Introduction to Gaussian process with inequality constraints - Application to coast flooding risk

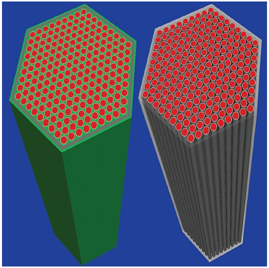
François Bachoc[1], Nicolas Durrande[2], Agnès Lagnoux[1], Andrés Felipe Lopez Lopera[3], Olivier Roustant[1]

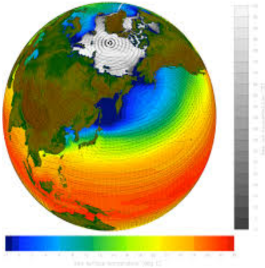
[1]Institut de Mathématiques de Toulouse
[2]Shift Lab (London)
[3]Université de Valenciennes

Workshop "Statistical methods for safety and decommissionning"
Avignon

November 2022

Computer models have become essential in science and industry !





For clear reasons : cost reduction, possibility to explore hazardous or extreme scenarios...

Computer models as expensive functions

A computer model can be seen as a deterministic function

$$f : \mathbb{X} \subset \mathbb{R}^d \to \mathbb{R}$$
$$x \mapsto f(x).$$

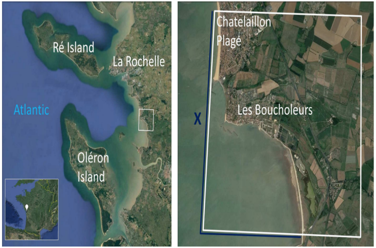
- x : tunable simulation parameter (e.g. geometry).
- $f(x)$: scalar quantity of interest (e.g. energetic efficiency).

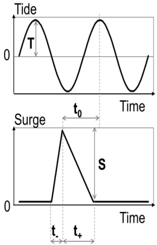
The function f is usually

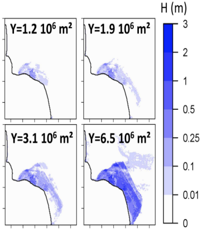
- continuous (at least)
- non-linear
- only available through evaluations $x \mapsto f(x)$.

$\Longrightarrow$ Black box model.

Figures from [Azzimonti et al., 2019].
Data and computer model MARS from BRGM.



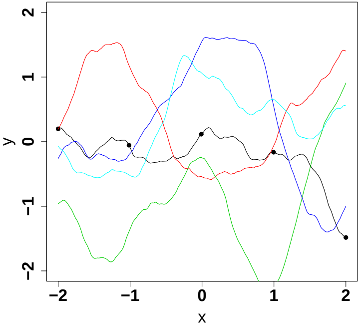




- Input x with $d = 5$.
  - : Tide (meter).
  - : Surge peak (meter).
  - : Phase difference between surge peak and high tide (hour).
  - : Time duration of raising part of surge (hour).
  - : Time duration of falling part of surge (hour).
- Onput $f(x)$.
  - Maximal flooding area (m^3).

Gaussian processes (Kriging model)

Modeling the **black box function** as a **single realization** of a Gaussian process $x \to \xi(x)$ on the domain $\mathbb{X} \subset \mathbb{R}^d$.



Usefulness

Predicting the continuous realization function, from a finite number of **observation points**.

Gaussian processes

Definition

A stochastic process $\xi : \mathbb{X} \to \mathbb{R}$ is Gaussian if for any $x_1, ..., x_n \in \mathbb{X}$, the vector $(\xi(x_1), ..., \xi(x_n))$ is a Gaussian vector.

Mean and covariance functions

The distribution of a Gaussian process is characterized by :

- Its mean function : $x \mapsto m(x) = \mathbb{E}(\xi(x))$ Can be any function $\mathbb{X} \to \mathbb{R}$.
- Its covariance function $(x_1, x_2) \mapsto k(x_1, x_2) = Cov(\xi(x_1), \xi(x_2))$.

