

# INSA – Gaussian processes

## Introduction

---

Andrés F. López-Lopera

The French Aerospace Lab ONERA, France  
Information Processing and Systems Department (DTIS)  
Multidisciplinary Methods, Integrated Concepts (M2CI) Research Unit

# Who am I?



Andrés F. López-Lopera

## Colombia

2008-2013

**Electrical Eng., Universidad Tecnológica de Pereira**

- Machine learning and signal processing

2014-2015

**M.Sc. in Electrical Eng., Universidad Tecnológica de Pereira**

- Probabilistic modelling using Gaussian processes (GPs)

## France

2016-2019

**PhD in Applied Mathematics, Mines Saint-Étienne**

- Joint supervision: *Institut de Mathématiques de Toulouse*
- GPs under inequality constraints
- Applications: nuclear risk assessment, coastal flooding

2019-2020

**Postdoctoral Research, Institut de Mathématiques de Toulouse**

- Joint supervision: *The French Geological Survey BRGM*
- Multi-output GPs & coastal flooding

2020-2021

**Postdoctoral Research, The French Aerospace Lab ONERA**

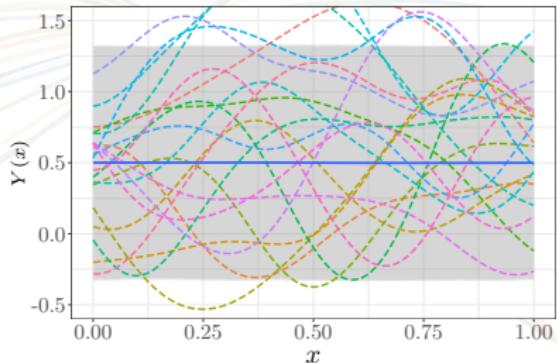
- Multi-fidelity GPs & aerodynamics (wind tunnel tests)

# Research interests

- My research interests include:
  - Applied mathematics
  - Machine learning & computer science
  - Probabilistic modelling, Bayesian inference and optimisation, **GPs**, etc.
- With applications to:
  - Electrical engineering and signal processing
  - Risk assessment (nuclear, coastal, etc.)
  - Aerodynamics (wind tunnel test)
  - Artificial intelligence

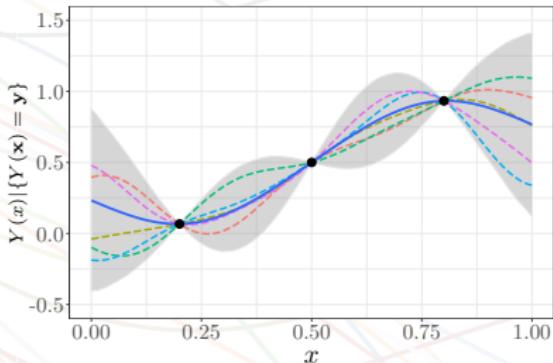
# Gaussian processes (GPs) as flexible priors over functions

GP prior



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

GP posterior

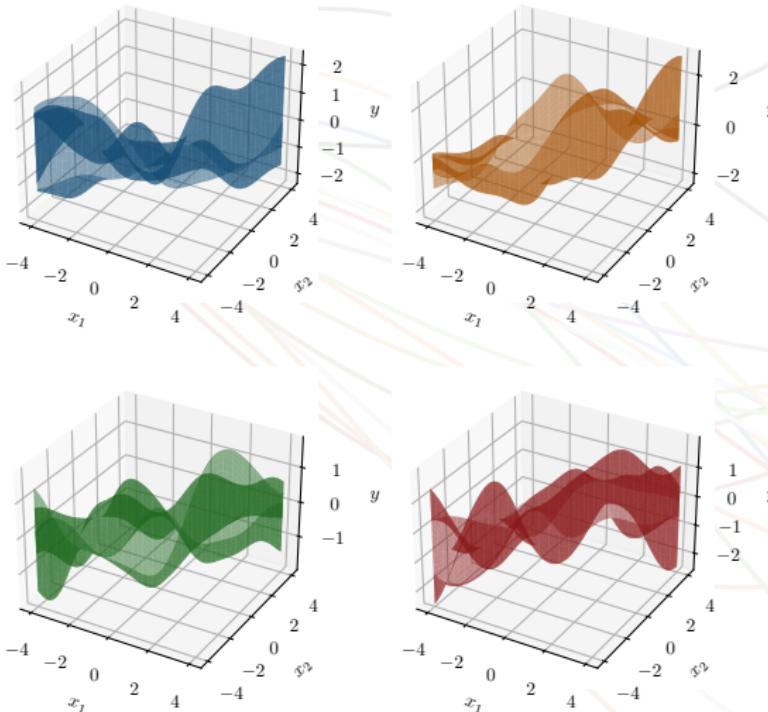


$$Y | \{Y(x) = y\} \sim \mathcal{GP}(m_{\text{cond}}, c_\theta)$$

■ mean function ■ prediction intervals ■ ■ ... ■ samples

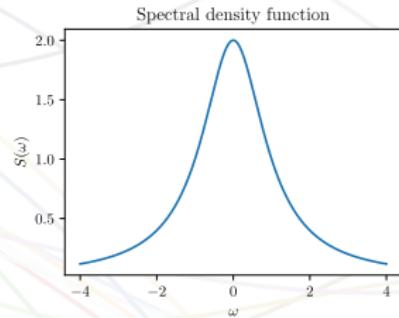
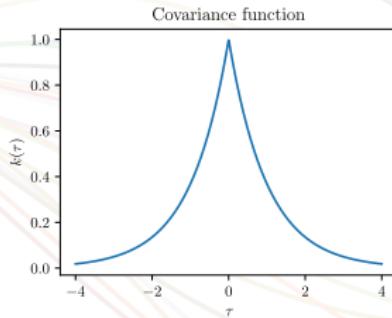
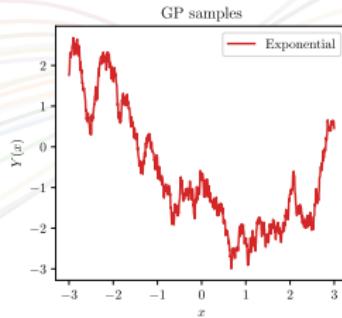
- Interpolation conditions:  $(x, y) = (x_i, y_i)_{i=1}^n$

