# Introduction to Gaussian process metamodel - Kriging

May 2020



# Outline

Random process

Gaussian process metamodel

Conclusions

# Outline

# Random variable and random vector

**Random variable**: variable whose values depend on outcome of a random phenomenon

A random variable $X$ is a function from a set of possible outcomes $\Omega$ to a measurable space $E$:

$$X : \Omega \to E$$

$\Omega$ being a sample space of the probability triple $(\Omega, \mathcal{F}, \mathcal{P})$ in which:

- $\mathcal{F}$: set of events, each event contains zero or more outcomes
- $\mathcal{P}$: probability measure, assignment of probability to events

Example: rolling a fair dice, outcome $\omega$, set of possible outcomes: six faces $\Omega = \{1, \ldots, 6\}$. Random variable $X$: $X = 1$ if $\omega \in \{1, 2\}$, $X = 2$ if $\omega \in \{3, 4\}$, $X = 3$ if $\omega \in \{5, 6\}$. Probabilities assigned to its values $\mathbb{P}[X = 1] = \frac{1}{3}$

**Random vector**: a vector of random variables

$$\boldsymbol{X} = (X_1, \ldots, X_n)$$

# Random process

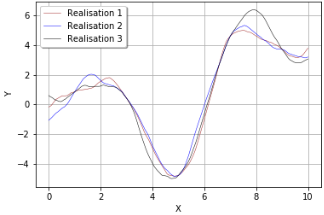
**Random process** Y: set of random variables indexed by $x$ and defined in the probability space $(\Omega, \mathcal{F}, \mathcal{P})$

$$Y : \Omega \, x \, \mathcal{D} \rightarrow E$$

$\mathcal{D} \subset \mathbb{R}^d$: space of indices (e.g. spatial, temporal domains)

- At a given point $x_0 \in \mathcal{D}$, $Y(\omega, x_0)$ is a random variable.
- With a given random event $\omega_0 \in \Omega$ and index $x \in \mathcal{D}$, one obtains a function (a.k.a realization, trajectory):

$$y(\omega_0, x) : x \in \mathcal{D} \rightarrow \mathbb{R}$$



# Random process

**Mean**:

$$\mu_x = \mathbb{E}\left[Y(x)\right]$$

**Covariance**:

$$C(x, x') := C(Y(x), Y(x')) = \mathbb{E}\left[(Y(x) - m_x)(Y(x') - m_{x'})\right]$$

**Stationary random process**: the covariance function $C(x, x')$ depends only on $\tau = x - x'$, not on the position in the space

$$C(x, x') = C(x - x') = C(\tau)$$

***Gaussian process***: the random process $Y : \Omega \, x \, \mathcal{D} \rightarrow E$ is called a gaussian process if every finite collection of random variables is a Gaussian random vector (i.e. has a multi-variate normal distribution)

$$\forall k, \forall \{x_1, \ldots, x_k\} \in \mathcal{D}^k, \{Y(x_1), \ldots, Y(x_k)\} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{C}); \; \boldsymbol{C}_{ij} = C(x_i, x_j)$$

# Covariance function of a stationary random process

Global form of a unidimensional covariance function (Schlather 2009):

$$C(x, x') = \delta_0 + \sigma^2 \, \rho \left( \frac{|x - x'|}{\theta} \right)$$

- $\delta_0$: nugget effect
- $\sigma^2$: constant variance of the random process
- $\theta$: correlation length

Examples of covariance functions:

| Kernel | Function |
|---|---|
| Matérn | $C_\nu(\tau) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{\sqrt{2\nu}\|\tau\|}{\theta} \right)^\nu K_\nu \left( \frac{\sqrt{2\nu}\|\tau\|}{\theta} \right)$ |
| Generalized exponential | $C(\tau) = \sigma^2 \exp\left( -\frac{\|\tau\|^\gamma}{\theta^\gamma} \right)$ |
| Squared exponential | $C(\tau) = \sigma^2 \exp\left( -\frac{1}{2} \frac{\|\tau\|^2}{\theta^2} \right)$ |

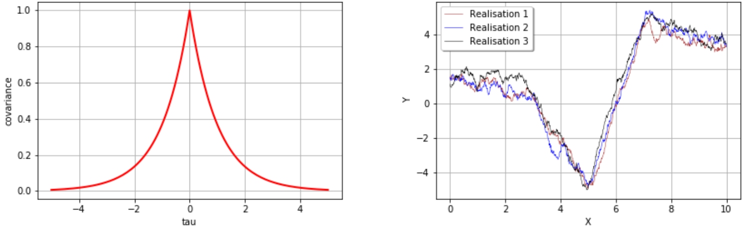
# Covariance function

The regularity of the process is determined by the differentiability of $C(\tau)$ at $\tau = 0$.

