Introduction to Gaussian-process based Kriging models for metamodeling and validation of computer codes

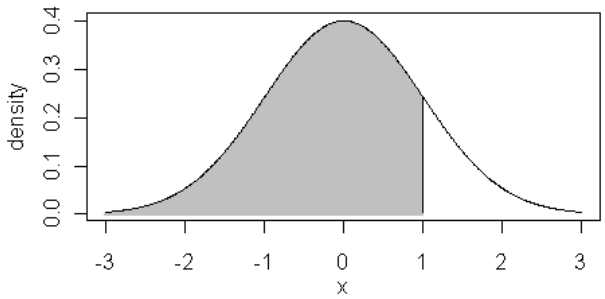
François Bachoc

Department of Statistics and Operations Research, University of Vienna
(Former PhD student at CEA-Saclay, DEN, DM2S, STMF, LGLS, F-91191 Gif-Sur-Yvette, France
and LPMA, Université Paris 7)

LRC MANON - INSTN Saclay - March 2014

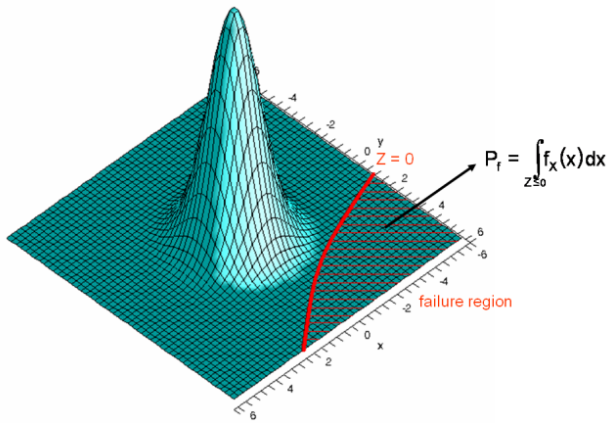
- A random Variable X is a random number, defined by a probability density function $f_X : \mathbb{R} \to \mathbb{R}^+$ for which, for $a, b \in \mathbb{R}$:

  "probability of $a \leq X \leq b$" $= \int_a^b f_X(x)dx$



- Similarly a random Vector $V = (V_1, ..., V_n)^t$ is a vector of random variables. It is also defined by a probability density function $f_V : \mathbb{R}^n \to \mathbb{R}^+$ for which, for $E \in \mathbb{R}^n$:

  "probability of $V \in E$" $= \int_E f_V(v)dv$



Remark

Naturally we have $\int_{-\infty}^{+\infty} f_X(x)dx = \int_{\mathbb{R}^n} f_V(v)dv = 1$

Mean, variance and covariance

- The mean of a random variable X with density f_X is denoted $\mathbb{E}(X)$ and is

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

- Let X be a random variable. The variance of X is denoted $var(X)$ and is

$$var(X) = \mathbb{E}\left\{(X - \mathbb{E}(X))^2\right\}$$

  - $var(X)$ is large $\rightarrow$ X can be far from its mean $\rightarrow$ more uncertainty.
  - $var(X)$ is small $\rightarrow$ X is close to its mean $\rightarrow$ less uncertainty.

- Let X, Y be two random variables. The covariance between X and Y is denoted $cov(X, Y)$ and is

$$cov(X, Y) = \mathbb{E}\left\{(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right\}$$

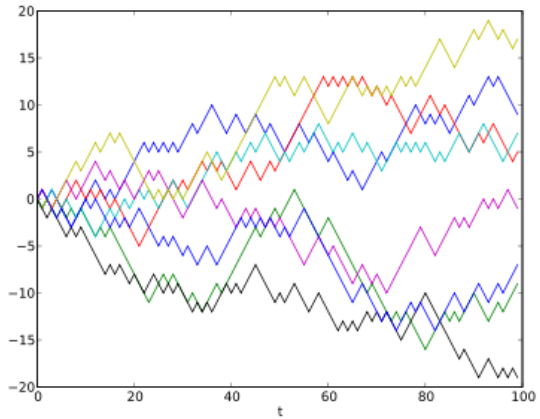
  - $|cov(X, Y)| \approx \sqrt{var(X)var(Y)} \rightarrow$ X and Y are almost proportional to one another.
  - $|cov(X, Y)| << \sqrt{var(X)var(Y)} \rightarrow$ X and Y are almost independent (when they are Gaussian).

Mean vector and covariance matrix

- Let $V = (V_1, ..., V_n)^t$ be a random vector. The mean vector of V is denoted $\mathbb{E}(V)$ and is the $n \times 1$ vector defined by
$$(\mathbb{E}(V))_i = \mathbb{E}(V_i)$$
- Let $V = (V_1, ..., V_n)^t$ be a random vector. The covariance matrix of V is denoted $cov(V)$ and is the $n \times n$ matrix defined by
$$(cov(V))_{i,j} = cov(V_i, V_j)$$
  - The diagonal terms show which components are the most uncertain.
  - The non-diagonal terms show the dependence between the components.

Stochastic processes

A stochastic process is a function $Z : \mathbb{R}^n \to \mathbb{R}$ such that $Z(x)$ is a random variable. Alternatively a stochastic process is a function on $\mathbb{R}^n$ that is unknown, or that depends of underlying random phenomena.



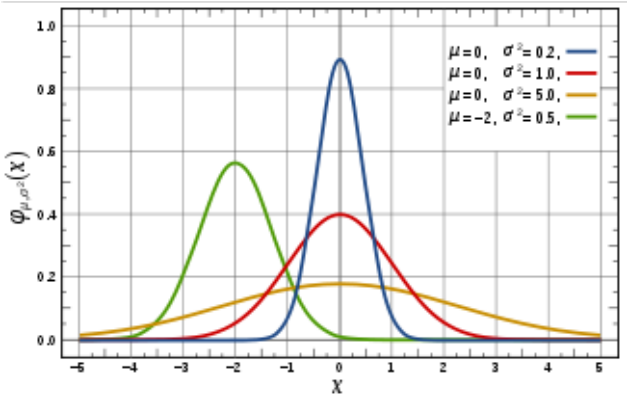
We explicit the randomness of $Z(x)$ by writing it $Z(\omega, x)$ with ω in a probability space Ω. For a given ω_0, we call the function $x \to Z(\omega_0, x)$ a realization of the stochastic process Z.

Mean function $M : x \to M(x) = \mathbb{E}(Z(x))$
Covariance function $C : (x_1, x_2) \to C(x_1, x_2) = cov(Z(x_1), Z(x_2))$

Gaussian variables and vectors

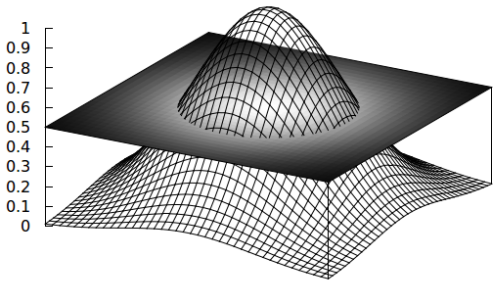
A random variable X is a Gaussian variable with mean μ and variance $\sigma^2 > 0$ when its probability density function is

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$



A n-dimensional random vector V is a Gaussian vector with mean vector m and invertible covariance matrix R when its multidimensional probability density function is

$$f_{m,R}(v) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det(R)}} \exp\left(-\frac{1}{2}(v-m)^t R^{-1}(v-m)\right)$$



E.g. for Gaussian variables : μ and σ^2 are both parameters of the probability density function and the mean and variances of it. That is $\int_{-\infty}^{+\infty} x f_{\mu,\sigma^2}(x)dx = \mu$ and $\int_{-\infty}^{+\infty}(x-\mu)^2 f_{\mu,\sigma^2}(x)dx = \sigma^2$

Gaussian processes

A stochastic process Z on $\mathbb{R}^d$ is a Gaussian process when for all $x_1, ..., x_n$, the random vector $(Z(x_1), ..., Z(x_n))$ is Gaussian.

- A Gaussian process is characterized by its mean and covariance functions.

Why are Gaussian processes convenient ?

