# Gaussian Process Regression

MATH-414: Stochastic Simulation

Francesco Pettenon, Francesca Venturi

École Polytechnique Fédérale de Lausanne

December 8, 2024

Professor: Fabio Nobile

# Outline


1 Introduction
- Context and Motivation

2 Theoretical Background
- Covariance Kernels

3 Recovering a simple function
- Gaussian Process Regression
- Results
- Optimize hyperparameters

4 GP Regression on a permeability field
- Dataset and computational considerations
- Optimize hyperparameters
- 2D Gaussian Process Regression
- Generate Processes: Circulant Embedding
- Monte Carlo Estimator
- Variance Reduction

5 Main References

6 Discussion

# Introduction

# Context and Motivation

- Context:
  - Permeability fields in geoscience
  - Limited data collected through boreholes
- Motivation:
  - Uncertainty quantification
  - Small datasets
  - Prior and posterior distributions

# Theoretical Background

# Theoretical Background

Gaussian Process Regression (GPR)

- $Z, Y$ finite set of positions $\rightarrow$ $\mathbf{f}(Y)$ to predict
- Gaussian random field: $\mathbf{f}(Z) \sim \mathcal{N}(\mathbf{m}(Z), \mathbf{K}(Z))$
- Noisy observations $\tilde{\mathbf{f}}(Z) = \mathbf{f}(Z) + \epsilon$, $\epsilon \sim \mathcal{N}(\mathbf{0}, s^2\mathbf{I}_{N_Z})$
  covariance of $\tilde{\mathbf{f}}(Z)$ is $\mathbf{K}(Z) + s^2\mathbf{I}$

Conditional distribution, mean and covariance functions:

$$\mathbf{f}(Y)|\tilde{\mathbf{f}}(Z) \sim \mathcal{N}(\tilde{\mathbf{m}}(Y \mid Z), \mathbf{K}(Y|Z))$$

$$\tilde{\mathbf{m}}(Y \mid Z) = \mathbf{m}(Y) + \mathbf{K}(Y, Z)\left(\mathbf{K}(Z) + s^2\mathbf{I}\right)^{-1}(\tilde{\mathbf{f}}(Z) - \mathbf{m}(Z))$$

$$\mathbf{K}(Y \mid Z) = \mathbf{K}(Y) - \mathbf{K}(Y, Z)\left(\mathbf{K}(Z) + s^2\mathbf{I}\right)^{-1}\mathbf{K}(Z, Y)$$

# Covariance Kernels

1. The Exponential (EXP) kernel:

$$K_{exp}\left(\mathbf{x}, \mathbf{x}'\right) = \sigma^2 \exp\left\{-\frac{\|\mathbf{x} - \mathbf{x}'\|_2}{\ell}\right\}$$

   - Linear exponent $\rightarrow$ rougher function
   - Isotropic kernel

2. The Squared Exponential (SE) kernel:

$$K_{se}\left(\mathbf{x}, \mathbf{x}'\right) = \sigma^2 \exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\frac{\left(x_i - x_i'\right)^2}{\ell_i^2}\right\}$$

   - Squared exponent $\rightarrow$ smoother function
   - Anisotropic kernel

# Visual Comparison between kernels

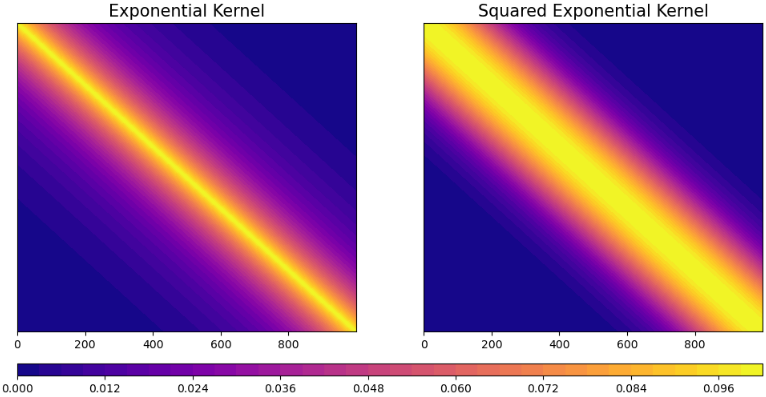


Figure: Visual comparison between the EXP and the SE kernels 1D with fixed parameters $\sigma^2$ and $\ell$.