Conditional distribution

Gaussian process ξ observed at $x_1, ..., x_n$, without noise.

Notation

- $y = (\xi(x_1), ..., \xi(x_n))^\top$.
- R is the $n \times n$ matrix $[k(x_i, x_j)]$.
- $r(x) = (k(x, x_1), ..., k(x, x_n))^\top$.

Conditional mean

The conditional mean is $m_n(x) = \mathbb{E}(\xi(x)|\xi(x_1), ..., \xi(x_n)) = r(x)^\top R^{-1} y$.

Conditional variance

The conditional variance is $k_n(x, x) = var(\xi(x)|\xi(x_1), ..., \xi(x_n)) =$
$\mathbb{E}\left[(\xi(x) - m_n(x))^2\right] = k(x, x) - r(x)^\top R^{-1} r(x)$.

Conditional distribution

Conditionally to $\xi(x_1), ..., \xi(x_n)$, ξ is a Gaussian process with (conditional) mean function m_n and (conditional) covariance function
$(u, v) \mapsto k_n(u, v) = k(u, v) - r(u)^\top R^{-1} r(v)$.

Illustration of conditional mean and variance

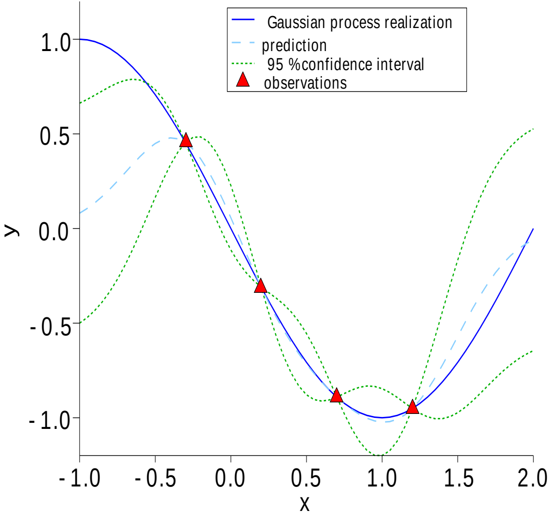
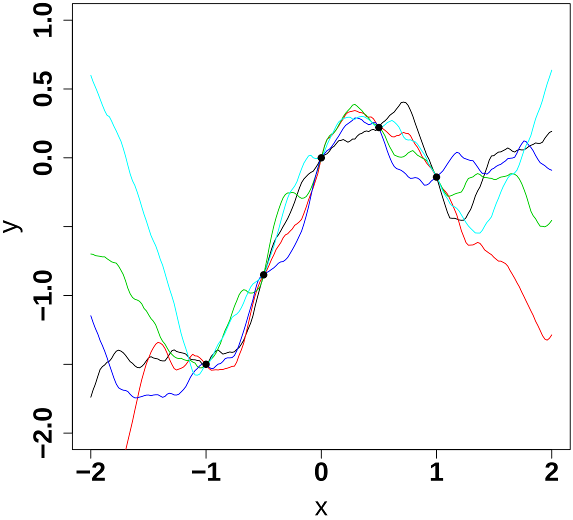


Illustration of the conditional distribution



Notation : parametric covariance function estimation

Parameterization

Covariance function model $\{k_\theta, \theta \in \Theta\}$ for the Gaussian process ξ.

- $\Theta \subset \mathbb{R}^p$.
- θ is the multidimensional covariance parameter.
- k_θ is a covariance function.

Observations

ξ is observed at $x_1, ..., x_n \in \mathbb{X}$, yielding the Gaussian vector $y = (\xi(x_1), ..., \xi(x_n))^\top$.

Estimation

Objective : build estimator $\hat{\theta}(y)$.

Notation : maximum likelihood (ML) for estimation

Explicit Gaussian likelihood function for the observation vector y.