- Gaussian distribution is reasonable for modeling a large variety of random variables
- Gaussian processes are simple to define and simulate
- They are characterized by their mean and covariance functions
- As we will see, Gaussian properties simplify the resolution of problems
- Gaussian processes have been the most studied theoretically

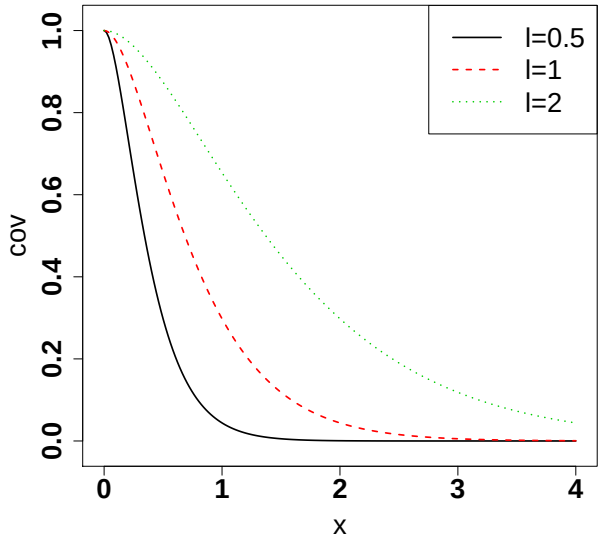
Example of the Matérn $\frac{3}{2}$ covariance function on $\mathbb{R}$

The Matérn $\frac{3}{2}$ covariance function, for a Gaussian process on $\mathbb{R}$ is parameterized by

- A variance parameter $\sigma^2 > 0$
- A correlation length parameter $\ell > 0$

It is defined as

$$C(x_1, x_2) = \left(1 + \sqrt{6}\frac{|x_1 - x_2|}{\ell}\right) e^{-\sqrt{6}\frac{|x_1 - x_2|}{\ell}}$$

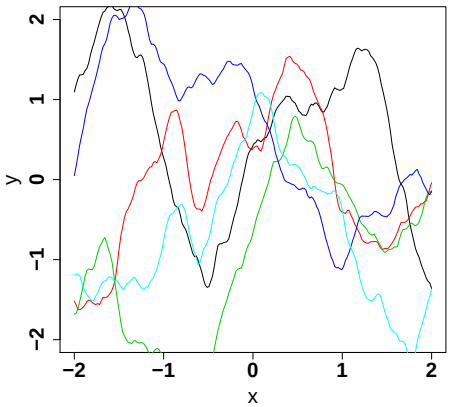


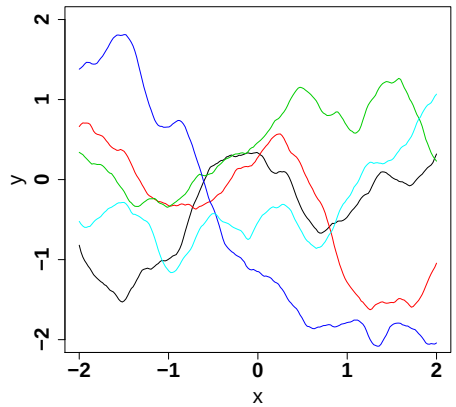
Interpretation

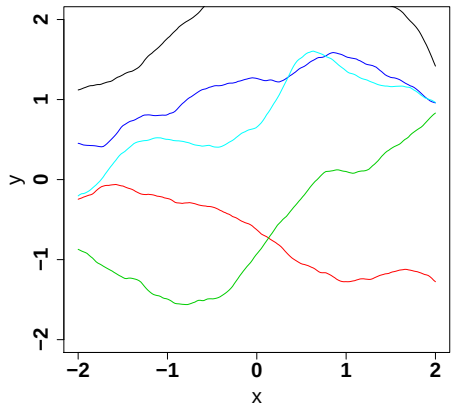
- The Matérn $\frac{3}{2}$ function is stationary : $C(x_1 + h, x_2 + h) = c(x_1, x_2) \Rightarrow$ The behavior of the corresponding Gaussian process is invariant by translation.
- σ^2 corresponds to the order of magnitude of the functions that are realizations of the Gaussian process
- ℓ corresponds to the speed of variation of the functions that are realizations of the Gaussian process

The Matérn $\frac{3}{2}$ convariance function on $\mathbb{R}$: illustration of ℓ

Plot of realizations of a Gaussian process having the Matérn $\frac{3}{2}$ covariance function for $\sigma^2 = 1$ and $\ell = 0,5, 1, 2$ from left to right







The Matérn $\frac{3}{2}$ convariance function : generalization to $\mathbb{R}^d$

We now consider a Gaussian process on $\mathbb{R}^d$.
The corresponding multidimensional Matérn $\frac{3}{2}$ convariance function is parameterized by

- A variance parameter $\sigma^2 > 0$
- d correlation length parameters $\ell_1 > 0, ..., \ell_d > 0$

It is defined as

$$C(x,y) = \left(1 + \sqrt{6}||x-y||_{\ell_1,...,\ell_d}\right) e^{-\sqrt{6}||x-y||_{\ell_1,...,\ell_d}}$$

with

$$||x-y||_{\ell_1,...,\ell_d} = \sqrt{\sum_{i=1}^{d} \frac{(x_i - y_i)^2}{\ell_i^2}}$$

Interpretation

- Still stationary
- σ^2 still drives the order of magnitudes of the realizations
- $\ell_1, ..., \ell_d$ correspond to the speed of variation of the realizations $x \to Z(\omega, x)$ when only the corresponding variable $x_1, ..., x_d$ varies.
- $\Rightarrow$ when ℓ_i is particularly small, then the variable x_i is particularly important $\Rightarrow$ hierarchy of the input variables $x_1, ...x_d$ according to their correlation lengths $\ell_1, ..., \ell_d$

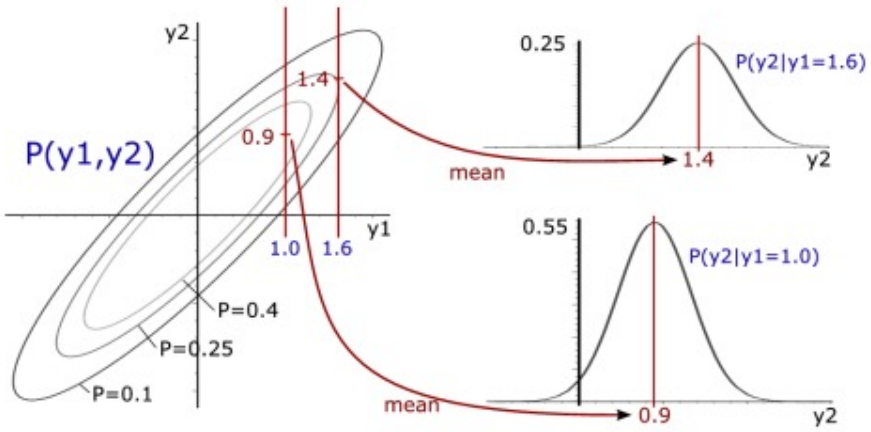
Summary

- A Gaussian process can be seen as a random phenomenon yielding realizations, i.e. specific functions $\mathbb{R}^d \to \mathbb{R}$
- The standard probability tools enable to model and quantify the uncertainty we have on these realizations
- The choice of the covariance function (e.g. Matérn $\frac{3}{2}$) enables to synthesize the information we have (get) on the nature of the realizations with a small number of parameters

Conditional probability density function

Consider a partitioned random vector $(Y_1, Y_2)^t$ of size $(n_1 + 1) \times 1$, with probability density function $f_{Y_1,Y_2} : \mathbb{R}^{n_1+1} \to \mathbb{R}^+$. Then, Y_1 has the probability density function
$f_{Y_1}(y_1) = \int_{\mathbb{R}} f_{Y_1,Y_2}(y_1, y_2) dy_2$.
The conditional probability density function of Y_2 given $Y_1 = y_1$ is then

$$f_{Y_2|Y_1=y_1}(y_2) = \frac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_1}(y_1)}$$



Interpretation

It is the continuous generalization of the Bayes formula

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

Conditional mean

Consider a partitioned random vector $(Y_1, Y_2)^t$ of size $(n_1 + 1) \times 1$, with conditional probability density function of Y_2 given $Y_1 = y_1$ given by $f_{Y_2|Y_1=y_1}(y_2)$.
Then the conditional mean of Y_2 given $Y_1 = y_1$ is

$$\mathbb{E}(Y_2|Y_1 = y_1) = \int_{\mathbb{R}} y_2 f_{Y_2|Y_1=y_1}(y_2) dy_2$$

$\mathbb{E}(Y_2|Y_1 = y_1)$ is in fact a function of Y_1. Thus it is also a random variable. We emphasize this by writing $\mathbb{E}(Y_2|Y_1)$. Thus $\mathbb{E}(Y_2|Y_1 = y_1)$ is a realization of $\mathbb{E}(Y_2|Y_1)$.

