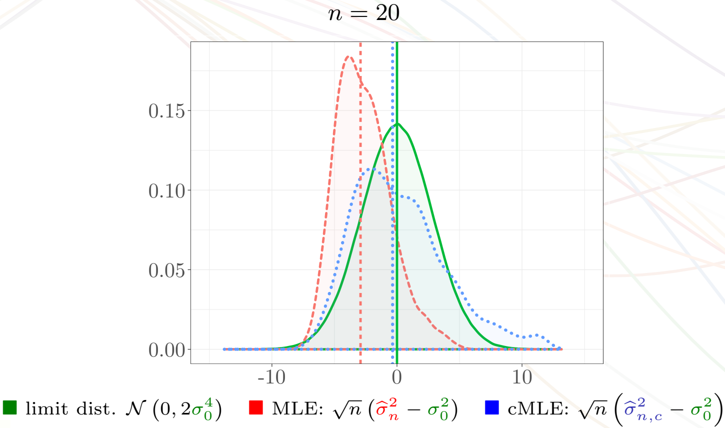
$$\sqrt{n}\left(\widehat{\sigma}_{n,c}^2 - \sigma_0^2\right) \xrightarrow[n\to+\infty]{\mathcal{L}|Y\in\mathcal{E}_\kappa} \mathcal{N}(0, 2\sigma_0^4).$$

The results can be extended for Matérn models (Bachoc et al., 2019).

Numerical illustration

Assessment of the MLE and cMLE normality for 10^3 **trajectories** drawn from a constrained GP $Y \in [0, 1]$ with Matérn 5/2 kernel and parameters $\boldsymbol{\theta}_0 = (\sigma_0^2 = 0.2, \ell = 0.2)$.

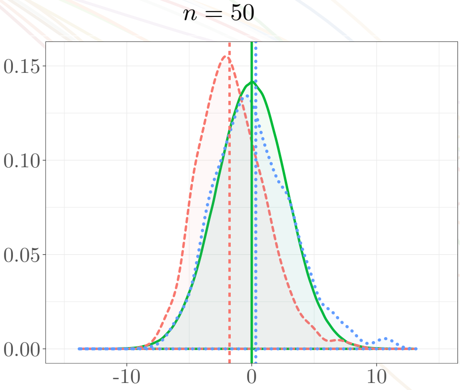
For small values of n, cMLE seems to be more accurate.



Numerical illustration

Assessment of the MLE and cMLE normality for $\mathbf{10^3}$ **trajectories** drawn from a constrained GP $Y \in [0,1]$ with Matérn 5/2 kernel and parameters $\boldsymbol{\theta}_0 = (\sigma_0^2 = 0.2, \ell = 0.2)$.

cMLE converges faster to the limit Gaussian distribution.

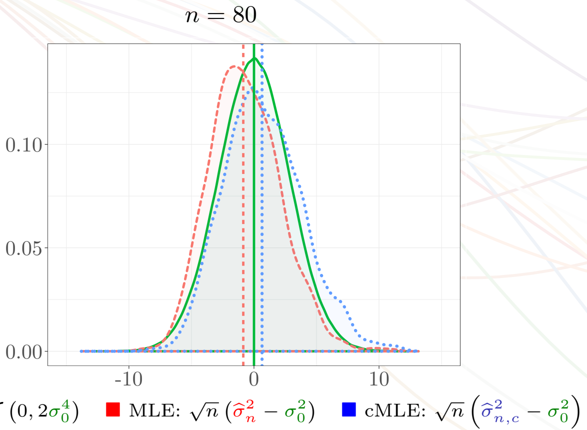


limit dist. $\mathcal{N}(0, 2\sigma_0^4)$ MLE: $\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right)$ cMLE: $\sqrt{n}\left(\widehat{\sigma}_{n,c}^2 - \sigma_0^2\right)$

Numerical illustration

Assessment of the MLE and cMLE normality for $\mathbf{10^3}$ **trajectories** drawn from a constrained GP $Y \in [0,1]$ with Matérn 5/2 kernel and parameters $\boldsymbol{\theta}_0 = (\sigma_0^2 = 0.2, \ell = 0.2)$.

For large values of n, MLE and cMLE provide similar performances.