# Hyperparameters: $\ell$ and $\sigma^2$

- **The Correlation Length** $\ell$: degree of smoothness in the function being modeled. The larger $\ell$, the slower the drop off, leading to smoother shape.
- **The Vertical Scale** $\sigma^2$: variance, deviation from the average prediction at any given point. The higher $\sigma^2$, the greater the flexibility, leading to more fluctuations.

# Recovering a simple function

# Recovering a simple function

- Scalar function to recover: $f(x) = sin(x)$
- $N_Z$ noise-free observations, i.e. $y = sin(x)$.

**Goal**: generate $N_Y$ Gaussian random variables whose mean approximates $f(x) = sin(x)$, to predict the entire Gaussian random process.

# 1D Gaussian Process Regression

**Algorithm** 1D Gaussian Process Regression

1: Generate $N_Z$ *Points* $Z$ uniformly distributed in $[0, 2\pi]$
2: Generate 1000 *Points* $Y$ in a 1D uniform grid of $[0, 2\pi]$
3: Given $f(x) = sin(x)$, evaluate the Points Z, obtaining $\mathbf{f}(Z)$
4: Build the covariance matrices $\mathbf{K}(Y)$, $\mathbf{K}(Z)$, $\mathbf{K}(Y, Z)$ for both kernels and for a given set of parameters $\boldsymbol{\theta} = (\ell, \sigma^2, s^2)$
5: Compute the prediction $\mathbf{m}(Y|Z)$ and $\mathbf{K}(Y|Z)$
6: Draw 3 independent realizations of the Gaussian process $\mathbf{f}(Y)|\mathbf{f}(Z)$

**Next**: analysis of the newly generated Gaussian processes, changes upon the amount of observations, i.e. $N_z$, and the choice of the kernel.