# Gaussian processes (GPs) as flexible priors over functions



Gaussian random fields

# Kernel functions

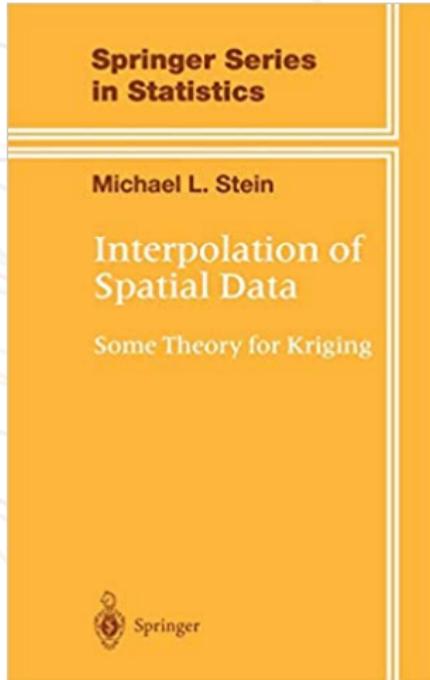
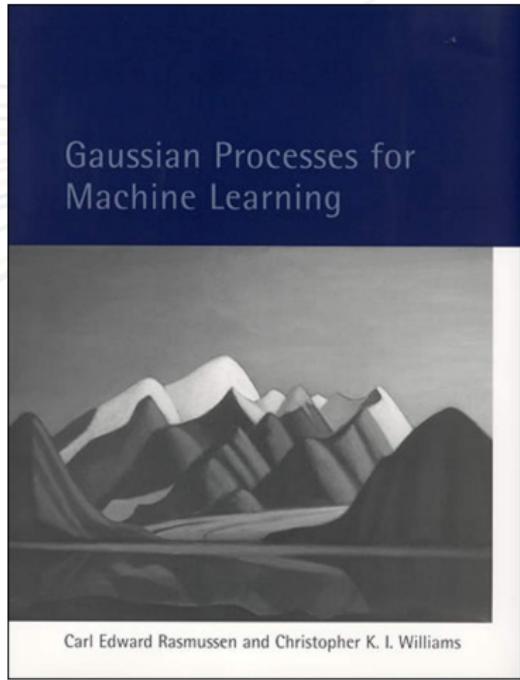


- Every kernel  $k$  is the **covariance function** of some centred Gaussian **stochastic process**  $Y$ : e.g. *Ornstein-Uhlenbeck process*
- Every **spectral density**  $S(\omega)$  defines a (stationary) kernel
- If  $k$  is a kernel, there exists a unique **RKHS** with  $k$  as its reproducing kernel. If  $k$  is a reproducing kernel, then it is a covariance function

# Outline

- In these lectures:
  1. A recap of Gaussian processes with applications
  2. Spectral representation and Bochner's theorem
  3. Regularity conditions (e.g. continuity, differentiability)
  4. An introduction to reproducing kernel Hilbert-spaces (RKHS)
- 2 practical sessions (~3.5h): Python (Jupyter) or R (Jupyter + IRkernel)
- Material can be found at: <https://anfelopera.github.io/teaching/>

## Main references



<http://www.gaussianprocess.org/gpml/>

<https://www.springer.com/gp/book/9780387986296>

## Additional references

- Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.
- Chris Chatfield. *The Analysis of Time Series: An Introduction*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2016.
- Harald Cramér and M. Ross Leadbetter. *Stationary and Related Stochastic Processes - Sample Function Properties and Their Applications*. Wiley, 1967.
- Marc G. Genton. Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research*, 2001.
- Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2021.
- Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. MIT Press, 2005.
- Olivier Roustant. Statistical models and methods for computer experiments. Technical report, 2011. HDR – Université Jean Monnet, Saint-Étienne, France.
- Arno Solin. *Machine learning with signal processing*. ICML – TUTORIAL, 2020.
- Michael L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, 1999.
- Akiva M. Yaglom. *Correlation Theory of Stationary and Related Random Functions*. Springer, 1987.

## Gaussian processes

---

# Gaussian processes

- A GP  $\{Y(x), x \in \mathbb{R}^d\}$  is a collection of random variables, any finite number of which have a joint Gaussian distribution [Rasmussen and Williams, 2005]
- $Y$  is completely defined by its mean  $m$  and covariance (kernel)  $k$  functions:

$$Y \sim \mathcal{GP}(m, k), \quad (1)$$

where

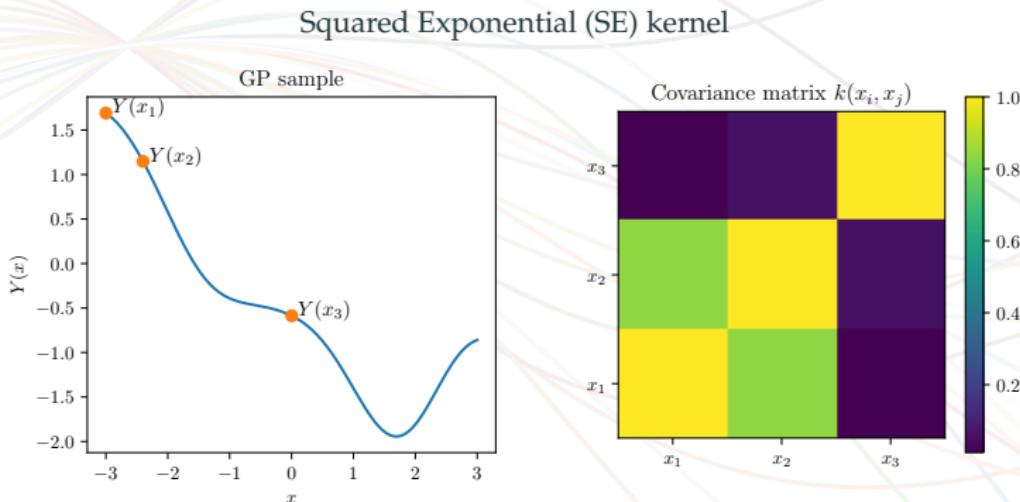
$$\begin{aligned} & \text{(trend)} \quad m(x) = \mathbb{E} \{Y(x)\}, \\ & \text{(correlation)} \quad k(x, x') = \text{cov} \{Y(x), Y(x')\}, \quad \text{for } x, x' \in \mathbb{R}^d. \end{aligned} \quad (2)$$

- The operator  $\mathbb{E}$  denotes the expectation of random variables (r.v's), and the covariance operator is given by

$$\text{cov} \{Y(x), Y(x')\} = \mathbb{E} \{[Y(x) - m(x)][Y(x') - m(x')]\}.$$

# Gaussian processes

**Note.** Independence between  $Y(x)$ ,  $Y(x')$  implies  $k(x, x') = 0$ .



- If  $Y(x)$ ,  $Y(x')$  are correlated, then  $k(x, x') \neq 0$
- If  $Y(x)$ ,  $Y(x')$  are non-correlated, then  $k(x, x') = 0$