Maximum likelihood

Define R_θ as the covariance matrix of $y = (\xi(x_1), ..., \xi(x_n))^\top$ with covariance function $k_\theta : R_\theta = [k_\theta(x_i, x_j)]_{i,j=1,\ldots,n}$.
The maximum likelihood estimator of θ is

$$\hat{\theta}_{ML} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \, \mathcal{L}_n(\theta)$$

with

$$\mathcal{L}_n(\theta) = \log(p_\theta(y)) = \log\left(\frac{1}{(2\pi)^{n/2}|R_\theta|} e^{-\frac{1}{2} y^\top R_\theta^{-1} y}\right).$$

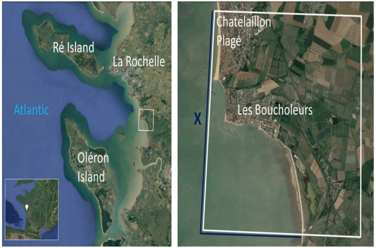
Inequality constraints

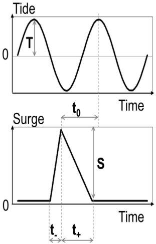
We consider a Gaussian process ξ on $\mathbb{X} = [0,1]^d$ for which we assume that additional information is available.

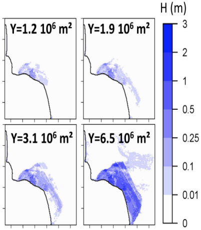
- $\xi(x)$ belongs to $[\ell, u]$ for $x \in [0,1]^d$ (**boundedness constraints**).
- $\partial\xi(x)/\partial x_i \geq 0$ for $x \in [0,1]^d$ and $i = 1, \ldots, d$ (**monotonicity constraints**).
- ξ is convex on $[0,1]^d$ (**convexity constraints**).
- Modifications and/or combinations of the above constraints.

Application examples in computer experiments.

- **Boundedness :** computer model output belongs to $\mathbb{R}^+$ (energy) or $[0,1]$ (concentration, energetic efficiency).
- **Monotonicity :** inputs are known to have positive effects (more input power $\rightarrow$ more output energy).







- Input *x*.
  - : Tide (meter). Output increases when tide increases !
  - : Surge peak (meter). Output increases when surge increases !
  - : Phase difference between surge peak and high tide (hours).
  - : Time duration of raising part of surge (hours).
  - : Time duration of falling part of surge (hours).
- Onput $f(x)$.
  - Maximal flooding area (m^3).

Impact of the constraints

Generic form of the constraints :

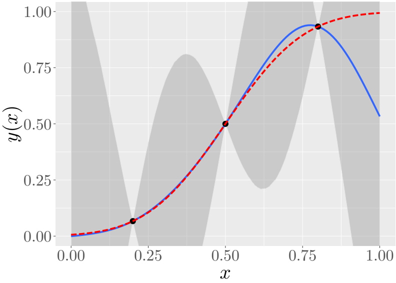
$$\xi \in \mathcal{E}$$

where $\mathcal{E}$ is a set of functions from $[0,1]^d \to \mathbb{R}$ such that $P(\xi \in \mathcal{E}) > 0$.

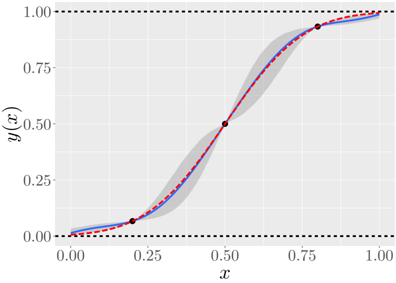
Impact.

- **New stochastic model :** The law of the realization function is $P(\xi \in . | \xi \in \mathcal{E})$.
- **New conditional distribution :** Conditional distribution of ξ given
  - $\xi(x_1) = y_1, \ldots, \xi(x_n) = y_n$ (data interpolation),
  - $\xi \in \mathcal{E}$ (inequality constraints).
- **New estimation** of the covariance parameters θ in the covariance model $\{k_\theta ; \theta \in \Theta\}$.