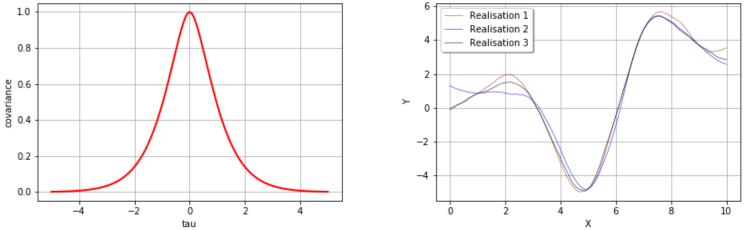
For stationary processes, the trajectories $y(x)$ are $p$-times differentiable if $C(\tau)$ is $2p$ times differentiable at $\tau = 0$.

| $\nu$ | Matérn covariance function |
|---|---|
| $\nu = 1/2$ | $C_{1/2}(\tau) = \sigma^2 exp(-\frac{\|\tau\|}{\theta})$ |
| $\nu = 3/2$ | $C_{3/2}(\tau) = \sigma^2 (1 + \frac{\sqrt{3}\|\tau\|}{\theta}) exp(-\frac{\sqrt{3}\|\tau\|}{\theta})$ |
| $\nu = 5/2$ | $C_{5/2}(\tau) = \sigma^2 (1 + \frac{\sqrt{5}\|\tau\|}{\theta} + \frac{5\|\tau\|^2}{3\theta^2}) exp(-\frac{\sqrt{5}\|\tau\|}{\theta})$ |