Optimality

The function $y_1 \to \mathbb{E}(Y_2|Y_1 = y_1)$ is the best prediction of Y_2 we can make, when observing only Y_1. That is, for any function $f : \mathbb{R}^{n_1} \to \mathbb{R}$:

$$\mathbb{E}\left\{(Y_2 - f(Y_1))^2\right\} \geq \mathbb{E}\left\{(Y_2 - \mathbb{E}(Y_2|Y_1))^2\right\}$$

Conditional variance

Consider a partitioned random vector $(Y_1, Y_2)^t$ of size $(n_1 + 1) \times 1$, with conditional probability density function of Y_2 given $Y_1 = y_1$ given by $f_{Y_2|Y_1=y_1}(y_2)$.
Then the conditional variance of Y_2 given $Y_1 = y_1$ is

$$var(Y_2|Y_1 = y_1) = \int_{\mathbb{R}} (y_2 - \mathbb{E}(Y_2|Y_1 = y_1))^2 f_{Y_2|Y_1=y_1}(y_2) dy_2$$

Summary

- The conditional mean $\mathbb{E}(Y_2|Y_1)$ is the best possible prediction of Y_2 given Y_1
- The conditional probability density function $y_2 \to f_{Y_2|Y_1=y_1}(y_2)$ can give the probability density function of the corresponding error ($\Rightarrow$ most probable value, probability of threshold exceedance...)
- The conditional variance $var(Y_2|Y_1 = y_1)$ summarizes the order of magnitude of the prediction error

Gaussian conditioning theorem

Theorem

Let $(Y_1, Y_2)^t$ be a $(n_1 + 1) \times 1$ Gaussian vector with mean vector $(m_1^t, \mu_2)^t$ and covariance matrix

$$\begin{pmatrix} R_1 & r_{1,2} \\ r_{1,2}^t & \sigma_2^2 \end{pmatrix}$$

Then, conditionaly on $Y_1 = y_1$, Y_2 is a Gaussian vector with mean

$$\mathbb{E}(Y_2|Y_1 = y_1) = \mu_2 + r_{1,2}^t R_1^{-1}(y_1 - m_1)$$

and variance

$$var(Y_2|Y_1 = y_1) = \sigma_2^2 - r_{1,2}^t R_1^{-1} r_{1,2}$$

Illustration

When $(Y_1, Y_2)^t$ be a 2×1 Gaussian vector with mean vector $(\mu_1, \mu_2)^t$ and covariance matrix

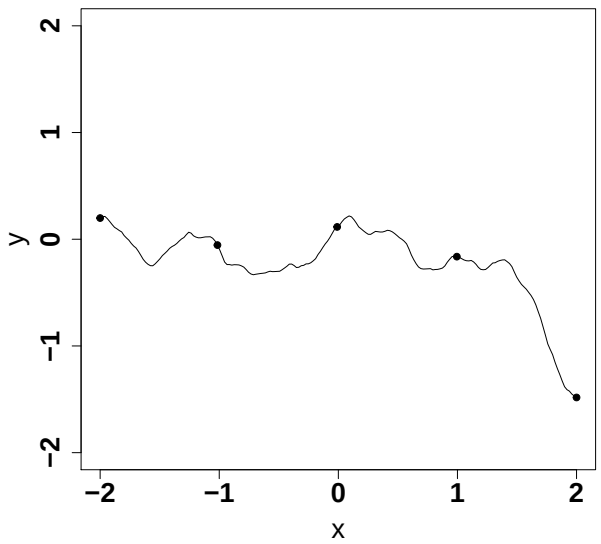
$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Then

$$\mathbb{E}(Y_2|Y_1 = y_1) = \mu_2 + \rho(y_1 - \mu_1) \quad \text{and} \quad var(Y_2|Y_1 = y_1) = 1 - \rho^2$$

A problem of function approximation

We want to approximate a deterministic function, from a finite number of observed values of it.



A possibility : deterministic approximation : polynomial regression, neural networks, splines, RKHS, ...

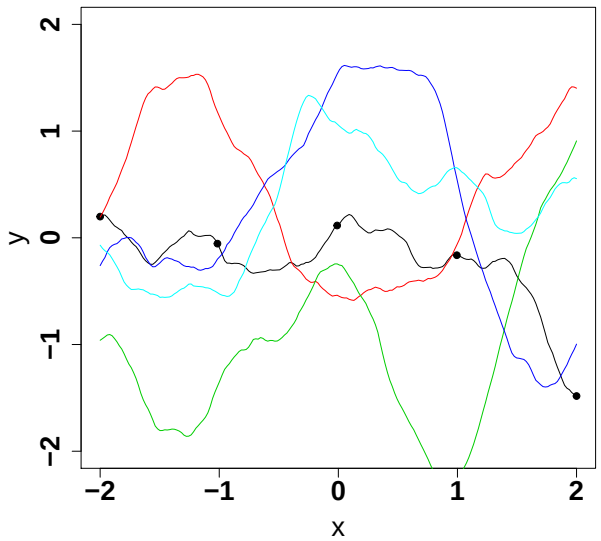
$\longrightarrow$ we can have a deterministic error bound

With a Kriging model : stochastic method

$\longrightarrow$ gives a stochastic error bound

Kriging model with Gaussian process realizations

Kriging model : representing the deterministic and unknown function by a realization of a Gaussian process.



Bayesian interpretation

In statistics, a Bayesian model generally consists in representing a deterministic and unknown number by the realization of a random variable ($\Rightarrow$ enables to incorporate expert knowledge, gives access to Bayes formula...). Here, we do the same with functions

Kriging prediction

We let Y be the Gaussian process, on $\mathbb{R}^d$. Y is observed at $x_1, ..., x_n \in \mathbb{R}^d$. We consider here that we know the covariance function C of Y, and that the mean function of Y is zero

Notations

- Let $Y_n = (Y(x_1), ..., Y(x_n))^t$ be the observation vector. It is a Gaussian vector
- Let R be the $n \times n$ covariance matrix of Y_n : $(R)_{i,j} = C(x_i, x_j)$.
- Let $x_{new} \in \mathbb{R}^d$ be a new input point for the Gaussian process Y. We want to predict $Y(x_{new})$.
- Let r be the $n \times 1$ covariance vector between y and $Y(x_{new})$: $r_i = C(x_i, x_{new})$

Then the Gaussian conditioning theorem gives the conditional mean of $Y(x_{new})$ given the observed values in Y_n :

$$\hat{y}(x_{new}) := \mathbb{E}(Y(x_{new})|Y_n) = r^t R^{-1} Y_n$$

We also have the conditional variance :

$$\hat{\sigma}^2(x_{new}) := var(Y(x_{new})|Y_n) = C(x_{new}, x_{new}) - r^t R^{-1} r$$

Exact reproduction of known values

Assume, $x_{new} = x_1$. Then , $R_{i,1} = C(x_i, x_1) = C(x_i, x_{new}) = r_i$. Thus

$$r^t R^{-1} Y_n = r^t \times \begin{pmatrix} r^t \\ * \\ \vdots \\ * \end{pmatrix}^{-1} \times \begin{pmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix} = (1, 0, \ldots, 0) \begin{pmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix} = Y(x_1)$$

Conservative extrapolation

Let x_{new} be far from $x_1, ..., x_n$. Then, we generally have $r_i = C(x_i, x_{new}) \approx 0$. Thus

$$\hat{y}(x_{new}) = r^t R^{-1} Y_n \approx 0$$

and

$$\hat{\sigma}^2(x_{new}) = C(x_{new}, x_{new}) - r^t R^{-1} r \approx C(x_{new}, x_{new})$$