Target function : bounded and monotonic.



Unconstrained Gaussian process.



Constrained Gaussian process.

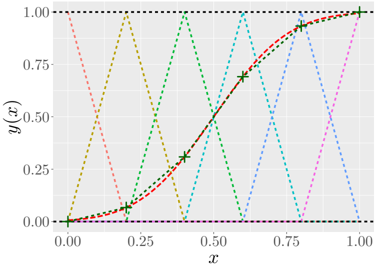
- true function
- predictive mean
- • training points
- confidence intervals

Discussion of computational aspects : handling the constraints

- For boundedness constraints, it is possible to consider models of the form $y_i = T(\xi(x_i))$ with T bijective from $\mathbb{R}$ to $[\ell, u]$ and ξ a Gaussian process.
  - No computational problem.
- For monotonicity and convexity constraints, the model $P(\xi \in .|\xi \in \mathcal{E})$ has become standard.
  - But the constraint $\xi \in \mathcal{E}$ needs to be approximated.
  - $\xi \in \mathcal{E}$ is replaced by a finite number of constraints on inducing points in [Da Veiga and Marrel, 2012, Golchi et al., 2015].

$$(\partial_i \xi)(s) \geq 0, s \in [0,1]^d \qquad \approx \qquad (\partial_i \xi)(s_j) \geq 0, j = 1, \ldots, m.$$

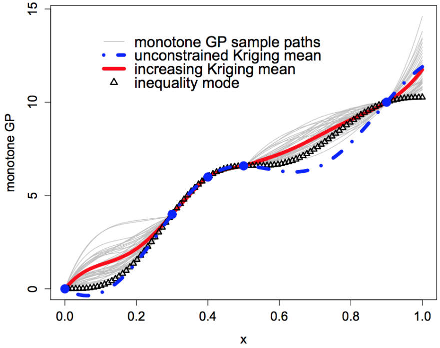
  - ξ is replaced by a finite-dimensional approximation ξ_m in [López-Lopera et al., 2018, Maatouk and Bay, 2017].



Discussion of computational aspects : conditional distribution

In the frame of [López-Lopera et al., 2018, Maatouk and Bay, 2017].

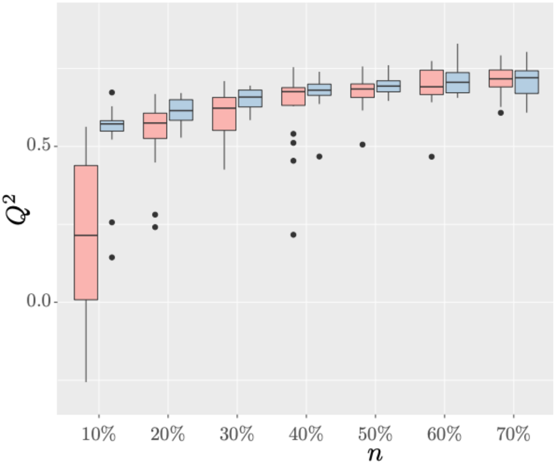
- The mode is the “most likely” function for ξ_m, obtained by quadratic optimization with linear constraints.
- Conditional realizations of ξ_m can be sampled approximately, for instance by Hamiltonian Monte Carlo for truncated Gaussian vectors [Pakman and Paninski, 2014].



Results on coastal flooding example

Gaussian process predictive score.

- Without constraints.
- With constraints.