# Covariance function

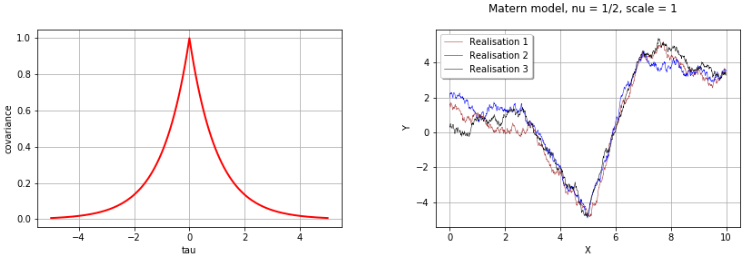



$$\nu = 1/2$$

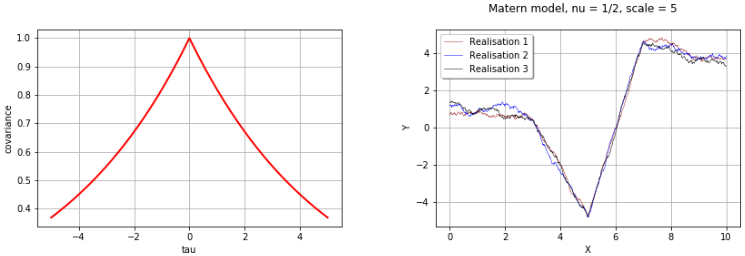



$$\nu = 3/2$$

# Covariance function

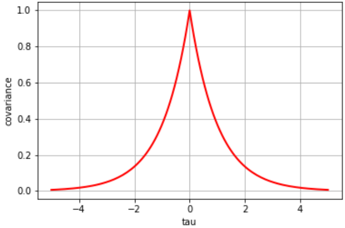


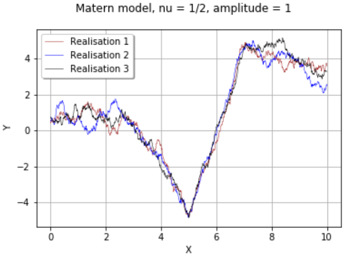

$\rho = 1$



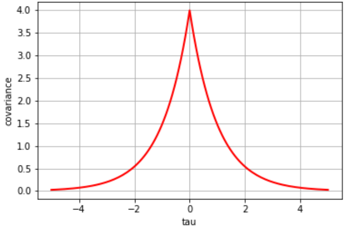

$\rho = 5$

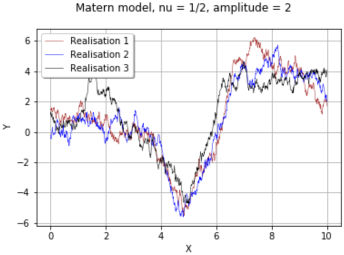
# Covariance function




$\sigma = 1$




$\sigma = 2$

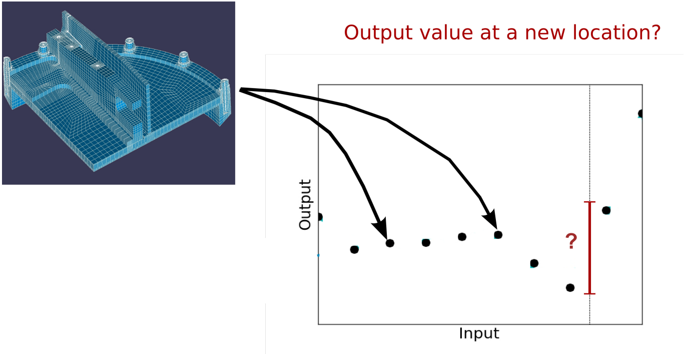
# Outline

Random process

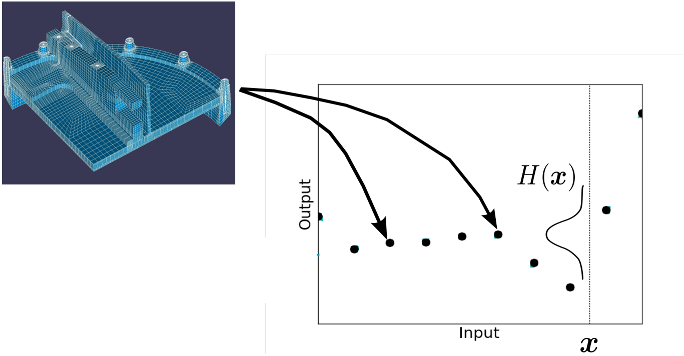
Gaussian process metamodel

Conclusions

# Prediction at a new point



# Prediction at a new point




**Assumption:** The response is a realization of a Gaussian random variable whose moments depend on the design points

# Gaussian process assumption

The model output is a realization of a Gaussian random process of the form :

$$Y(\boldsymbol{x},\omega) = \boxed{\boldsymbol{r}(\boldsymbol{x})\cdot\boldsymbol{\beta}} + \boxed{Z(\boldsymbol{x},\omega)}$$

Trend (deterministic)

Linear regression
on a fixed basis

Random fluctuations

Gaussian process
with zero mean and
stationary

$$\mathbb{C}\mathrm{ov}_Z(\boldsymbol{x},\boldsymbol{x}') = \sigma^2\rho(\|\boldsymbol{x}-\boldsymbol{x}'\|)$$

Kriging

# Conditional mean and variance

Experimental design: $\mathcal{X} = \left\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}\right\}$
$\mathcal{Y} = \left\{Y(\boldsymbol{x}^{(1)}), \dots, Y(\boldsymbol{x}^{(N)})\right\}$

Notations: $\boldsymbol{k}(\boldsymbol{x}^*) \equiv \left\{\rho\left(\boldsymbol{x}^*, \boldsymbol{x}^{(1)}\right), \dots, \rho\left(\boldsymbol{x}^*, \boldsymbol{x}^{(N)}\right)\right\}^{\mathsf{T}}$

$$\boldsymbol{R} \equiv \left(r_j(\boldsymbol{x}^{(i)})\right)_{1\leq i\leq N, 1\leq j\leq p} \quad , \quad \boldsymbol{K} \equiv \left(\rho(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})\right)_{1\leq i,j\leq N}$$

Conditional mean: $\boxed{\mu(\boldsymbol{x}^*) = \boldsymbol{r}^{\mathsf{T}}(\boldsymbol{x}^*)\boldsymbol{\beta} + \boldsymbol{k}^{\mathsf{T}}(\boldsymbol{x}^*)\boldsymbol{K}^{-1}(\mathcal{Y} - \boldsymbol{R}\boldsymbol{\beta})}$