$\Rightarrow$ conservative

Illustration of Kriging prediction

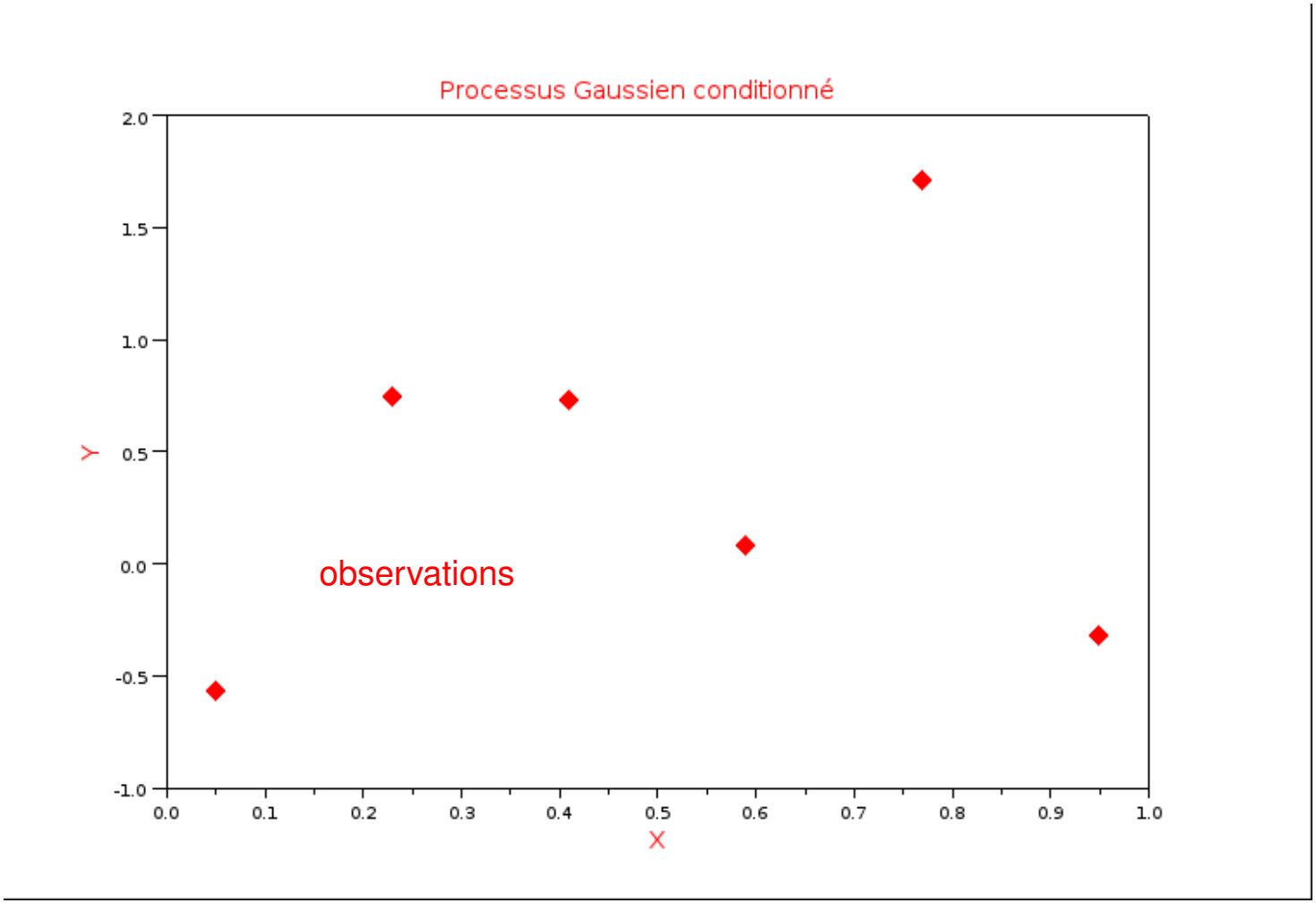


Illustration of Kriging prediction

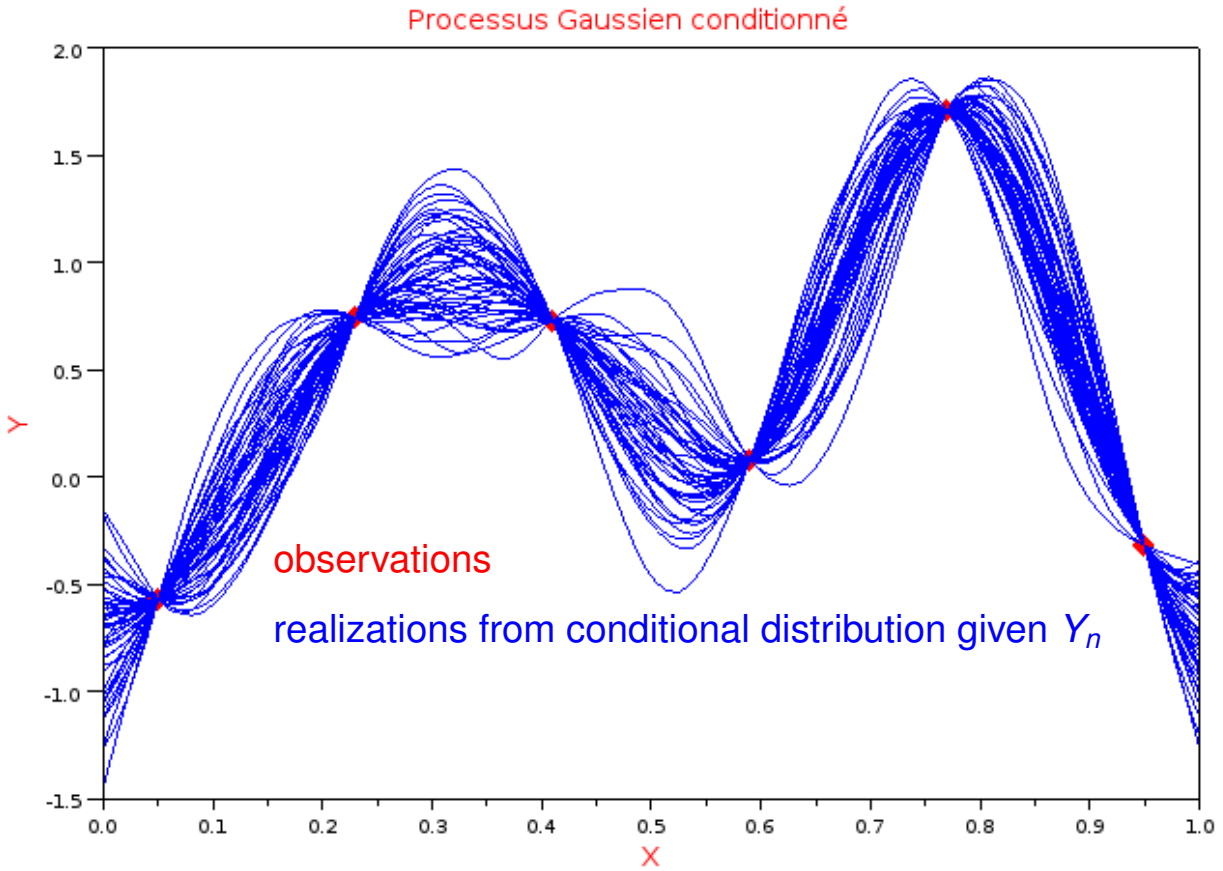


Illustration of Kriging prediction

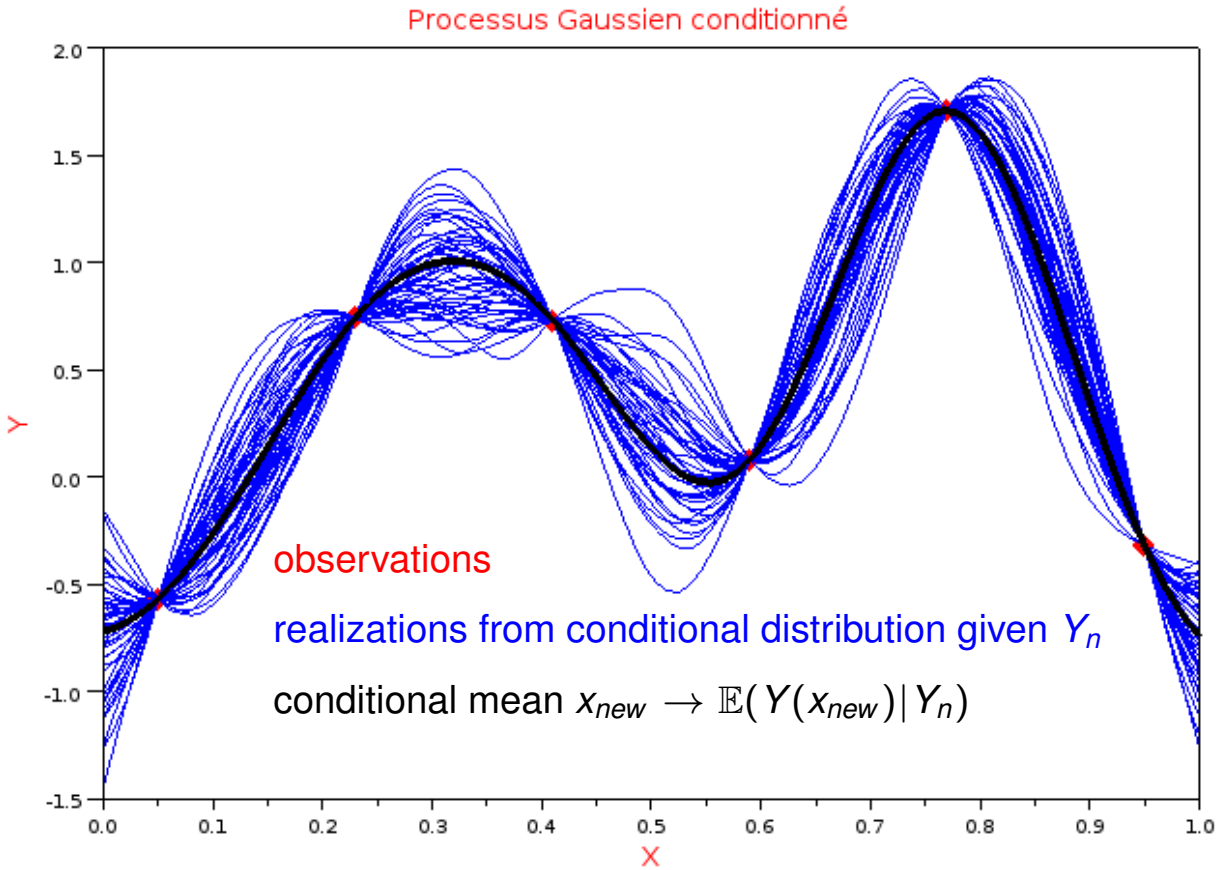
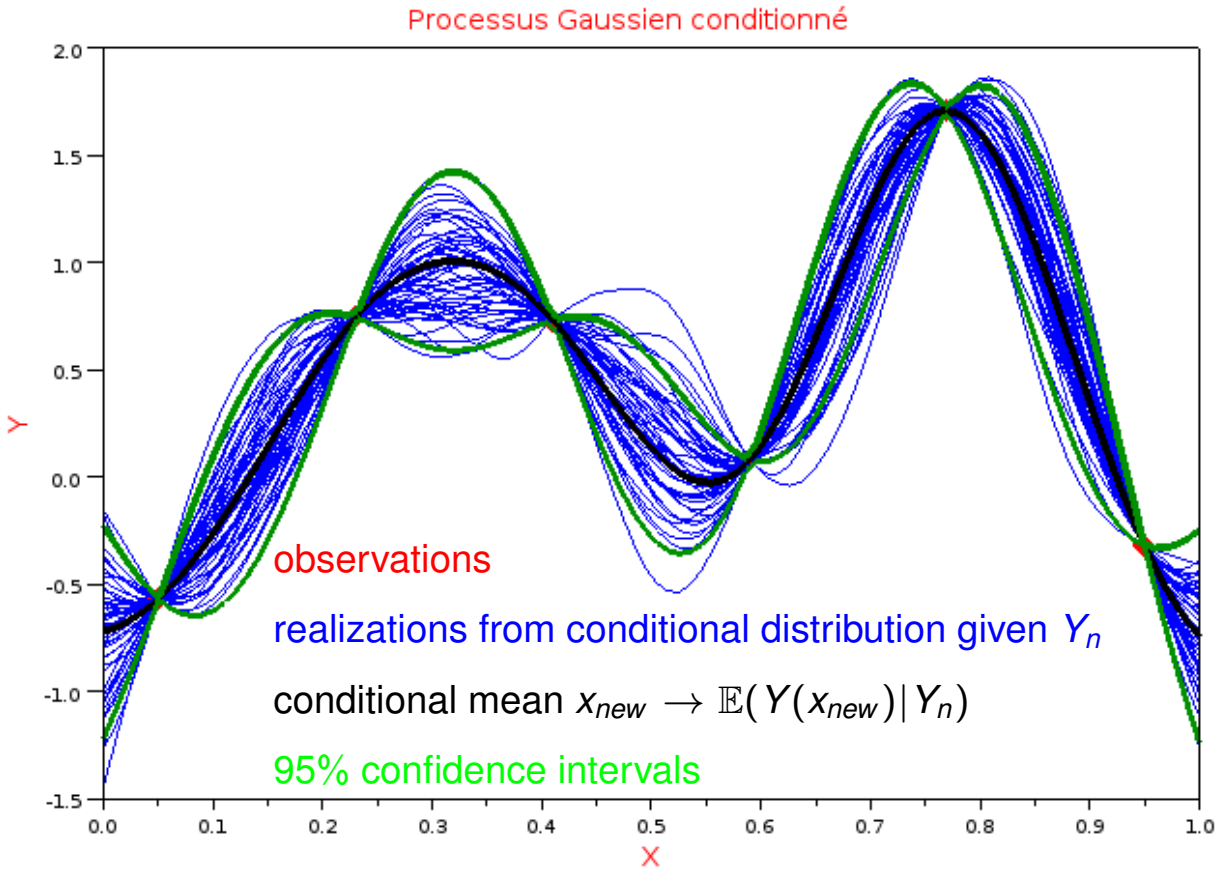