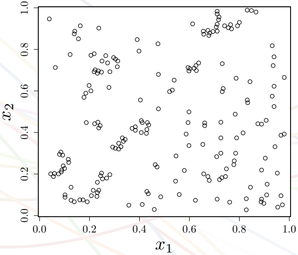


Main contributions

1. The improvement of the **applicability of GPs under constraints**: ✓
   - applicable up to 5D without further assumptions.
2. The **scalability** of constrained GPs **to high dimensions**: ✓
   - i.e. involving hundreds of input variables under additive assumptions.
3. **Parameter estimation under inequality constraints**: ✓
   - the use of **cMLE** is suggested for small amount of data.

Point Poisson process

- **Random countable subset** where *point patterns occur independently*.



Applications

- Geology (Møller and Waagepetersen, 2004).
- Astronomy (Baddeley et al., 2015).
- Lifetime of items or failure rates (Cha and Finkelstein, 2018).

- Unconditional likelihood under an **inhomogeneous Poisson process**:

$$f_{(N,\mathbf{X}_1,\cdots,\mathbf{X}_n)}(n,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n) = \exp\left(-\int_{\boldsymbol{x}\in\mathcal{D}} \lambda(\boldsymbol{x})d\boldsymbol{x}\right)\prod_{i=1}^{n}\frac{\lambda(\boldsymbol{x}_i)}{n!}, \qquad (14)$$

with

- $N \in \mathbb{N}$ a r.v. denoting the number of *point patterns*,
- $\mathbf{X}_1,\cdots,\mathbf{X}_n$ a set of n i.i.d. random vectors on $\mathcal{D}$, and
- $\lambda : \mathcal{D} \to [0,\infty]$ the (positive) intensity function.

- In real applications, λ is *unknown and must to be inferred.*

- Unconditional likelihood under an **inhomogeneous Poisson process**:

$$f_{(N,\mathbf{X}_1,\cdots,\mathbf{X}_n)}(n,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n) = \exp\left(-\int_{\boldsymbol{x}\in\mathcal{D}} \lambda(\boldsymbol{x})d\boldsymbol{x}\right)\prod_{i=1}^{n}\frac{\lambda(\boldsymbol{x}_i)}{n!}, \qquad (14)$$

with

- $N \in \mathbb{N}$ a r.v. denoting the number of *point patterns*,
- $\mathbf{X}_1,\cdots,\mathbf{X}_n$ a set of n i.i.d. random vectors on $\mathcal{D}$, and
- $\lambda : \mathcal{D} \to [0,\infty]$ the (positive) intensity function.

- In real applications, λ is *unknown and must to be inferred.*

Cox process (Cox, 1955)

- Cox processes considers λ as a positive stochastic process:

$$\Lambda = \{\Lambda(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}.$$

MINES Saint-Étienne

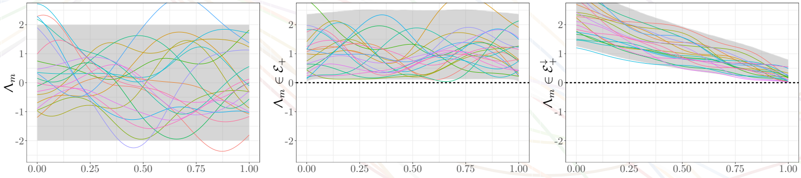
INSTITUT DE MATHÉMATIQUES DE TOULOUSE

GP-modulated Cox processes under inequality constraints

- Consider the finite-dimensional GP prior (López-Lopera et al., 2019):

$$\Lambda_m \in \mathcal{E},$$

with $\mathcal{E}$ a convex set composed by a set of linear inequalities.



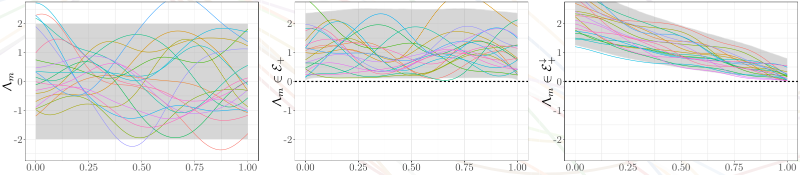
unconstrained GP prior | positive GP prior | monotone GP prior

GP-modulated Cox processes under inequality constraints

- Consider the finite-dimensional GP prior (López-Lopera et al., 2019):

$$\Lambda_m \in \mathcal{E},$$

with $\mathcal{E}$ a convex set composed by a set of linear inequalities.