Conditional variance: $\boxed{\sigma^2(\boldsymbol{x}^*) = \sigma^2 - \boldsymbol{k}^{\mathsf{T}}(\boldsymbol{x}^*)\, \boldsymbol{K}^{-1}\, \boldsymbol{k}^{\mathsf{T}}(\boldsymbol{x}^*) - \boldsymbol{U}(x^*)^{\mathsf{T}}\, \boldsymbol{F}\, \boldsymbol{U}(x^*)}$

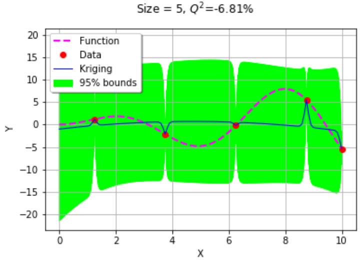
with $\boldsymbol{F} = (\boldsymbol{R}^{\mathsf{T}}\boldsymbol{K}^{-1}\boldsymbol{R})^{-1}$, $\boldsymbol{U}(x^*) = \boldsymbol{R}^t K^{-1}\boldsymbol{k}(x^*) - \boldsymbol{r}(x^*)^t$

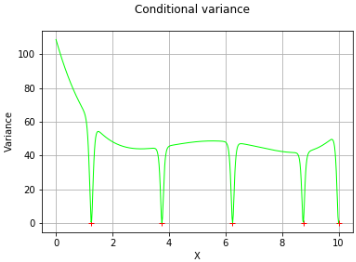
# Conditional mean and variance

Consider an instructive model: $y = f(x) = x\sin(x)$

Gaussian process metamodel:

$$H(x,\omega) = \boldsymbol{r}(x)\cdot\boldsymbol{\beta} + Z(x,\omega) \quad , \quad \mathbb{C}\mathrm{ov}_Z(x,x') = \sigma^2 e^{-\theta(x-x')^2}$$



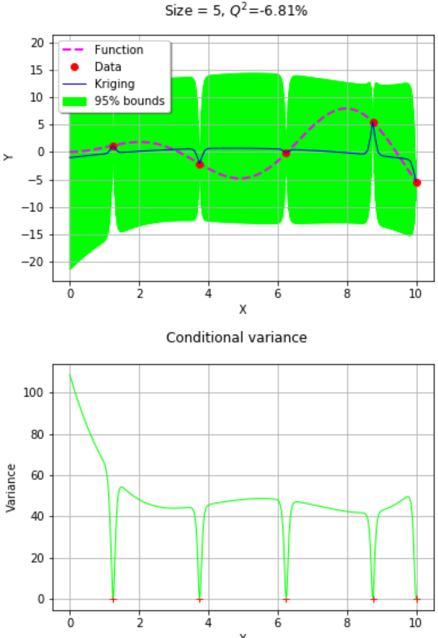

- The conditional mean is used as a metamodel (interpolator)
- The conditional variance is used as an error indicator

# Parameter fitting

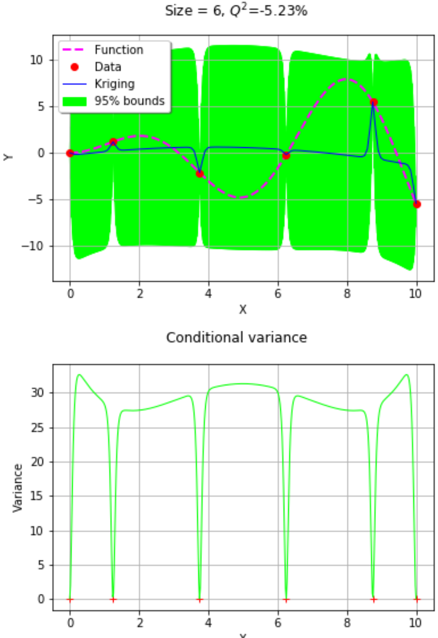
To apply the previous formulas, the parameters $(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta})$ have to be estimated from the design points

- Optimal correlation parameter $\hat{\boldsymbol{\theta}}$ estimated by the maximum likelihood estimate (Marrel et al. 2008) or cross validation (Bachoc 2013)
- Parameters $(\hat{\boldsymbol{\beta}}, \hat{\sigma})$ estimated by empirical best linear unbiased estimator (BLUE) (Santner et al. 2003)