An application to nuclear engineering

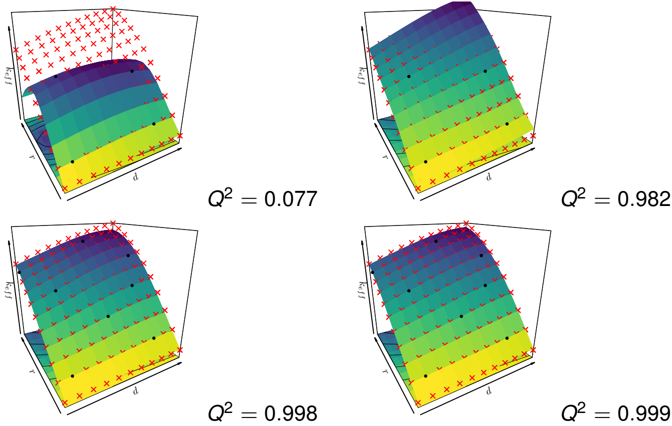


Figure – Two dimensional nuclear engineering example. **Radius** and **density** of uranium sphere $\Longrightarrow$ **criticality coefficient**. Mononoicity constraints. Left : unconstrained Gaussian process models. Right : constrained Gaussian process models. The Q^2 measures the prediction quality and should be close to 1.

Constrained maximum likelihood estimator

The constrained maximum likelihood estimator for θ is

$$\hat{\theta}_{cML} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \, \mathcal{L}_{\mathcal{C},n}(\theta)$$

with

$$\begin{aligned}\mathcal{L}_{\mathcal{C},n}(\theta) &= \log(p_\theta(y|\xi \in \mathcal{E})) \\ &= \log(p_\theta(y)) - \log(\mathbb{P}_\theta(\xi \in \mathcal{E})) + \log(\mathbb{P}_\theta(\xi \in \mathcal{E}|y)).\end{aligned}$$

- The additional terms $\log(\mathbb{P}_\theta(\xi \in \mathcal{E}))$ and $\log(\mathbb{P}_\theta(\xi \in \mathcal{E}|y))$ have no explicit expressions.
- They need to be approximated by numerical integration or Monte Carlo : [Genz, 1992, Botev, 2017].

Main questions :

- $\hat{\theta}_{ML}$ ignores the constraints. Is it biased conditionally to the constraints ?
  - For instance if $\hat{\theta}_{ML}$ is the variance estimator, if the true variance is 4 and if the constraints are $\xi \in [-1, 1]$, does $\hat{\theta}_{ML}$ underestimate the variance ?
- Does $\hat{\theta}_{cML}$ improve over $\hat{\theta}_{ML}$ by taking the constraints into account ?

Asymptotic normality result : Matérn model

Matérn family of covariance functions :

$$K_\theta(u,v) = K_{\sigma^2,\rho}(u,v) = \sigma^2 K_{\text{Matérn}}\left(\frac{u-v}{\rho}\right).$$

Shown in [Kaufman and Shaby, 2013] using results from [Du et al., 2009, Wang and Loh, 2011] :

$$\sqrt{n}\left(\frac{\widehat{\sigma}^2_{ML}}{\widehat{\rho}^{2\nu}_{ML}} - \frac{\sigma_0^2}{\rho_0^{2\nu}}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, 2\left(\frac{\sigma_0^2}{\rho_0^{2\nu}}\right)^2\right).$$

Theorem [Bachoc et al., 2019]

We have

$$\sqrt{n}\left(\frac{\widehat{\sigma}^2_{ML}}{\widehat{\rho}^{2\nu}_{ML}} - \frac{\sigma_0^2}{\rho_0^{2\nu}}\right) \xrightarrow[n\to+\infty]{\mathcal{L}|\xi\in\mathcal{E}} \mathcal{N}\left(0, 2\left(\frac{\sigma_0^2}{\rho_0^{2\nu}}\right)^2\right)$$

and

$$\sqrt{n}\left(\frac{\widehat{\sigma}^2_{cML}}{\widehat{\rho}^{2\nu}_{cML}} - \frac{\sigma_0^2}{\rho_0^{2\nu}}\right) \xrightarrow[n\to+\infty]{\mathcal{L}|\xi\in\mathcal{E}} \mathcal{N}\left(0, 2\left(\frac{\sigma_0^2}{\rho_0^{2\nu}}\right)^2\right).$$

- Same conclusions as for the estimation of a variance parameter.