- It is common to assume that  $Y$  is centred, i.e.  $m(\cdot) = 0$ .
- Then,  $Y$  is completely defined by its kernel  $k$ :

$$k(x, x') = \text{cov} \{ Y(x), Y(x') \} = \mathbb{E} \{ Y(x)Y(x') \}, \quad (3)$$

**Exercise.** Show that  $Z \sim \mathcal{GP}(m, k)$  can be written in terms of  $Y \sim \mathcal{GP}(0, k)$ :

$$Z(x) = m(x) + Y(x). \quad (4)$$

# Gaussian process regression

- Let  $\{Y(x), x \in \mathbb{R}^d\}$  be a centred GP with covariance function  $k$
- Consider a set of observations  $(x_i, y_i)_{1 \leq i \leq n}$  for  $n \in \mathbb{N}$
- In regression tasks, we aim at computing the distribution of the conditional process:

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

- This conditional process is also GP-distributed with conditional mean and covariance functions given by

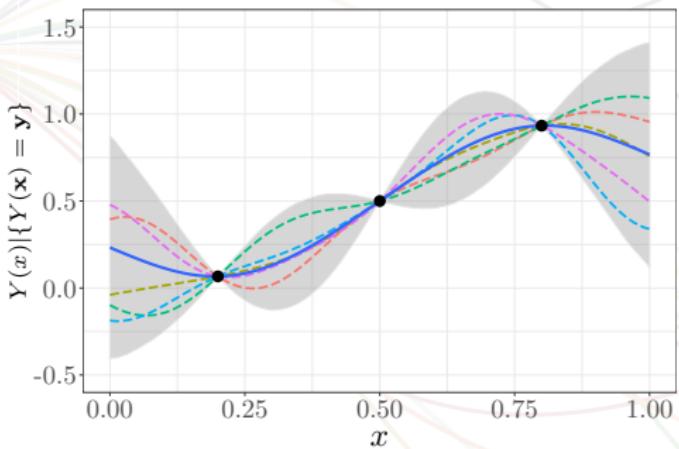
$$\mu(x) = k^\top(x) \mathbf{K}_n^{-1} y,$$

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

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

# Gaussian process regression

GP regression



■ conditional mean ■ confidence intervals ■ GP realisations

● training data:  $(x, y) = (x_i, y_i)_{i=1}^n$

[Link]

# Gaussian process regression with noisy observation

- For noisy observations, we have the conditional process:

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

with additive noises  $\varepsilon_i \sim \mathcal{N}(0, \tau^2)$ , for  $i = 1, \dots, n$ , and noise variance  $\tau^2$

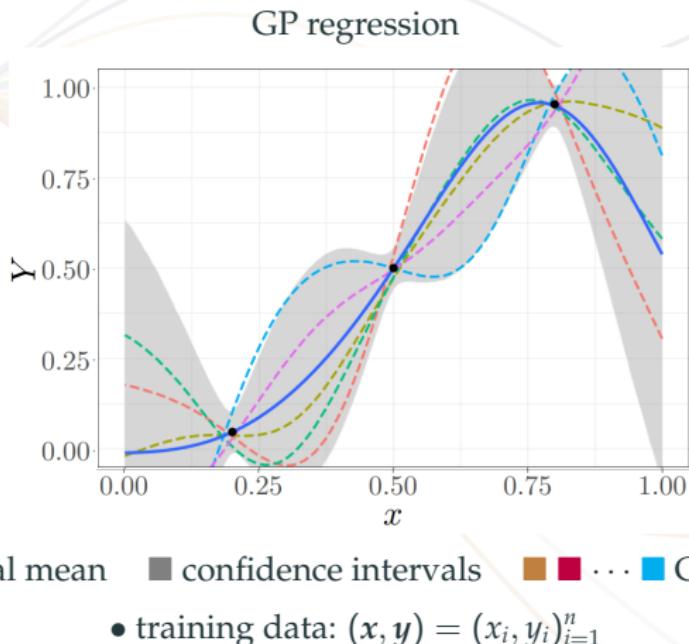
- This conditional process is also GP-distributed with conditional mean and covariance functions given by

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

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

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

# Gaussian process regression with noisy observation



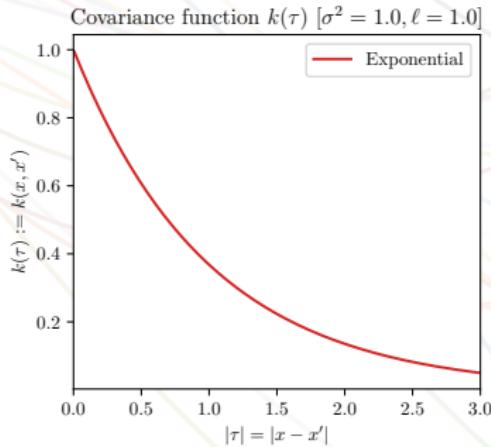
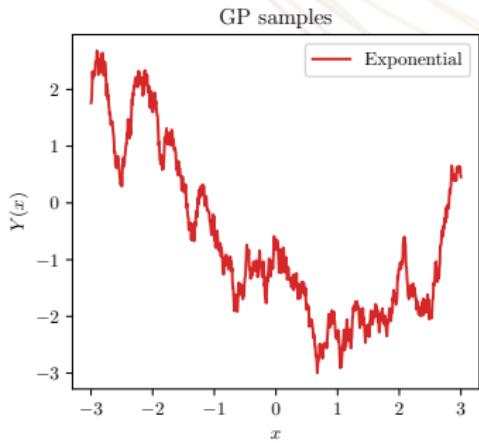
## Kernel functions

---

# Kernel functions

- In previous lectures, the exponential (Ornstein-Uhlenbeck) kernel function has been studied:

$$k(x, x') = \sigma^2 \exp \left\{ -\frac{|x - x'|}{\ell} \right\}.$$



## Definition (Symmetry)

Let  $\mathcal{X}$  be a non-empty set (e.g.  $\mathcal{X} \subset \mathbb{R}^d$ ). A function  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is symmetric if, for all  $x, x' \in \mathcal{X}$ :

$$k(x, x') = k(x', x).$$

## Definition (Positive semi-definiteness, p.s.d)

$k$  is p.s.d. if for all  $n \in \mathbb{N}$ , and for all  $a_1, \dots, a_n \in \mathbb{R}, x_1, \dots, x_n \in \mathcal{X}$ :

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j k(x_i, x_j) \geq 0.$$

## Definition (Covariance functions)

$k$  is a valid covariance function (or kernel) on  $\mathcal{X}$  if it is symmetric and p.s.d.