# Results for the EXP kernel

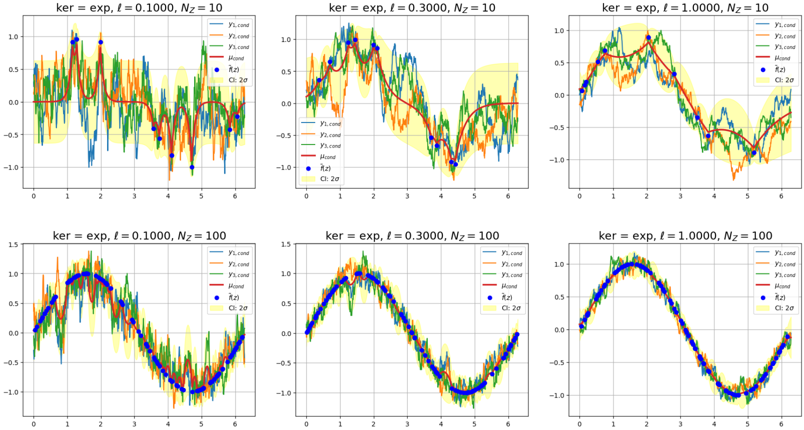


Figure: 1D Gaussian Process: Exponential Kernel

# Results for the SE kernel

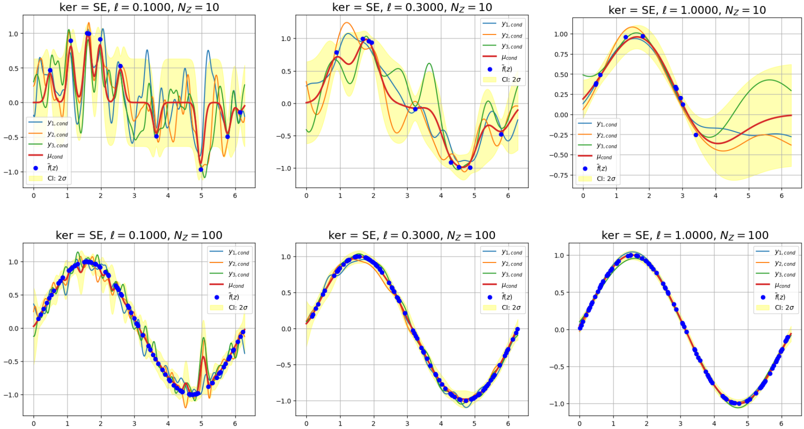


Figure: 1D Gaussian Process: Squared Exponential Kernel

# Optimize hyperparameters

**Optimization criterion**: maximization of the marginal likelihood
**Numerical optimization**: gradient-based optimization algorithm `L-BFGS-B`

| $N_Z = 10$ | $\ell$ | $\sigma^2$ | $s^2$ |
|---|---|---|---|
| EXP | 3.406 | 0.411 | 0.000 |
| SE | 1.964 | 0.444 | 0.000 |

Table: Optimized hyperparameters for $N_Z = 10$

| $N_Z = 100$ | $\ell$ | $\sigma^2$ | $s^2$ |
|---|---|---|---|
| EXP | 12.781 | 0.182 | 0.000 |
| SE | 1.146 | 1.089 | 0.001 |

Table: Optimized hyperparameters for $N_Z = 100$

# Results for optimized hyperparameters

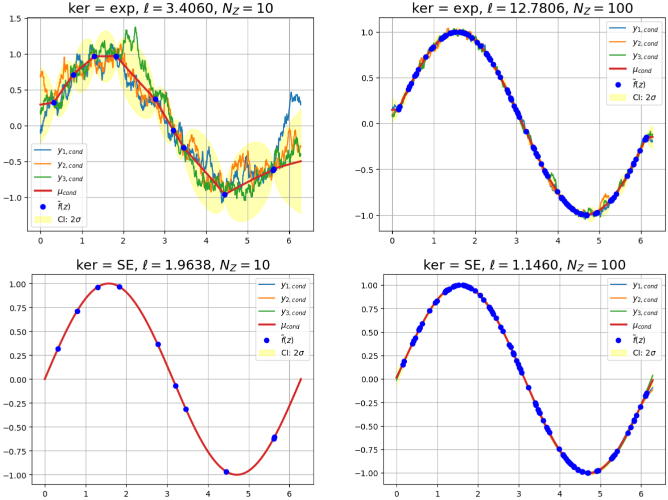


Figure: Predictions and observations with optimized hyperparameters

# Gaussian Process Regression on a Permeability Field

# Dataset and computational considerations

**Dataset**: Permeability field of $60 \times 110$ noisy observations, $\epsilon \sim \mathcal{N}\left(\mathbf{0}, s^2 \mathbf{I}_{N_Z}\right)$
**Goal**: Approximation of the field with a limited training set for model generation. The two uniform rectilinear grids are:

| | | | | | |
|---|---|---|---|---|---|
| $x$ | 10 | 20 | 30 | 40 | 50 |
| $y$ | 15 | 35 | 55 | 75 | 95 |

Table: First dataset $5 \times 5$

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 |
| $y$ | 5 | 15 | 25 | 35 | 45 | 55 | 65 | 75 | 85 | 95 | 105 |

Table: Second dataset $11 \times 11$

# Optimize hyperparameters

Kernel SE has two parameters for the length scale $\ell$ (anistropy).