An illustration

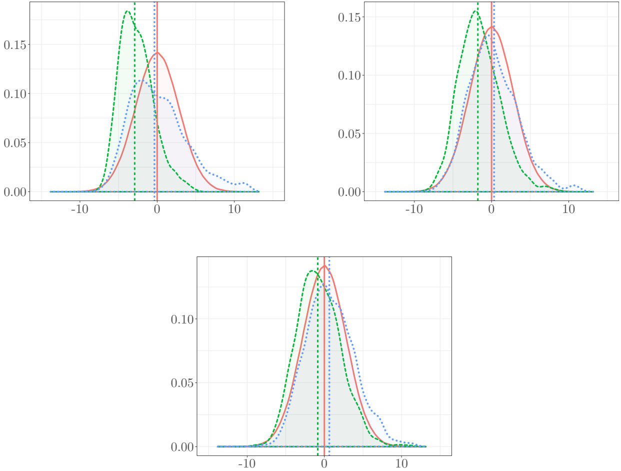
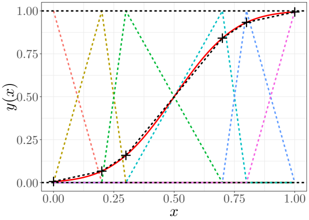


Figure – An example with the estimation of σ_0^2 with boundedness constraints. Distribution of $n^{1/2}(\hat{\sigma}^2 - \sigma_0^2)$. $n = 20$ (top left), $n = 50$ (top right) and $n = 80$ (bottom). Green : ML. Blue : cML. Red : Gaussian limit.



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

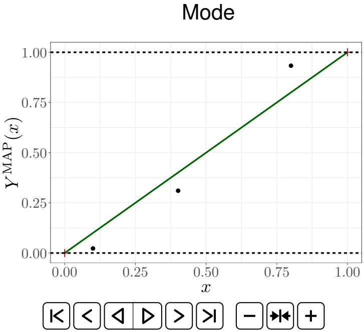
$$\{t_0, \dots, t_m\}, \quad \text{with} \quad 0 = t_0 < \dots < t_m = 1.$$

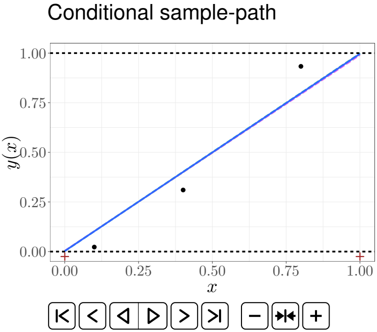
- Here, we aim at adding a new knot t (where ?).
- To do so, we aim at *maximising the total modification of the mode* :

$$I(t) = \int_{[0,1]} \left(\widehat{Y}_{+t}(x) - \widehat{Y}(x) \right)^2 dx. \tag{1}$$

The integral in (1) has a closed-form expression.