- **Remember.** Every kernel  $k$  is the covariance function of some centred (Gaussian) stochastic process.
- Then, it is possible to design dedicated kernels for encoding regularity assumptions in GPs [Genton, 2001], e.g.:
  - smoothness (continuity & differentiability)
  - periodicity, quasi-periodicity
  - stationarity
  - isotropy (homogeneity)

## Definition (Stationary kernel functions)

A kernel function  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ , with  $\mathcal{X} \subset \mathbb{R}^d$ , is **stationary** if, for all  $x, x' \in \mathcal{X}$ ,  $k(x, x')$  only depends on  $x - x'$ .

- We denote  $k(\tau) := k(x - x')$  (abuse of notation)

## Definition (Isotropic kernel functions)

A kernel  $k$  is **isotropic** (or homogeneous) if  $k(x, x')$  only depends on  $\|x - x'\|$ .

# Examples of 1D kernels

- Some classic 1D stationary kernels are [Genton, 2001]:

Squared Exponential (SE):  $k_{\sigma^2, \ell}(\tau) = \sigma^2 \exp \left\{ -\frac{1}{2} \frac{\tau^2}{\ell^2} \right\},$

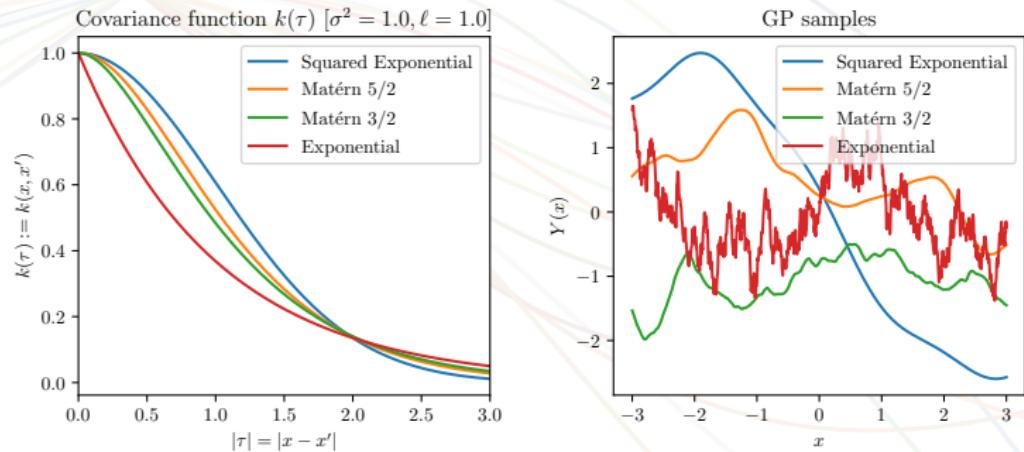
Matérn 5/2:  $k_{\sigma^2, \ell}(\tau) = \sigma^2 \left( 1 + \sqrt{5} \frac{|\tau|}{\ell} + \frac{5}{3} \frac{\tau^2}{\ell^2} \right) \exp \left\{ -\sqrt{5} \frac{|\tau|}{\ell} \right\},$

Matérn 3/2:  $k_{\sigma^2, \ell}(\tau) = \sigma^2 \left( 1 + \sqrt{3} \frac{|\tau|}{\ell} \right) \exp \left\{ -\sqrt{3} \frac{|\tau|}{\ell} \right\},$

Exponential:  $k_{\sigma^2, \ell}(\tau) = \sigma^2 \exp \left\{ -\frac{|\tau|}{\ell} \right\},$

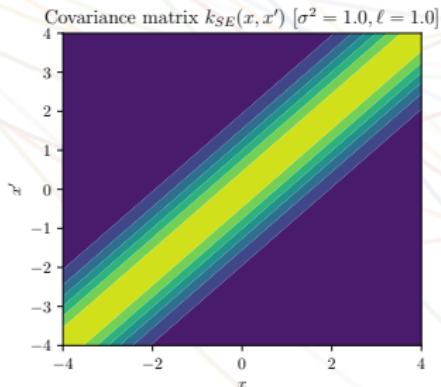
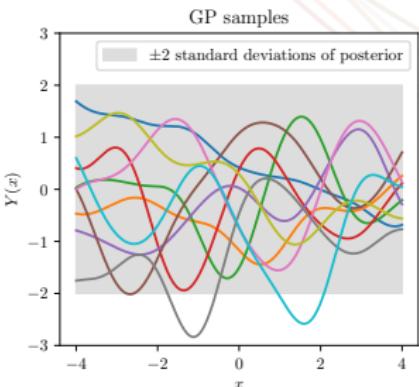
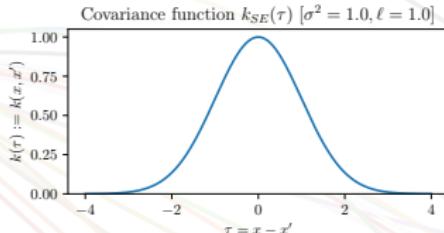
with variance parameter  $\sigma^2$  and length-scale parameter  $\ell$ .

# Examples of 1D kernels



# Examples of 1D kernels

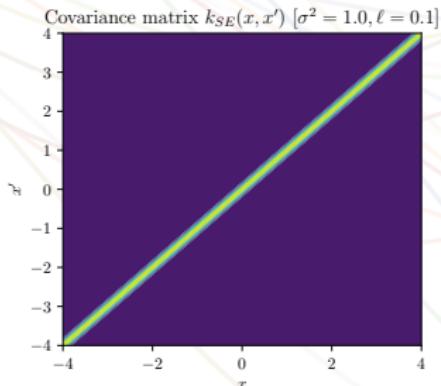
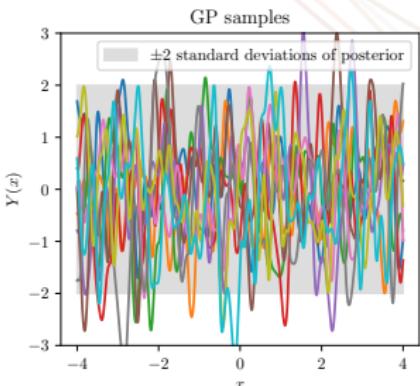
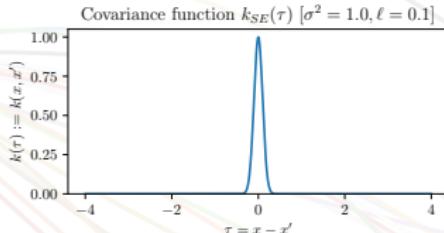
Squared Exponential (SE):  $k_{\sigma^2, \ell}(\tau) = \sigma^2 \exp \left\{ -\frac{1}{2} \frac{\tau^2}{\ell^2} \right\}$