| Z1 | $\sigma^2$ | $s^2$ | $\ell_1$ | $\ell_2$ |
|---|---|---|---|---|
| EXP | 12.395 | 1e-10 | 64.590 | - |
| SE | 9.047 | 0.589 | 16.330 | 21.891 |

Table: Optimal hyperparameters for Z1: $\theta = \{\sigma^2, s^2, \ell\}$

| Z2 | $\sigma^2$ | $s^2$ | $\ell_1$ | $\ell_2$ |
|---|---|---|---|---|
| EXP | 11.428 | 0.035 | 40.395 | - |
| SE | 7.005 | 1.120 | 11.720 | 18.266 |

Table: Optimal hyperparameters for Z2: $\theta = \{\sigma^2, s^2, \ell\}$

# Predict the Gaussian Process

**Algorithm** 2D Gaussian Process Regression

1. Create 2D-meshgrid with Points $Z$
2. Extract the Points Z in `true_perm`, obtaining $\mathbf{f}(Z)$
3. Create the link between the physical points (dim $= (110 \times 60)$) and the Gaussian Vector $Y$ (dim $= 6600$)
4. Optimize the parameters $\boldsymbol{\theta} = (\boldsymbol{\ell}, \sigma^2, s^2)$ of the kernel w.r.t. marginal likelihood
5. Build the covariance matrices $\mathbf{K}(Y)$, $\mathbf{K}(Z)$, $\mathbf{K}(Y, Z)$
6. Compute the prediction $\tilde{\mathbf{m}}(Y \mid Z)$ and $\mathbf{K}(Y|Z)$
7. **return** $\tilde{\mathbf{m}}(Y \mid Z)$, $\mathbf{K}(Y|Z)$

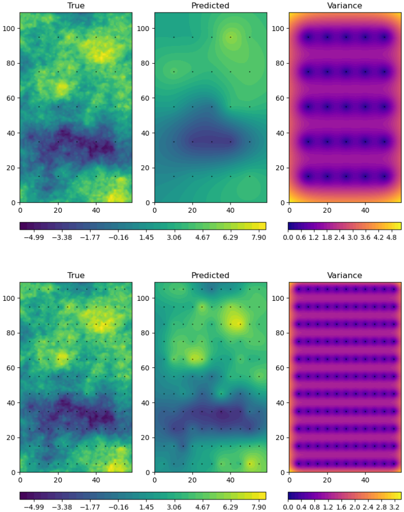


Figure: 2D Gaussian Process: Exponential Kernel

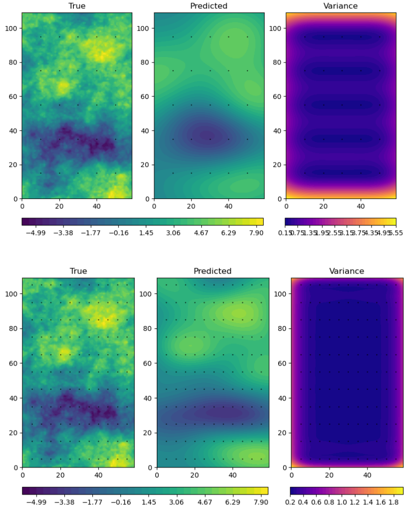


Figure: 2D Gaussian Process: Squared Exponential Kernel

# Generate Processes: Circulant Embedding

Stationary Gaussian Process $\rightarrow$ 2D-circulant matrix $\mathbf{P}$

---

**Algorithm** Circulant Embedding 2D

---

Build $\mathbf{P}$ starting from $\mathbf{K}(Y)$
Calculate $\mathbf{W} = \text{fft}(\mathbf{P})$, the 2-dimensional FFT of $\mathbf{P}$
Check that all elements of $\mathbf{W}$ are positive
Generate matrix $\mathbf{X}$ with size of $\mathbf{W}$ containing i.i.d. normal variables
Calculate $\bar{\mathbf{Z}} = \text{ifft}(\sqrt{\mathbf{W}} \odot \mathbf{X})$, the 2-dimensional inverse FFT of $\sqrt{\mathbf{W}} \odot \mathbf{X}$
Generate $\mathbf{f}(Y) = Re(\bar{\mathbf{Z}})|_Y + Im(\bar{\mathbf{Z}})|_Y$
**return** $\mathbf{f}(Y)$

---

# Generate Processes: Circulant Embedding

Within the 6600 values just generated, there are also the sampling points, whose true value is known. "A posteriori" conditioning can be applied to the observations.