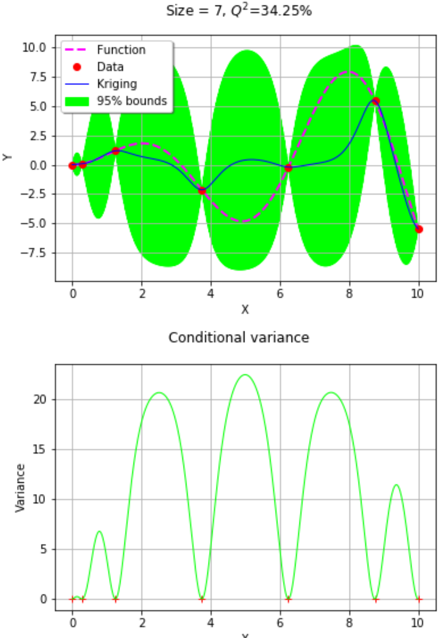
# Sequential enrichment of the experimental design



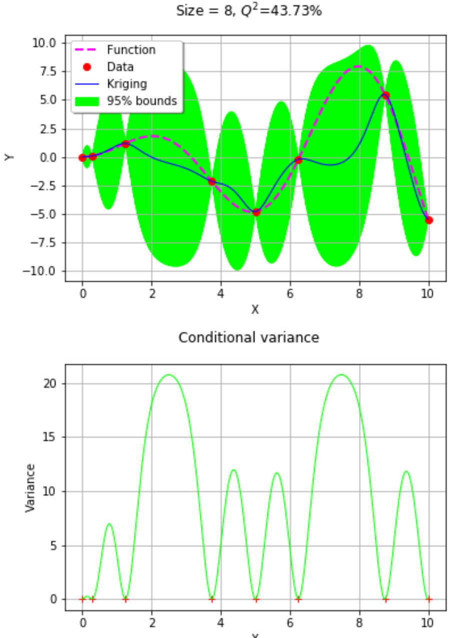
# Sequential enrichment of the experimental design



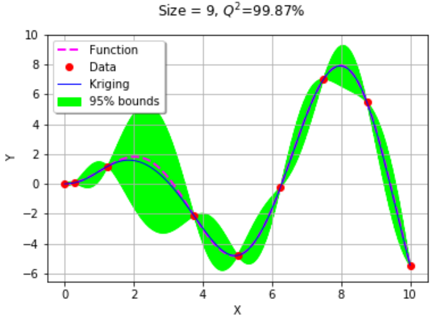
# Sequential enrichment of the experimental design

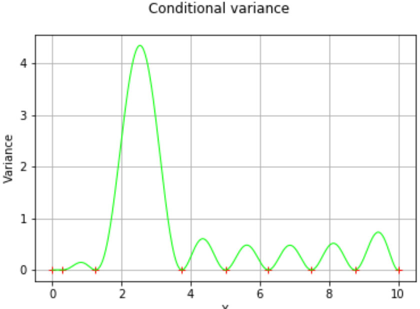


# Sequential enrichment of the experimental design

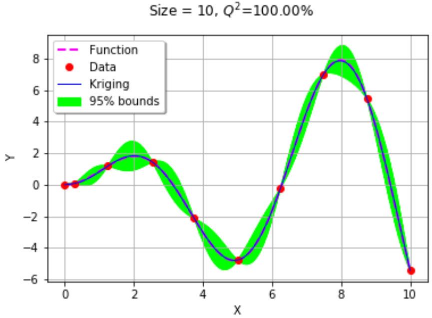


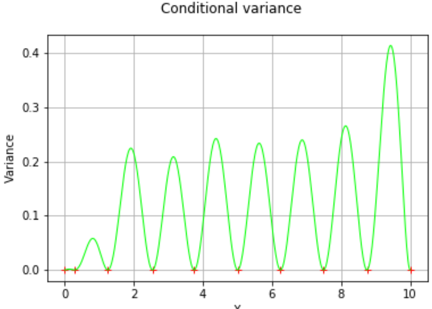
# Sequential enrichment of the experimental design



# Sequential enrichment of the experimental design



# Outline

# Gaussian process metamodel

- The regularity of the trajectories depends on the choice of covariance function
- Kriging allows to associate a measure of certainty to a prediction of the function
- Kriging allows the effective sequential enrichment of the experimental design

Thank you