Effect of the variance  $\sigma^2$  and the length-scale  $\ell$  on GP samples

# Examples of 1D kernels

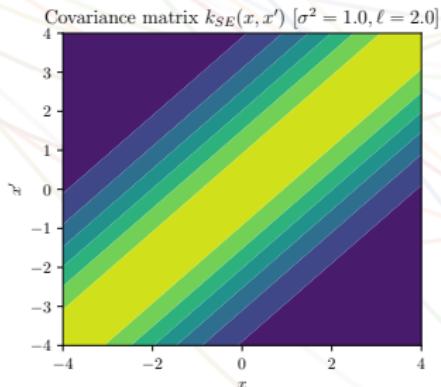
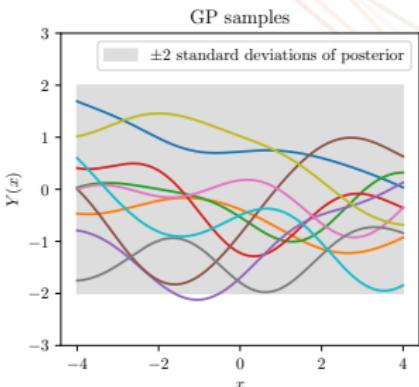
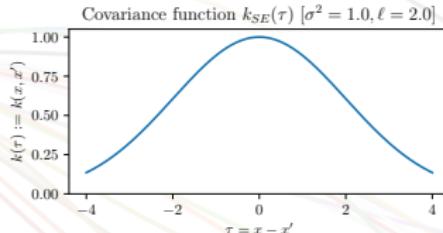
Squared Exponential (SE):  $k_{\sigma^2, \ell}(\tau) = \sigma^2 \exp \left\{ -\frac{1}{2} \frac{\tau^2}{\ell^2} \right\}$



Effect of the variance  $\sigma^2$  and the length-scale  $\ell$  on GP samples

# Examples of 1D kernels

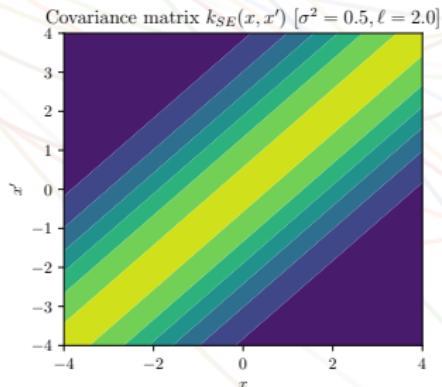
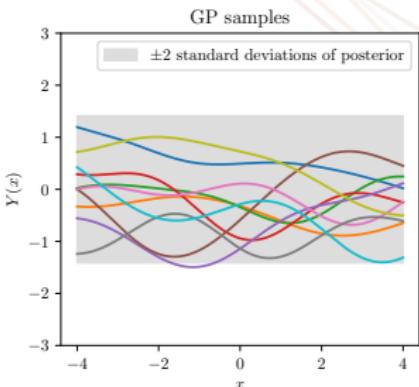
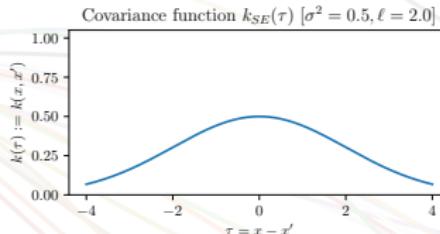
Squared Exponential (SE):  $k_{\sigma^2, \ell}(\tau) = \sigma^2 \exp \left\{ -\frac{1}{2} \frac{\tau^2}{\ell^2} \right\}$



Effect of the variance  $\sigma^2$  and the length-scale  $\ell$  on GP samples

# Examples of 1D kernels

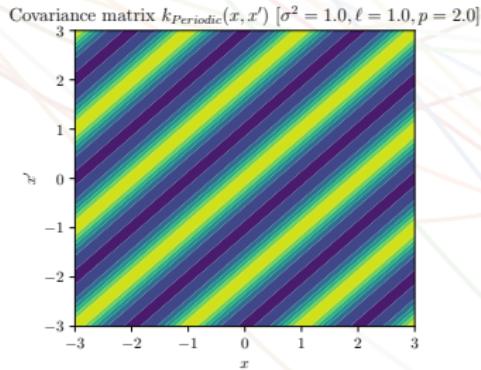
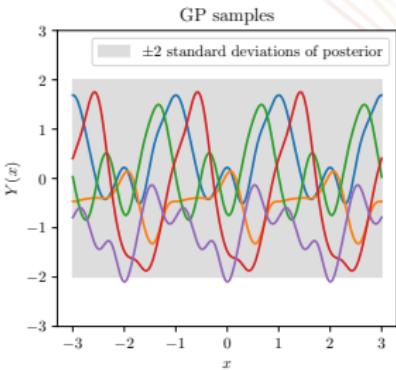
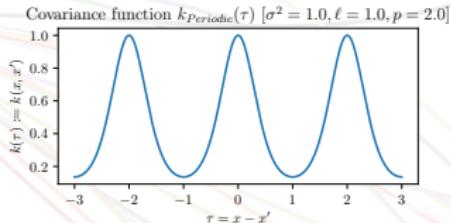
Squared Exponential (SE):  $k_{\sigma^2, \ell}(\tau) = \sigma^2 \exp \left\{ -\frac{1}{2} \frac{\tau^2}{\ell^2} \right\}$



Effect of the variance  $\sigma^2$  and the length-scale  $\ell$  on GP samples

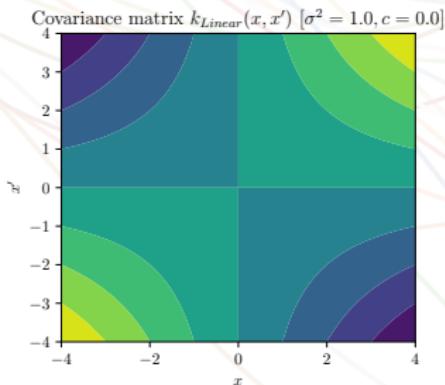
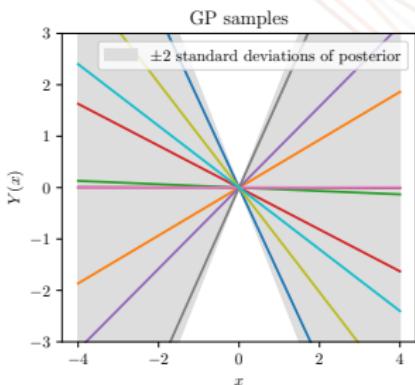
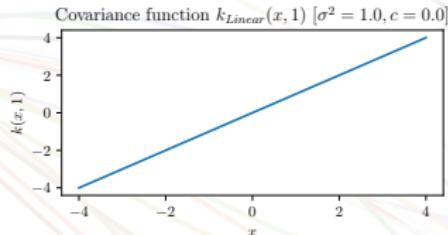
# Examples of 1D kernels