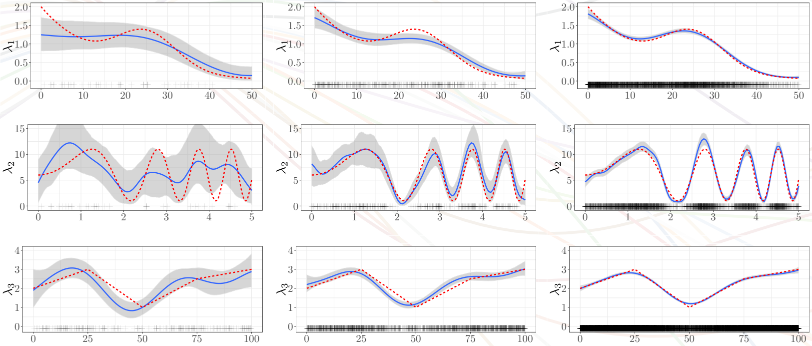


unconstrained GP prior | positive GP prior | monotone GP prior

Benefits:

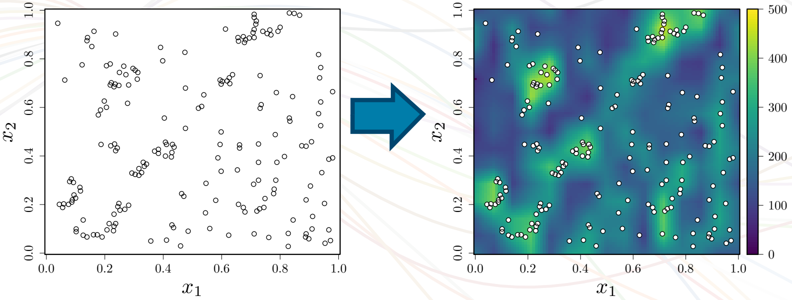
- **Tractable computations of the likelihood**.
- **Cox process inference via MCMC**:
  - *Metropolis-Hastings algorithm* with truncated Gaussian proposals.
- Applications to **other point processes**:
  - e.g. *Renewal processes* with monotone priors (Cha and Finkelstein, 2018).

Inference results for the examples in (Adams et al., 2009).



true intensity λ inferred intensity λ_m 90% confidence intervals

Location of redwood trees (dataset from Ripley, 1977).

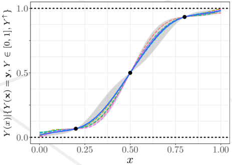


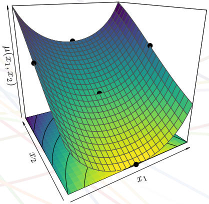
○ location of trees ■ inferred intensity λ

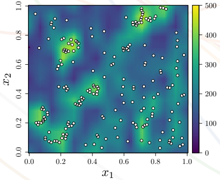
Competitive inference compared to approaches from the literature (e.g. log GPs, Møller et al., 2001).

In summary...

- We further investigated (Maatouk and Bay, 2017) to account for **linear inequality constraints**.
- We explored the use of the **HMC** for the *posterior approximation*.
- We proposed **variations of the GP model**:
  - *Noisy observations*,
  - *Free-knot (rectangular) designs* of knots,
  - *Additive (or block-additive)* functions.
- We studied a **constrained likelihood** analysing some *asymptotic properties*.
- We adapted the constrained GP framework to **modelling Poisson processes**.







Conclusions

- This thesis was dedicated to study **stochastic models based on GPs under inequality constraints**.
- Accounting for those constraints led to **more realistic uncertainty quantifications**, e.g.:
  - *Risk assessment in nuclear safety.*
  - *Coastal flooding modelling.*
  - *Spatial location of trees.*
  - ...
- Inference under constraints yielded **more accurate parameter estimation** for small (or moderate) amount of data.

Future work

- The **extension to high dimensions** is still challenging:
  - Alternative constructions can be explored (e.g. *Delaunay triangulations*).
- Existing methods for **approximating the orthant probabilities** of the cMLE are *time-consuming* and *unstable*.
- Coupling the proposed framework to other types of **GP-modulated processes** under inequality constraints: e.g. *Renewal processes*.

Publications in international journals

- L-L, A.F., Bachoc, F., Durrande, N., and Roustant, O. (2018). Finite-dimensional Gaussian approximation with linear inequality constraints. In ***SIAM/ASA Journal on Uncertainty Quantification***.
- Bachoc, F., Lagnoux, A., and L-L, A.F. (2019). Maximum likelihood estimation for Gaussian processes under inequality constraints. In ***Electronic Journal of Statistics***.