**Algorithm** Generation from conditional Gaussian distribution

1: Generate $\mathbf{f}(Y) \sim \mathcal{N}(\mathbf{0}, \mathbf{K}(Y))$ using 2D Circulant Embedding
2: Extract indexes of points $Z$ and evaluate $\mathbf{f_{new}}(Z) = \mathbf{f}(Y)|_Z$
3: Generate $\mathbf{f}(Y)|\mathbf{f}(Z) = \mathbf{f}(Y) + \mathbf{K}(Y, Z)\mathbf{K}(Z)^{-1}(\mathbf{f_{new}}(Z) - \mathbf{f}(Z))$
4: **return** $\mathbf{f}(Y)|\mathbf{f}(Z)$

Conditional Gaussian processes:

$$\mathbf{f}(Y)|\tilde{\mathbf{f}}(Z) \sim \mathcal{N}(\tilde{\mathbf{m}}(Y|Z), \mathbf{K}(Y|Z))$$

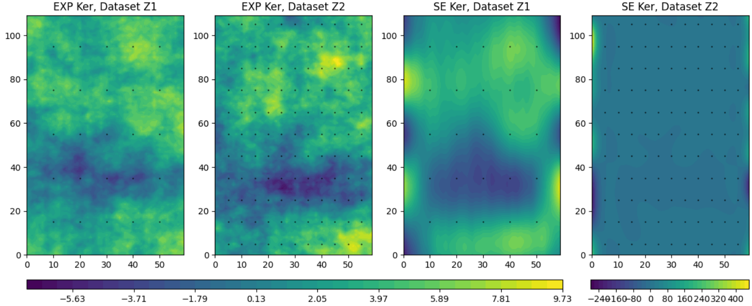


Figure: Gaussian Processed obtained via Circular Embedding technique

# Monte Carlo Estimator

Determine whether the permeability at certain coordinates exceeds a predetermined threshold.
Critical locations: $x = 35, 40, 45, 50$ and $y = 85$.

$$\mathbb{P}(\max_{i=1,\ldots,4} f(x_i, y_i) \geq 8) = \mathbb{E}(\psi(\mathbf{x}, \mathbf{y})),$$

where $\psi(\mathbf{x}, \mathbf{y}) = \mathbf{1}_{\max_{i=1,\ldots,4} f(x_i, y_i) \geq 8}$.

**Theorem**

Let $\mathbf{X}$ be a Gaussian vector of size $n$ with mean $\boldsymbol{\mu}$ and covariance matrix $\Sigma$. Then the linear transformation $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$ will transform $\mathbf{X}$ into a Gaussian vector $\mathbf{Y}$ with mean $A\boldsymbol{\mu} + \mathbf{b}$ and covariance matrix $A\Sigma A^T$.

# Monte Carlo Estimator

| Z2 | $\bar{N}$ | $\hat{\mu}_{\bar{N}}$ | $\hat{\sigma}_{\bar{N}}$ | tol |
|---|---|---|---|---|
| EXP | 9360182 | 0.024976 | 0.15605 | 1e-4 |
| SE | 2633285 | 6.7596e-5 | 0.00822 | 1e-5 |

Table: Estimation with Two stages MC for datasets Z1

| Z2 | $\bar{N}$ | $\hat{\mu}_{\bar{N}}$ | $\hat{\sigma}_{\bar{N}}$ | tol |
|---|---|---|---|---|
| EXP | 26223987 | 0.073693 | 0.26127 | 1e-4 |
| SE | 768284 | 2.6032e-5 | 0.00161 | 1e-5 |

Table: Estimation with Two stages MC for datasets Z2

# Variance Reduction

Reduce the variance of MC $\rightarrow$ accelerate convergence $O(\frac{1}{\sqrt{N}})$.