Periodic kernel:  $k_{\sigma^2, \ell, p}(\tau) = \sigma^2 \exp \left\{ -\frac{2}{\ell^2} \sin^2 \left[ \frac{\pi}{p} \tau \right] \right\}$

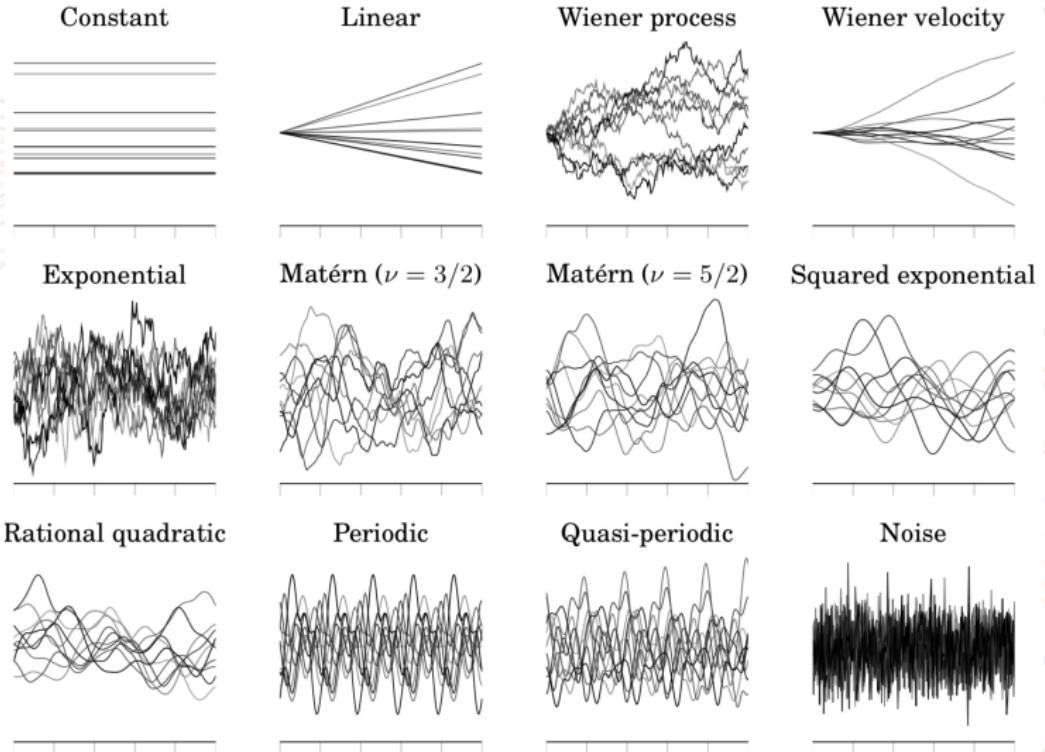


# Examples of 1D kernels

Linear kernel:  $k_{\sigma^2, c}(x, x') = \sigma^2(x - c)(x' - c)$



# Examples of 1D kernels



Examples of GP samples [Solin, 2020]

## Building new kernels from other ones

- We can also create new kernels by combining predefined ones, e.g.:

Sum of kernels:

$$k(x, x') = k_1(x, x') + k_2(x, x')$$

Product of kernels:

$$k(x, x') = k_1(x, x') \cdot k_2(x, x')$$

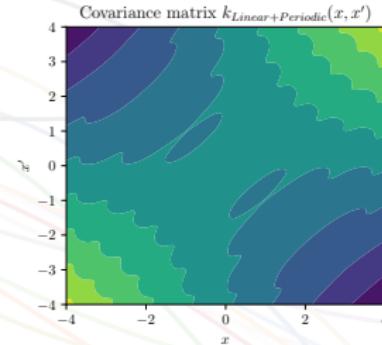
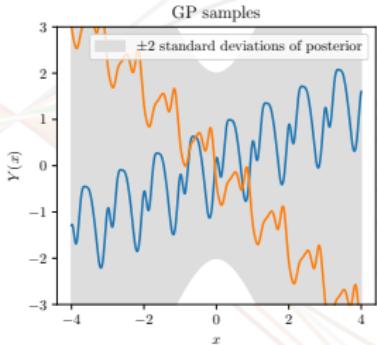
Composed with a function:

$$k(x, x') = k_1(\phi(x), \phi(x'))$$

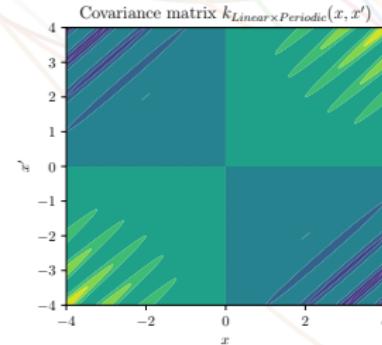
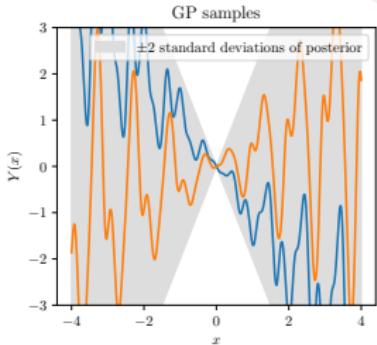
**Exercise.** Show that all the previous operations preserve the p.s.d.

# Building new kernels from other ones

$$k(x, x') = k_{\text{Linear}}(x, x') + k_{\text{Periodic}}(x, x')$$



$$k(x, x') = k_{\text{Linear}}(x, x') \times k_{\text{Periodic}}(x, x')$$



# A visual exploration of GPs

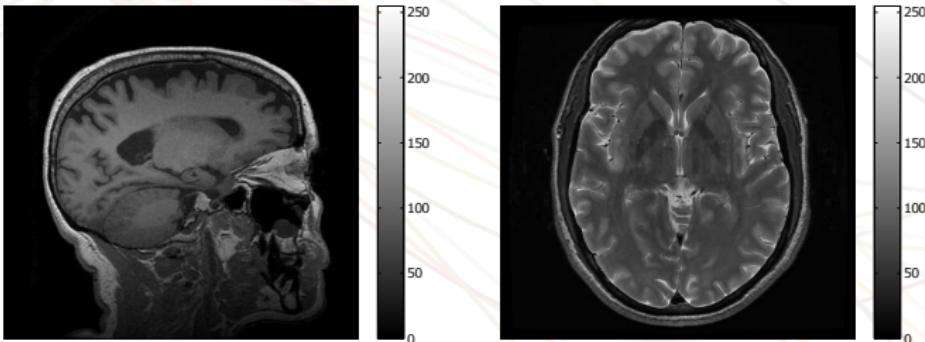
- **Durrande [2017]: Gaussian process playground** [[Link](#)]
- **Görtler et al. [2019]: A visual exploration of Gaussian processes** [[Link](#)]
- **Damianou [2016]: A Python notebook on Gaussian processes** [[Link](#)]