Proceedings in international conferences

- L-L, A.F., John, S., and Durrande, N. (2019). Gaussian process modulated Cox processes under linear inequality constraints. In *International Conference on Artificial Intelligence and Statistics (**AISTATS**)*.
- L-L, A.F., Bachoc, F., Durrande, N., Rohmer, J., Idier, D., and Roustant, O (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 (**MCQMC**)*.

R packages

- L-L, A.F. (2019). lineqGPR: Gaussian process regression models with inequality constraints. **CRAN R package**.


Lancement de la Chaire
en Mathématiques Appliquées
Oquaido : Optimisation et Quantification
d'Incertitudes pour les Données Onéreuses

Thanks!

Adams, R. P., Murray, I., and MacKay, D. J. (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *International Conference on Machine Learning*.

Azzimonti, D., Ginsbourger, D., Rohmer, J., and Idier, D. (2019). Profile extrema for visualizing and quantifying uncertainties on excursion regions. Application to coastal flooding. *Technometrics*, 0(ja):1–26.

Bachoc, F., Lagnoux, A., and López-Lopera, A. F. (2019). Maximum likelihood estimation for Gaussian processes under inequality constraints. *Electronic J. of Statistics*, 13(2):2921–2969.

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC Interdisciplinary Statistics. CRC Press.

Bay, X., Grammont, L., and Maatouk, H. (2016). Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *Electronic Journal of Statistics*, 10:1580–1595.

Botev, Z. I. (2017). The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B*, 79(1):125–148.

Cha, J. and Finkelstein, M. (2018). *Point Processes for Reliability Analysis: Shocks and Repairable Systems*. Springer.

Cousin, A., Maatouk, H., and Rullière, D. (2016). Kriging of financial term-structures. *European Journal of Operational Research*, 255(2):631–648.

Cox, D. R. (1955). Some Statistical Methods Connected with Series of Events. *Journal of the Royal Statistical Society. Series B*, 17(2):129–164.

Lawrence, N. D., Sanguinetti, G., and Rattray, M. (2007). Modelling transcriptional regulation using Gaussian processes. In *Neural Information Processing Systems*, pages 785–792.

López-Lopera, A. F., Bachoc, F., Durrande, N., and Roustant, O. (2018). Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):1224–1255.

López-Lopera, A. F., Bachoc, F., Durrande, N., and Roustant, O. (2019). Approximating Gaussian Process Emulators with Linear Inequality Constraints and Noisy Observations via MC and MCMC. *To appear in Proceedings in Monte Carlo and Quasi-Monte Carlo Methods*.

López-Lopera, A. F., John, S., and Durrande, N. (2019). Gaussian process modulated Cox processes under linear inequality constraints. In *International Conference on Artificial Intelligence and Statistics*, pages 1997–2006.

Maatouk, H. and Bay, X. (2016). *A New Rejection Sampling Method for Truncated Multivariate Gaussian Random Variables Restricted to Convex Sets*, pages 521–530. Springer.

Maatouk, H. and Bay, X. (2017). Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences*, 49(5):557–582.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (2001). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC Monographs on Stats. and Applied Probability. CRC Press.

Pakman, A. and Paninski, L. (2014). Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542.

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. MIT Press.

Riihimäki, J. and Vehtari, A. (2010). Gaussian processes with monotonicity information. In *Journal of Machine Learning Research: Workshop and Conference Proceedings*, pages 645–652.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B*, 39(2):172–212.

Rohmer, J. and Idier, D. (2012). A meta-modelling strategy to identify the critical offshore conditions for coastal flooding. *Natural Hazards and Earth System Sciences*, 12(9):2943–2955.

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.

Taylor, J. and Benjamini, Y. (2016). *RestrictedMVN: Multivariate Normal Restricted by Affine Constraints*. R package version 1.0.

Vanhatalo, J. and Vehtari, A. (2007). Sparse log Gaussian processes via MCMC for spatial epidemiology. In *Proceedings of Machine Learning Research*, pages 73–89.

Zhou, S., Giulani, P., Piekarewicz, J., Bhattacharya, A., and Pati, D. (2019). Reexamining the proton-radius problem using constrained Gaussian processes. *Physical Review C*, 99:055202.