1D example under boundedness and monotonicity constraints

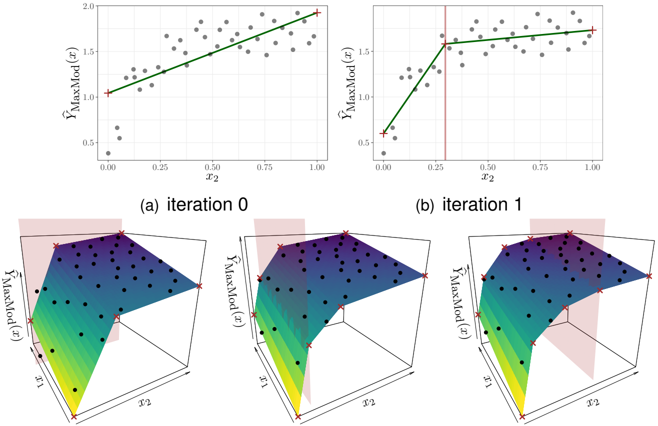




● Pbservation points + Knots ■ Mode
■ Predictive mean ■ 90% confidence intervals

2D example under monotonicity constraints

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



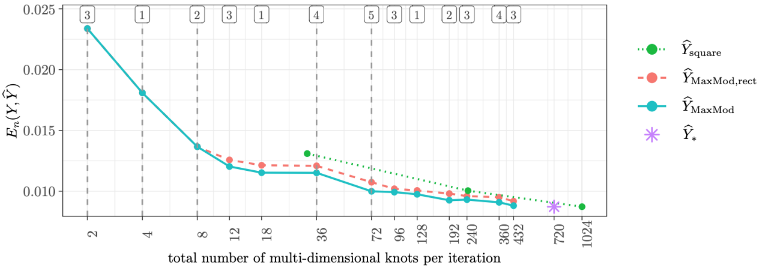
(a) iteration 0 (b) iteration 1

(c) iteration 2 (d) iteration 3 (e) iteration 4

● training points + knots ■ mode

MaxMod results on coastal example

- $E_n(Y, \hat{Y})$: relative square error.
- $\hat{Y}_{\text{square}}$: regularly spaced knots, identical number per variable.
- $\hat{Y}_{\text{MaxMod,rect}}$: regularly spaced knots, numbers per variable given by MaxMod.
- $\hat{Y}_*$: optimized by hand in a previous study.



Approach	m	$E_n(Y,\hat{Y})$ $[1 \times 10^{-3}]$	CPU time [s]		
			Training step	Computation of $\hat{Y}$	Sampling step with 100 realizations
$\hat{Y}_{\text{square}}$	1024	8.72	49.1	8.03	non converged after 1 day
$\hat{Y}_{\text{MaxMod}}$	432	8.81	949.5	0.58	108.72

Conclusion

Summary

- Inequality constraints correspond to additional information (e. g. physical knowledge).
- Taking them into account can significantly improve the predictions.
- With a computational cost (explicit $\Longrightarrow$ Monte Carlo).
- Asymptotically, we do not see an impact of the constraints and ML $\approx$ cML.
- MaxMod algorithm for higher dimension.

References

- Constrained Gaussian processes : [López-Lopera et al., 2018].
- Constrained Maximum Likelihood : [Bachoc et al., 2019].
- MaxMod : [Bachoc et al., 2022].
- Extension of MaxMod for additive models : [López-Lopera et al., 2022].
- R package LineqGPR : `https://github.com/anfelopera/lineqGPR`.

Thank you for your attention !

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.

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

Bachoc, F., López-Lopera, A. F., and Roustant, O. (2022).
Sequential construction and dimension reduction of Gaussian processes under inequality constraints.
SIAM Journal on Mathematics of Data Science, 4(2) :772–800.

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

Da Veiga, S. and Marrel, A. (2012).
Gaussian process modeling with inequality constraints.
Annales de la faculté des sciences de Toulouse Mathématiques, 21(6) :529–555.

Du, J., Zhang, H., and Mandrekar, V. (2009).
Fixed-domain asymptotic properties of tapered maximum likelihood estimators.
The Annals of Statistics, 37 :3330–3361.

Genz, A. (1992).
Numerical computation of multivariate normal probabilities.
Journal of Computational and Graphical Statistics, 1 :141–150.

Golchi, S., Bingham, D., Chipman, H., and Campbell, D. (2015).
Monotone emulation of computer experiments.
SIAM/ASA Journal on Uncertainty Quantification, 3(1) :370–392.

Kaufman, C. and Shaby, B. (2013).
The role of the range parameter for estimation and prediction in geostatistics.
Biometrika, 100 :473–484.

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., and Roustant, O. (2022).
High-dimensional additive Gaussian processes under monotonicity constraints.
In *NeurIPS*.

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

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

Wang, D. and Loh, W.-L. (2011).
On fixed-domain asymptotics and covariance tapering in Gaussian random field models.
Electronic Journal of Statistics, 5 :238–269.