**Theorem**

Assume that the random variable $Z$ has the expression $Z = \psi(X)$, with $X = (X_1, \dots, X_d)$ a random vector with independent components, such that

- $X$ has a symmetric distribution around its mean, i.e. $2\mathbb{E}[X] - X \sim X$
- $\psi$ is a monotone function in each of its arguments.

Then $Z = \psi(X)$ and $Z_a = \psi(2\mathbb{E}[X] - X)$ satisfy $\mathbb{E}[Z] = \mathbb{E}[Z_a]$ and $\mathrm{Cov}(Z, Z_a) < 0$.

# Variance Reduction

Hypothesys on $\psi$ hold: variance reduction with Antithetic Variables.

| Z1 | $\hat{\mu}_{CMC}$ | $\hat{\mu}_{AV}$ |
|---|---|---|
| EXP | 0.0249 $\pm$ 9.662e-5 | 0.0250 $\pm$ 9.560e-5 |
| SE | 7.4e-5 $\pm$ 5.331e-6 | 7.23e-5 $\pm$ 5.269e-6 |

Table: Confidence interval of CMC and AV for datasets Z1

| Z2 | $\hat{\mu}_{CMC}$ | $\hat{\mu}_{AV}$ |
|---|---|---|
| EXP | 0.0737 $\pm$ 1.6193e-4 | 0.0737 $\pm$ 1.5539e-4 |
| SE | 3.7e-6 $\pm$ 1.1922e-6 | 3.7e-6 $\pm$ 1.1922e-6 |

Table: Confidence interval of CMC and AV for datasets Z2

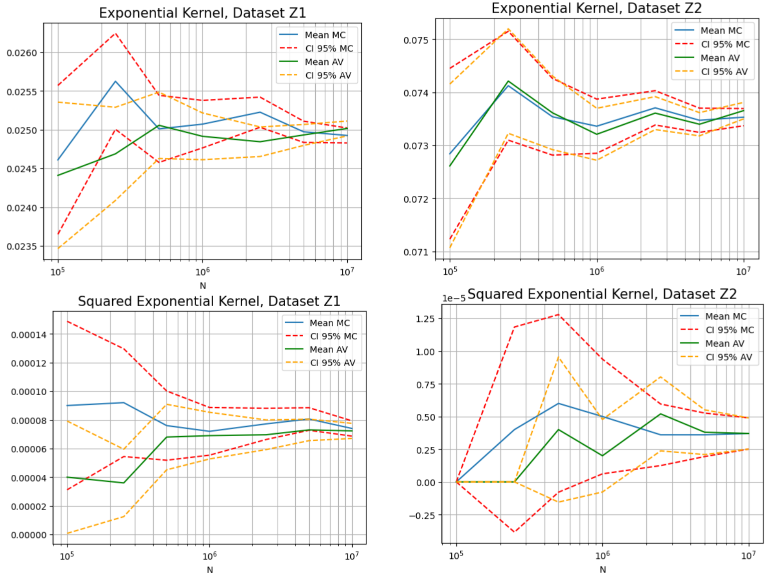


Figure: Monte Carlo estimations and Antithetic Variables to reduce variance

# References

Nobile, Fabio. *Stochastic Simulation: Lecture Notes*. 2022.

Van Barel, Andreas. “Multilevel Monte Carlo Methods for Robust Optimization of Partial Differential Equations”. In: (2021).

Williams, Christopher KI and Carl Edward Rasmussen. *Gaussian processes for machine learning*. Vol. 2. 3. MIT press Cambridge, MA, 2006.

# Questions?