Illustration of Kriging prediction



Kriging prediction with measure error

It can be desirable not to reproduce the observed value exactly :

- when the observations comes from experiments $\Rightarrow$ variability of the response for a fixed input point
- even when the response is fixed for a given input point, it can vary strongly between very close input points

Observations with measure error

We consider that at $x_1, ..., x_n$, we observe $Y(x_1) + \epsilon_1, ..., Y(x_n) + \epsilon_n$. $\epsilon_1, ..., \epsilon_n$ are independent and are Gaussian variables, with mean 0 and known variance σ^2_{mes}.

- We Let $Y_n = (Y(x_1) + \epsilon_1, ..., Y(x_n)\epsilon_n)^t$

Then the Gaussian conditioning theorem still gives the conditional mean of $Y(x_{new})$ given the observed values in Y_n :

$$\hat{y}(x_{new}) := \mathbb{E}(Y(x_{new})|Y_n) = r^t(R + \sigma^2_{mes} I_n)^{-1} Y_n$$

We also have the conditional variance :

$$\hat{\sigma}^2(x_{new}) := var(Y(x_{new})|Y_n) = C(x_{new}, x_{new}) - r^t(R + \sigma^2_{mes} I_n)^{-1} r$$

Illustration of Kriging prediction with measure error

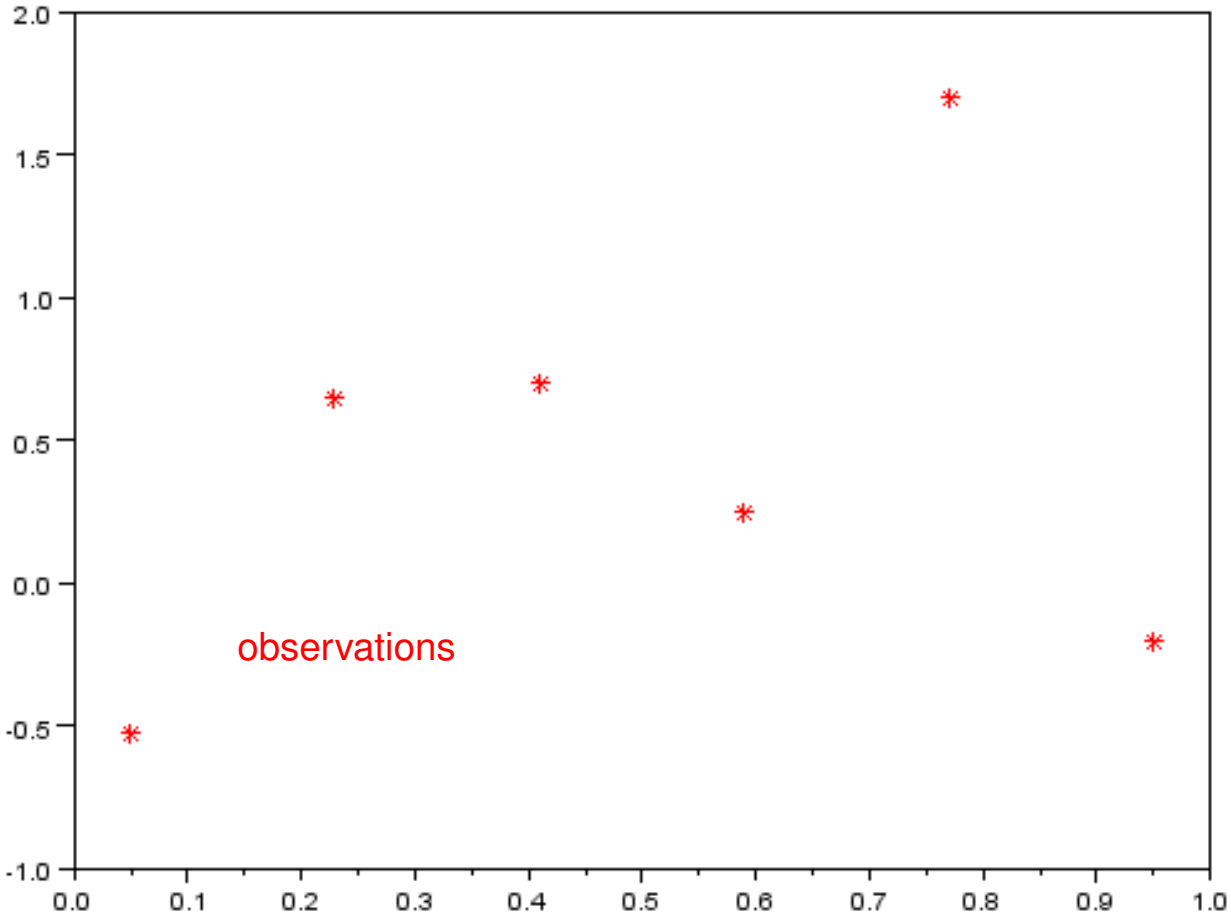


Illustration of Kriging prediction with measure error

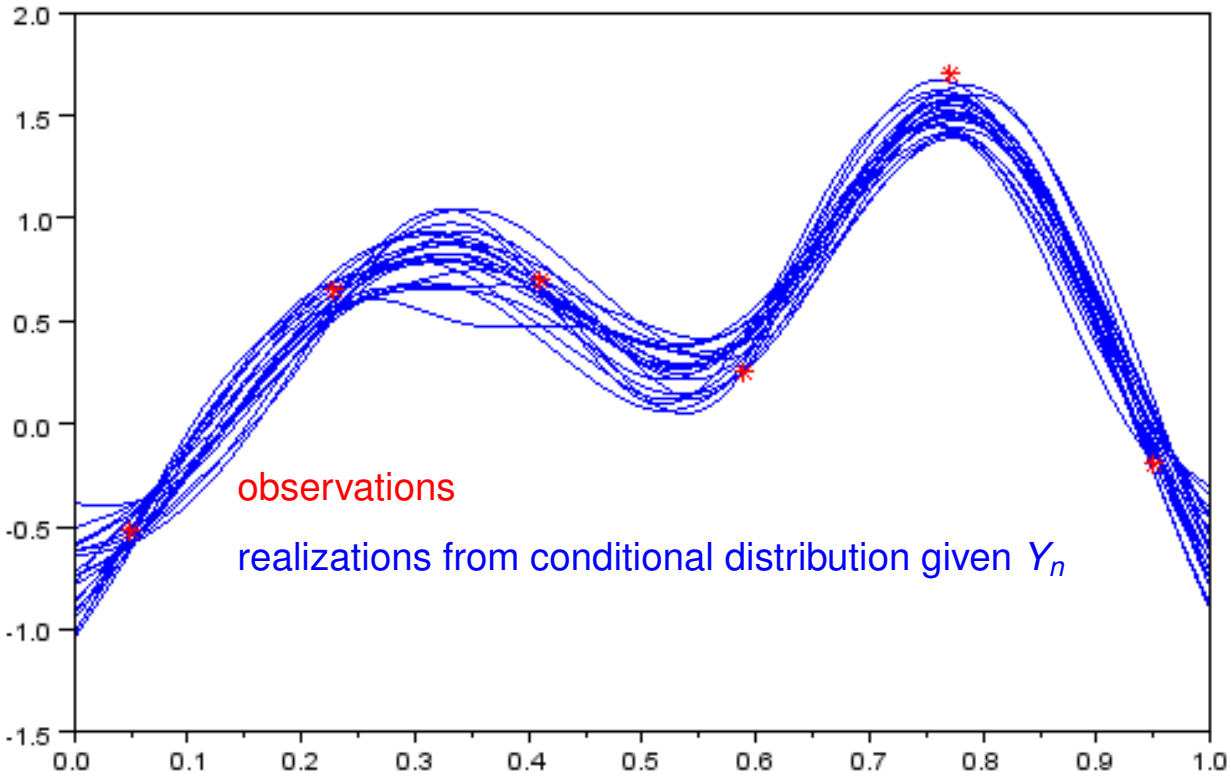


Illustration of Kriging prediction with measure error

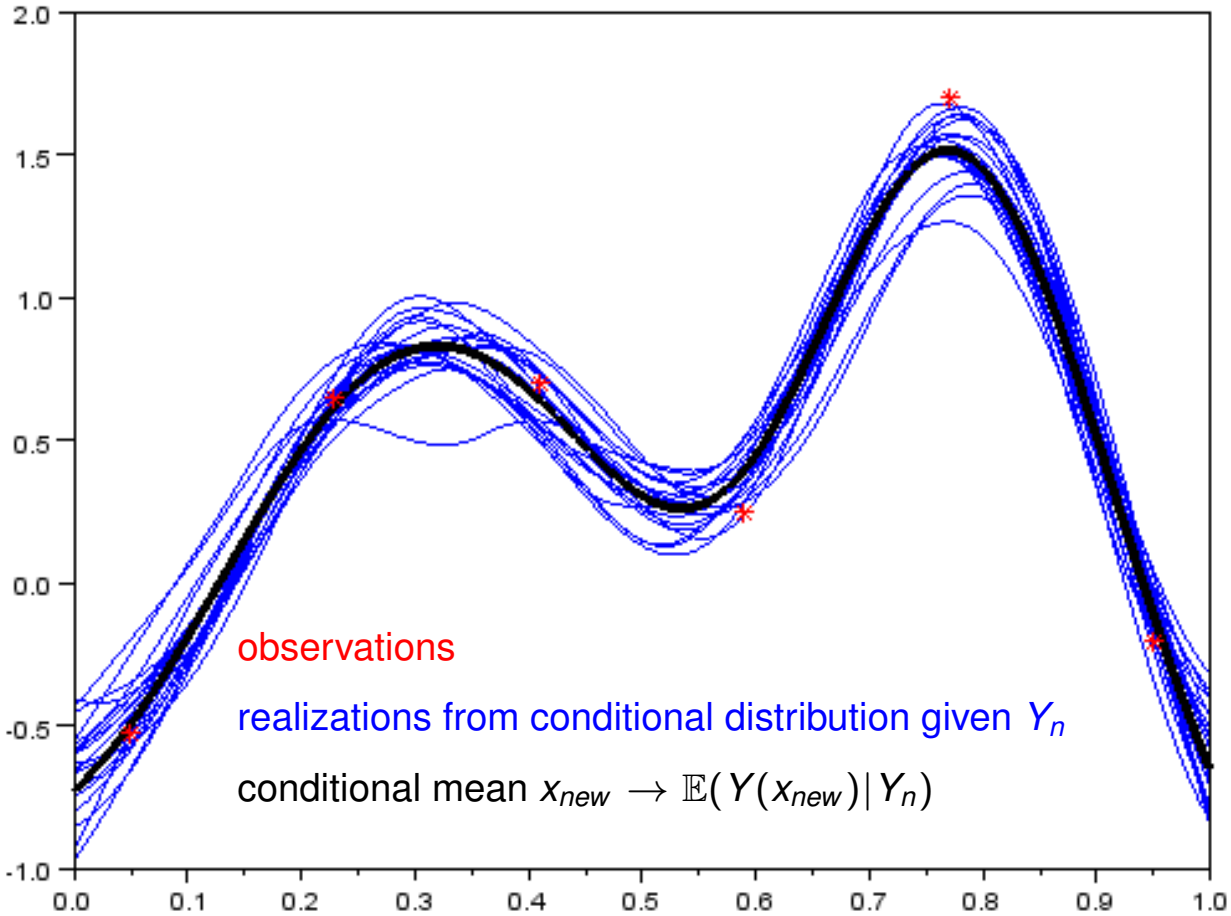
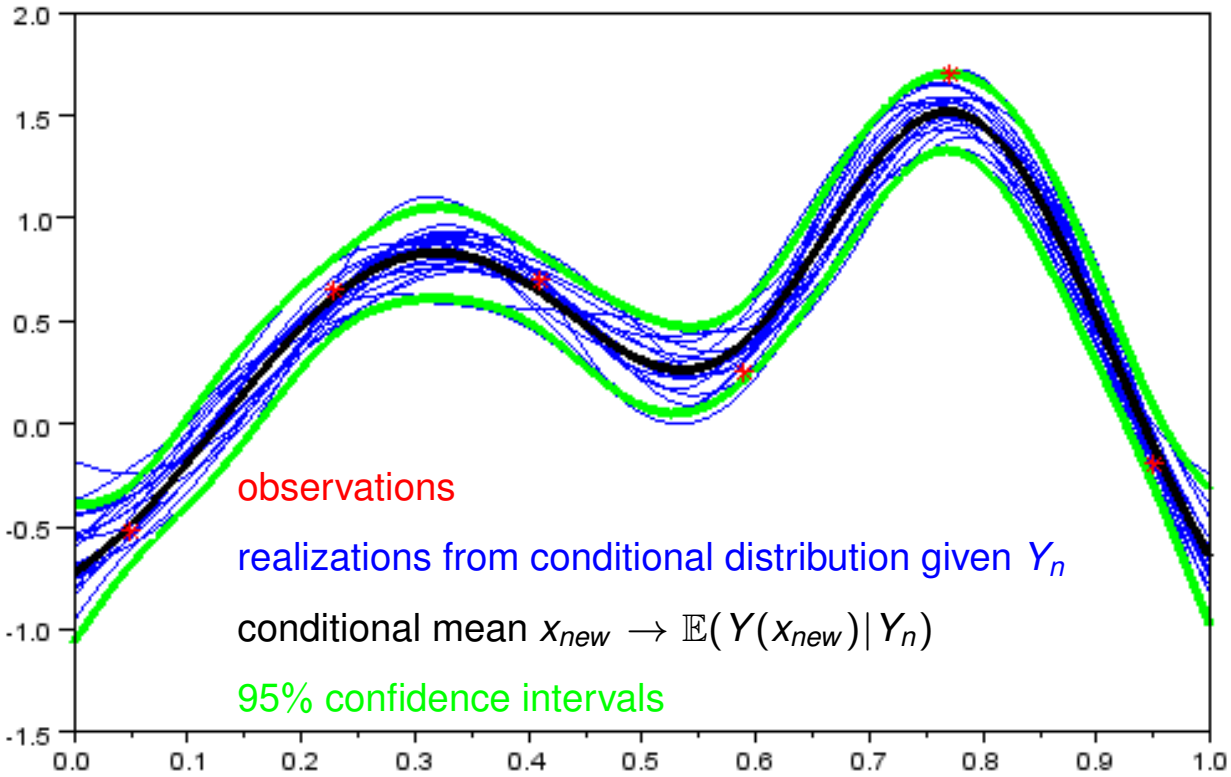


Illustration of Kriging prediction with measure error



Covariance function estimation

In most practical cases, the covariance function $(x_1, x_2) \to C(x_1, x_2)$ is unknown.
It is important to choose it correctly.
In practice, it is first constrained in a parametric family of the form

$$\{C_\theta, \theta \in \Theta\}, \quad \Theta \subset \mathbb{R}^p$$

$\Rightarrow$ E.g. the multidimensional Matérn $\frac{3}{2}$ covariance function model on $\mathbb{R}^d$, with $\theta = (\sigma^2, \ell_1, ..., \ell_d)$
Then, most classically, the covariance parameter θ is automatically selected by Maximum Likelihood
In the case without measure errors :

- Let $Y_n = (Y(x_1), ..., Y(x_n))^t$ be the $n \times 1$ observation vector
- Let R_θ, be the $n \times n$ covariance matrix of Y_n, under covariance parameter θ : $(R_\theta)_{i,j} = C_\theta(x_i, x_j)$.