## **Applications of Gaussian processes**

---

# Applications of Gaussian processes

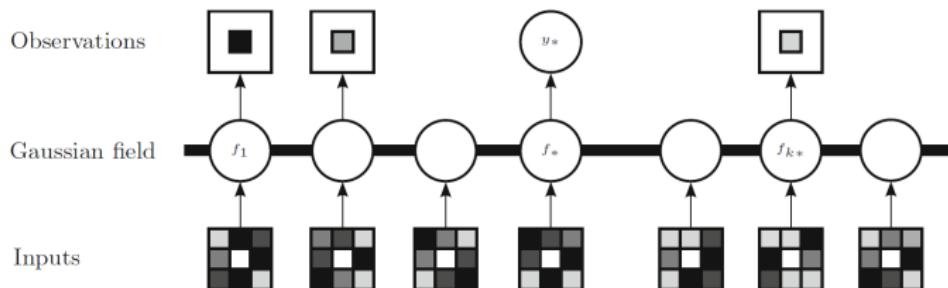
## Neuroscience: magnetic resonance imaging (MRI)



- H. Vargas, A. López-Lopera, M. A. Ivarez, A. Orozco, J. Hernández and N. Malpica:  
Gaussian processes for slice-based super-resolution MR images  
Lecture Notes in Computer Science (LNCC), 2015

# Applications of Gaussian processes

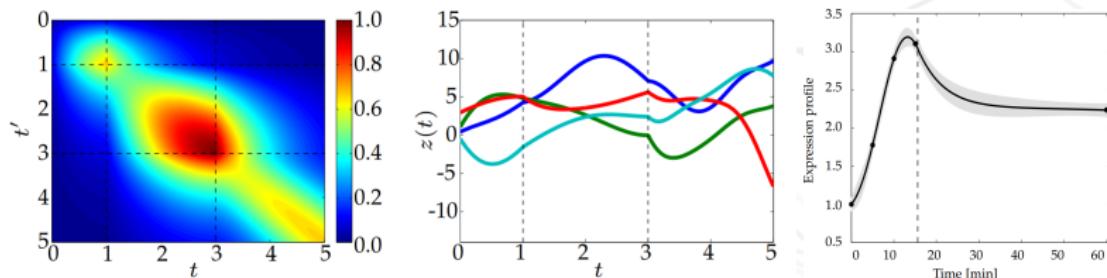
## Neuroscience: magnetic resonance imaging (MRI)



- H. Vargas, A. López-Lopera, M. A. Ivarez, A. Orozco, J. Hernández and N. Malpica:  
Gaussian processes for slice-based super-resolution MR images  
Lecture Notes in Computer Science (LNCC), 2015

# Applications of Gaussian processes

## Biology: prediction of protein concentrations



We considered a system of coupled differential equations:

$$\frac{dy_d(t)}{dt} + \gamma_d y_d(t) = B_d + \sum_{r=1}^R S_{r,d} u_r(t),$$

with  $u_1, \dots, u_R$  being independent GPs

- A. F. López-Lopera and M. A. Alvarez:

Switched latent force models for reverse-engineering transcriptional regulation in genes  
IEEE/ACM Transaction on Computational Biology and Bioinformatics, 2017

# Applications of Gaussian processes

- Consider the first-order differential equation given by

$$\frac{dY(t)}{dt} + \gamma Y(t) = SU(t), \quad (5)$$

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

- By assuming the initial condition  $Y(0) = 0$ , we obtain

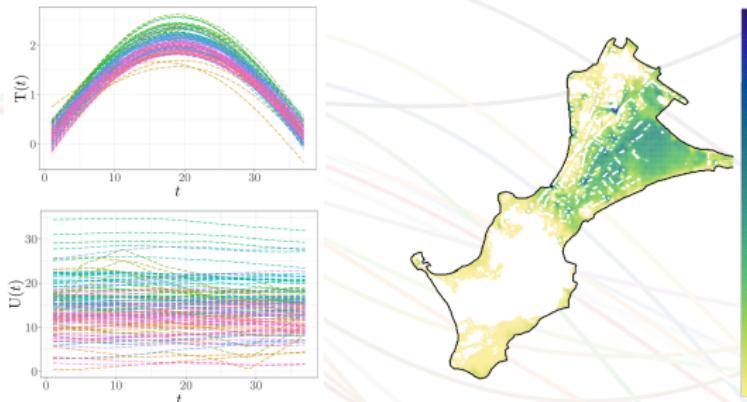
$$Y(t) = Sc(t) \int_0^t U(\tau) \exp(\gamma\tau) d\tau, \quad \text{avec } c(t) = \exp(-\gamma t)$$

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

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

# Applications of Gaussian processes

## Risk assessment: coastal flooding



We considered a separable kernel given by

$$k((\mathcal{F}, x), (\mathcal{F}', x')) = k_f(\mathcal{F}, \mathcal{F}') k_x(x, x'),$$

with  $k_f : \mathcal{F}(\mathcal{T}, \mathbb{R})^Q \times \mathcal{F}(\mathcal{T}, \mathbb{R})^Q \rightarrow \mathbb{R}$  (kernel for the functional inputs) and  $k_x : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$  (spatial kernel)

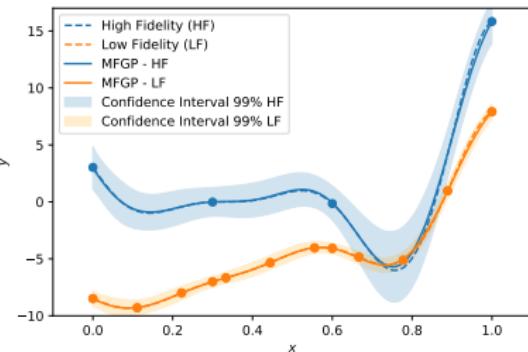
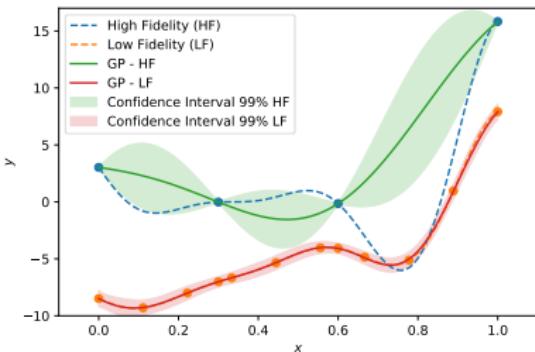
- A. F. López-Lopera, D. Idier, J. Rohmer and F. Bachoc:

Multi-output Gaussian processes with functional data: A study on coastal flood hazard assessment

Submitted, 2020

# Applications of Gaussian processes

## Aerodynamics: multi-fidelity Gaussian processes



- We considered the autoregressive model:

$$Y_1(x) = \rho Y_0(x) + \delta(x),$$

with  $Y_0$  and  $\delta$  independent (centred) GPs and  $\rho \in \mathbb{R}$ .