The Maximum Likelihood estimator $\hat{\theta}_{ML}$ of θ is then :

$$\hat{\theta}_{ML} \in \underset{\theta \in \Theta}{\operatorname{argmin}} \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{det(R_\theta)}} \exp\left(-\frac{1}{2} Y_n^t R_\theta^{-1} Y_n\right)$$

- We maximize the Gaussian probability density function of the observation vector, as a function of the covariance parameter
- Numerical optimization problem, where the cost function has a $O(n^3)$ computational cost

Summary for Kriging prediction

Most classical method :

1. From observed values gathered in the vector Y_n
2. Choose a covariance function family, parameterized by θ
   - Generally before investigating the observed values in detail and from a limited number of classical options (e.g. Matérn $\frac{3}{2}$)
3. Optimize the Maximum Likelihood criterion w.r.t $\theta \Rightarrow \hat{\theta}_{ML}$
   - Numerical optimization : gradient, quasi Newton, genetic algorithm... Potential condition-number problems
4. In the sequel, do as if the estimated covariance function $C_{\hat{\theta}_{ML}}(x_1, x_2)$ is the true covariance function (plug-in method).
5. Compute the conditional mean $x_{new} \to \mathbb{E}(Y(x_{new})|Y_n)$ and the conditional variance $x_{new} \to var(Y(x_{new})|Y_n)$ with explicit matrix vector formulas (Gaussian conditioning theorem)

1. Introduction to Gaussian processes

2. Kriging prediction
   - Conditional distribution and Gaussian conditioning theorem
   - Kriging prediction

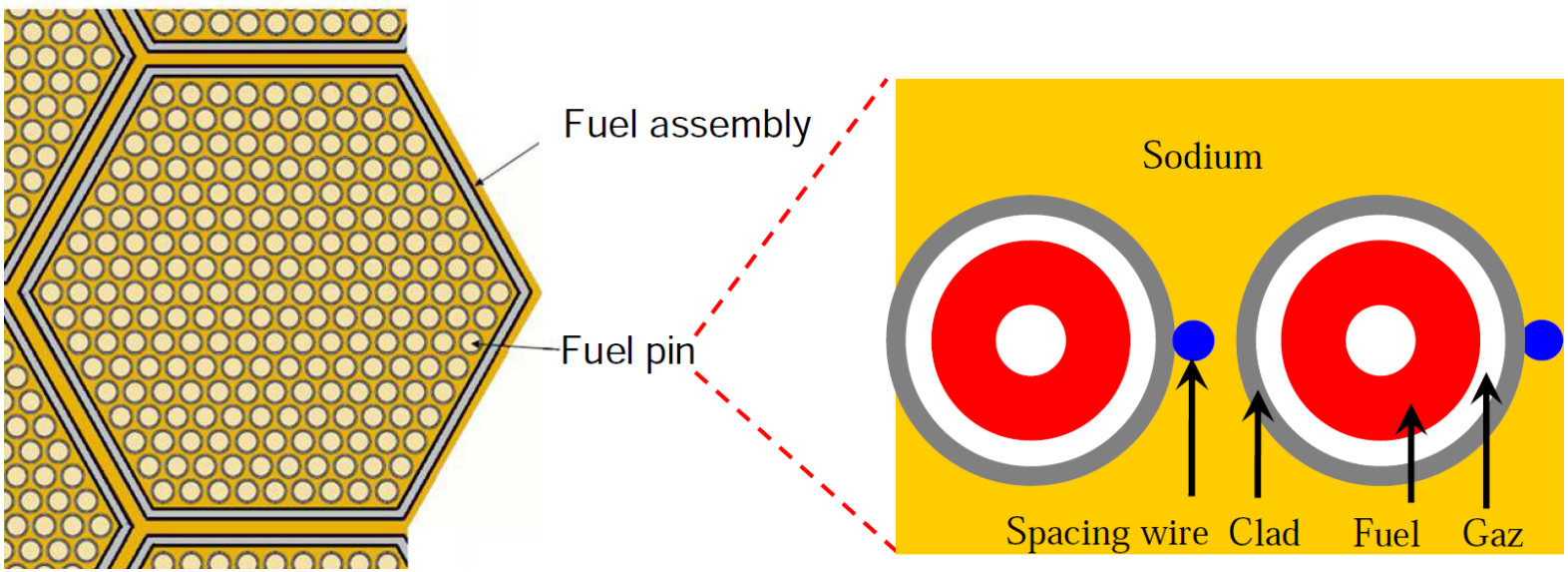
3. Application to metamodeling of the GERMINAL code

4. Application to validation of the FLICA 4 thermal-hydraulic code

GERMINAL code : presentation

Context

- GERMINAL code : simulation of the thermal-mechanical impact of the irradiation on a nuclear fuel pin
- Its utilization is part of a multi-physics and multi-objective optimization problem from reactor core design
- In collaboration with Karim Ammar (PhD student, CEA, DEN)



GERMINAL code : inputs and outputs

12 inputs $x_1, ..., x_{12} \in [0, 1]$ (normalization)

- x_1, x_2 : Schedule parameters for the exploitation of the fuel pin
- $x_3, ..., x_8$: Nature parameters of the fuel pin (geometry, plutonium concentration)
- x_9, x_{10}, x_{11} : parameters for the characterization of the power map in the fuel pin
- x_{12} : disposal volume for the fission gas produced in the fuel pin

2 scalar variable of interests

- g_1 : initial temperature. Maximum, over space, of the temperature at the initial time. Rather simple to approximate
- g_2 : fusion-margin. Minimum difference, over space and time, of the fusion temperature of the fuel and the current temperature. More difficult to approximate

general scheme

12 scalar inputs $\Rightarrow$ GERMINAL run $\Rightarrow$ spatio-temporal maps $\Rightarrow$ 2 scalar outputs

$\rightarrow$ We want to approximate 2 functions $g_1, g_2 : \mathbb{R}^{12} \rightarrow \mathbb{R}$

GERMINAL code : data bases and measure error

Data bases

For the first output g_1, we have a learning base of $n = 15722$ points (15722 couples $(x_1, g_1(x_1)), ..., (x_n, g_1(x_n))$ with $x_i \in \mathbb{R}^{12}$). We have a test base of $n_{test} = 6521$ elements.
For the second output g_2, we have $n = 3807$ and $n_{test} = 1613$

Measure errors

The GERMINAL computation scheme (GERMINAL + pre and post-treatment) had not been used for so many inputs $\rightarrow$ numerical instabilities (some very close inputs can give significantly distant outputs)
$\Rightarrow$ we incorporate the measure error parameter σ^2_{mes} to model numerical instabilities (estimated by Maximum Likelihood, together with covariance function parameters)

Metamodel

A metamodel of g is a function $\hat{g} : [0,1]^{12} \to \mathbb{R}$, that is built using the learning base only.
We consider 2 metamodels :

- The Kriging conditional mean (with Matérn $\frac{3}{2}$ covariance function and measure error variance estimated by Maximum Likelihood)
- A neural-network method, of the uncertainty platform URANIE

$\Rightarrow$ Once built, the cost of computing $\hat{g}(x_{new})$ for a new $x_{new} \in [0,1]^{12}$ is very small compared to a GERMINAL run.

Error indicator

Root Mean Square Error (RMSE) on the test base :

$$RMSE = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (\hat{g}(x_{test,i}) - g(x_{test,i}))^2}$$

Results

For initial temperature g_1 (standard deviation of 344˚)

	estimated σ_{mes}	RMSE
Kriging	7.8˚	9.03˚
Neural networks		11.9˚

For Fusion Margin g_2 (standard deviation of 342˚)

	estimated σ_{mes}	RMSE
Kriging	28˚	35.9˚
Neural networks		39.7˚

- Confirmation that output g_2 is more difficult to predict than g_1
- In both cases, a significant part of the RMSE comes from the numerical instability, of order of magnitude σ_{mes}
- The metamodels have overall quite good performances (3% and 10% relative error)
- The Kriging metamodel has here comparable to slightly larger accuracy than the neural networks
- On the other hand, the neural network metamodel is significantly faster than Kriging (computational cost in $O(n)$ with n large). Nevertheless both metamodels can be considered as fast enough

1. Introduction to Gaussian processes

2. Kriging prediction
   - Conditional distribution and Gaussian conditioning theorem
   - Kriging prediction

3. Application to metamodeling of the GERMINAL code

4. Application to validation of the FLICA 4 thermal-hydraulic code

Computer code and physical system

The computer code is represented by a function f :

$$\begin{array}{rll} f & : \mathbb{R}^d \times \mathbb{R}^m & \to \mathbb{R} \\ & (x, \beta) & \to f(x, \beta) \end{array}$$