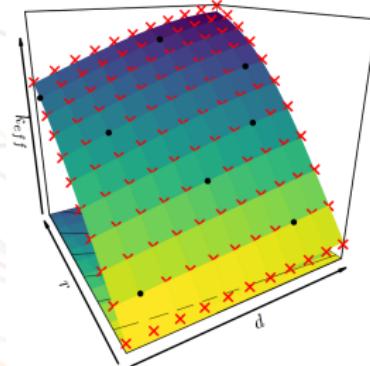
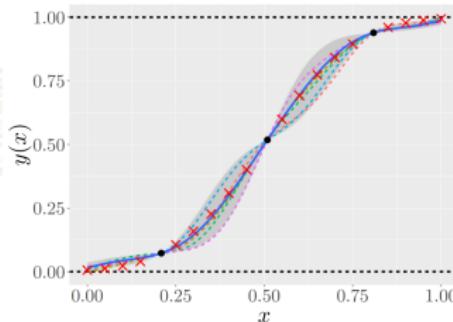
**Exercise.** Considering  $Y_0 \sim \mathcal{GP}(0, k_{y_0, y_0})$  and  $\delta \sim \mathcal{GP}(0, k_{\delta, \delta})$ , compute the covariance function of  $Y_1$ .

- A. F. López-Lopera, S. John and N. Durrande:

Data fusion with multifidelity Gaussian processes for aerodynamic experimental and numerical databases. Work in progress

# Applications of Gaussian processes

## Risk assessment: nuclear safety



We considered the piecewise-linear approximation given by

$$Y_m(x) = \sum_{j=1}^m \phi_j(x) Y(t_j), \text{ s.t. } \begin{cases} Y_m(x_i) = y_i, \text{ for } i = 1, \dots, n & \text{(interpolation constraints)} \\ Y_m \in \mathcal{E} & \text{(inequality constraints)} \end{cases}$$

with  $[Y(t_1), \dots, Y(t_m)] \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$ , and  $\phi_1, \dots, \phi_m$  hat basis functions

- A. F. López-Lopera, N. Durrande, F. Bachoc and O. Roustant:

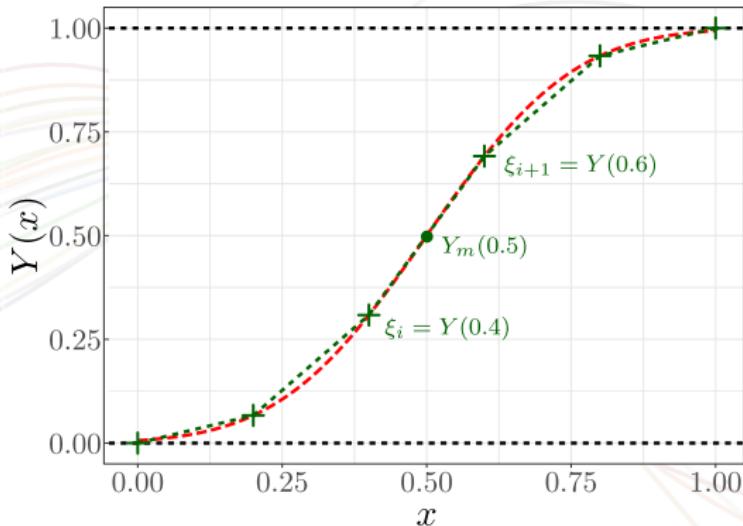
Finite-dimensional Gaussian approximation with linear inequality constraints

SIAM/ASA Journal on Uncertainty Quantification, 2018

ONERA

THE FRENCH AEROSPACE LAB

# Applications of Gaussian processes



- smooth function
- finite approximation

Observe that:

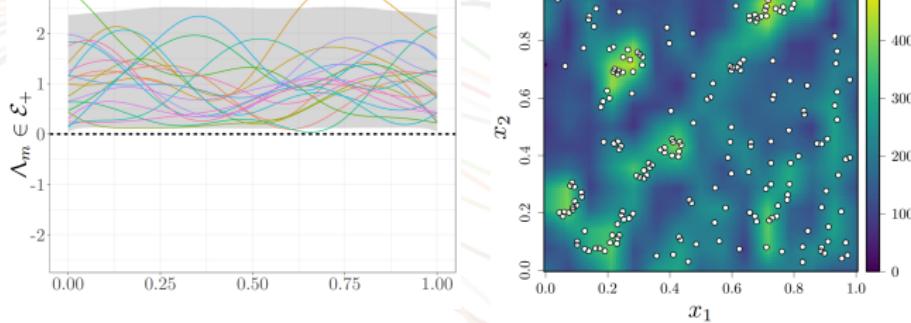
- If  $\alpha_i, \alpha_{i+1} \in [0, 1]$ , then  
 $Y_m(0.5) \in [0, 1]$
- If  $\alpha_i < \alpha_{i+1}$ , then  
 $\alpha_i < Y_m(0.5) < \alpha_{i+1}$

**Advantage :** It is enough imposing the inequality constraints over the knots

- Assuming that  $\xi \sim \mathcal{N}(\mathbf{0}, \Gamma)$ , such that  $\mathbf{l} \leq \Lambda \xi \leq \mathbf{u}$ , then we have

$$\xi \sim \mathcal{T}\mathcal{N}(\mathbf{0}, \Lambda \Gamma \Lambda^\top, \mathbf{l}, \mathbf{u})$$

## Geostatistics: spatial distribution of tree species



We considered a Poisson process with stochastic intensity function  
 $\Lambda \sim \mathcal{GP}(0, k)$  subject to  $\Lambda$  being positive

- A. F. López-Lopera, S. John and N. Durrande:

Gaussian process modulated Cox processes under linear inequality constraints  
International Conference on Artificial Intelligence and Statistics (AISTATS), 2019

## Conclusions

---

# Conclusions

- GPs provide a well-founded non-parametric (Bayesian) framework
- They have been successfully applied in diverse applications:
  - Geostatistics, physics, chemistry
  - Neuroscience, biology and medicine
  - Engineering fields
  - Econometrics
  - ...
- Regularity assumptions are encoded in kernel functions
  - stationarity, isotropy, periodicity, smoothness...