The physical system is represented by a function Y_{real}.

$$\begin{array}{rll} Y_{real} & : \mathbb{R}^d & \to \mathbb{R} \\ & x & \to Y_{real}(x) \\ Y_{obs} & : x & \to Y_{obs}(x) := Y_{real}(x) + \epsilon(x) \end{array}$$

- The inputs in x are the experimental conditions
- The inputs in β are the calibration parameters of the computer code
- The outputs $f(x, \beta)$ and $Y_{real}(x)$ are the variable of interest
- Measure error $\epsilon(x)$

Explaining discrepancies between simulation and experimental results

We have carried out experiments $Y_{obs}(x_1), ..., Y_{obs}(x_n)$. Discrepancies between simulations $f(x_i, \beta)$ and observations $Y_{obs}(x_i)$ can have 3 sources :

- misspecification of β
- measure errors on the observations $Y_{obs}(x_i)$
- Errors on the specifications of the experimental conditions x_i

$\longrightarrow$ These 3 errors can insufficient to explain the differences between simulations and experiments

Gaussian process modeling of the model error

Gaussian process model : unknown physical system $\rightarrow$ represented by a realization of a Gaussian process

$$\begin{aligned} Y_{real}(x) &= f(x,\beta) + Z(x) \\ Y_{obs}(x) &= Y_{real}(x) + \epsilon(x) \end{aligned}$$

- β : calibration parameter
  incorporation of expert knowledge with the Bayesian framework
- Z is the model error of the code. Z is modeled as the realization of a Gaussian process

Universal Kriging model

Linear approximation of the code

$$\forall x : \ f(x, \beta) = \sum_{i=1}^{m} h_i(x)\beta_i$$

$\rightarrow$ small uncertainty on β
Observations stem from a Gaussian process with linearly parameterized mean function with unknown coefficients $\Rightarrow$ universal Kriging model.

3 steps

With similar matrix vector formula and interpretation as for the 0 mean function case :

- Estimation of the covariance function of Z
- Code calibration : conditional probability density function of β
- Prediction of the physical system : conditional mean $\mathbb{E}(Y_{real}(x_{new})|Y_n)$ and conditional variance $var(Y_{real}(x_{new})|Y_n)$

Universal Kriging model : calibration and prediction formula (1/2)

Y_{real} is observed at $x_1, ..., x_n \in \mathbb{R}^d$. We consider here that we know the covariance function C of the model error Z

Notations

- Let $Y_n = (Y_{real}(x_1), ..., Y_{real}(x_n))^t$ be the observation vector. It is a Gaussian vector
- Let R be the $n \times n$ covariance matrix of $(Z(x_1), ..., Z(x_n)) : (R)_{i,j} = C(x_i, x_j)$.
- Let $x_{new} \in \mathbb{R}^d$ be a new input point for the Gaussian process Y_{real}. We want to predict $Y(x_{new})$.
- Let r be the $n \times 1$ covariance vector between $Z(x_1), ..., Z(x_n)$ and $Z(x_{new}) : r_i = C(x_i, x_{new})$
- Let H be the $n \times m$ matrix of partial derivatives of f at $x_1, ..., x_n : H_{i,j} = h_j(x_i)$
- Let h be the $m \times 1$ vector of partial derivatives of f at $X_{new} : h_i = h_i(x_{new})$
- Let σ^2_{mes} be the variance of the measure error

Then the Gaussian conditioning theorem gives the conditional mean of β given the observed values in Y_n :

$$\beta_{post} := \mathbb{E}(\beta | Y_n) = \beta_{prior} + (Q^{-1}_{prior} + H^T(R + \sigma^2_{mes} I_n)^{-1} H)^{-1} H^T (R + \sigma^2_{mes} I_n)^{-1} (Y_n - H\beta_{prior}).$$

Universal Kriging model : calibration and prediction formula (2/2)

We also have the conditional mean of $Y_{real}(x_{new})$:

$$\hat{y}_{real}(x_{new}) := \mathbb{E}(Y_{real}(x_{new})|Y_n) = h^t\beta_{post} + r^t(R + \sigma^2_{mes}I_n)^{-1}(Y_n - H\beta_{post})$$

The conditional variance of $Y_{real}(x_{new})$ is

$$\begin{aligned}\hat{\sigma}^2(x_{new}) &:= var(Y_{real}(x_{new})|Y_n) \\ &= C(x_{new}, x_{new}) - r^t(R + \sigma^2_{mes}I_n)^{-1}r \\ &\quad + (h - H^t(R + \sigma^2_{mes}I_n)^{-1}h)^t(H^t(R + \sigma^2_{mes}I_n)^{-1}H + Q^{-1}_{prior})^{-1}(h - H^t(R + \sigma^2_{mes}I_n)^{-1}r)\end{aligned}$$

Interpretation

- The prediction expression is decomposed into a calibration term and a Gaussian inference term of the model error
- When the code has a small error on the n observations, the prediction at x_{new} uses almost only the calibrated code
- The conditional variance is larger than when the mean function is known

Application to FLICA 4 thermal-hydraulic code

The experiment consists in pressurized and possibly heated water passing through a cylinder. We measure the pressure drop between the two ends of the cylinder.
Quantity of interest : The part of the pressure drop due to friction : ΔP_{fro}
Two kinds of experimental conditions :

- System parameters : Hydraulic diameter D_h, Friction height H_f, Channel width e
- Environment variables : Output pressure P_s, Flowrate G_e, Parietal heat flux Φ_p, Liquid enthalpy h_e^l, Thermodynamic title X_{th}^e, Input temperature T_e

We have 253 experimental results

Friction model of FLICA 4

Code based on the (local) analytical model

$$\Delta P_{fro} = \frac{H}{2\rho D_h} G^2 f_{iso} f_h.$$

with

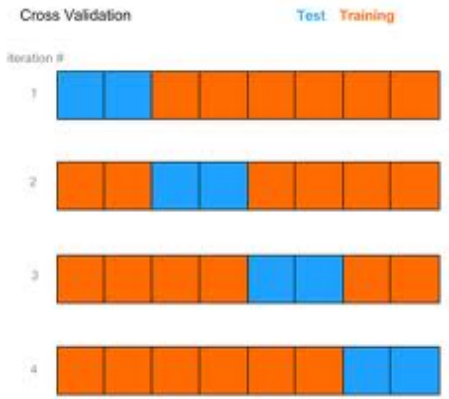
- f_{iso} : Isothermal model. Parameterized a_t and b_t.
- f_h : Monophasic model.

Prior information case with

$$\beta_{prior} = \begin{pmatrix} 0.22 \\ 0.21 \end{pmatrix}, Q_{prior} = \begin{pmatrix} 0.11^2 & 0 \\ 0 & 0.105^2 \end{pmatrix}$$

Results

We compare predictions to observations using Cross Validation



We dispose of :

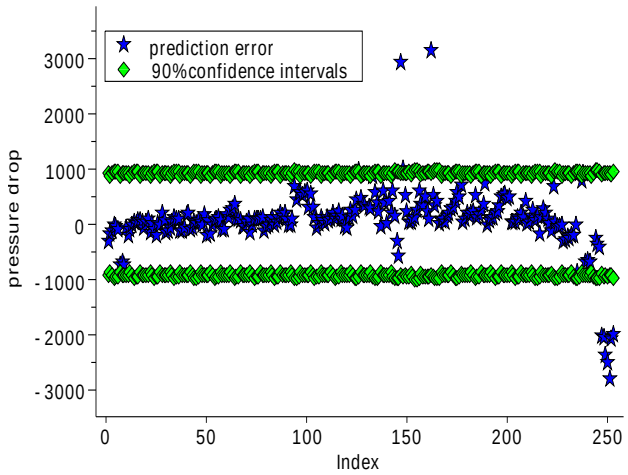
- The vector of posterior mean $\Delta\hat{P}_{fro}$ of size n.
- The vector of posterior variance σ^2_{pred} of size n.

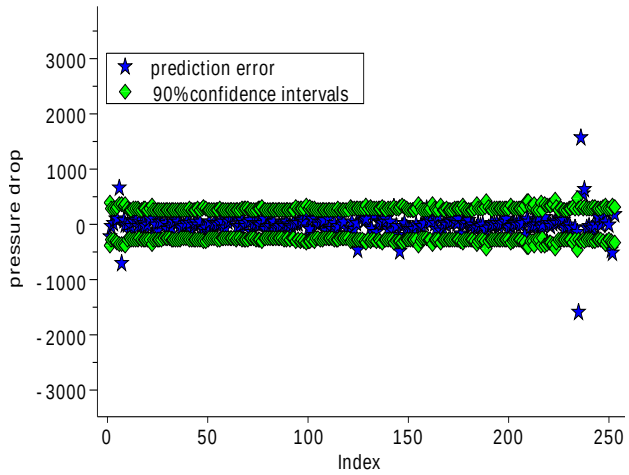
2 quantitative criteria :

- RMSE : $\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\Delta P_{fro,i} - \hat{\Delta} P_{fro}(x_i)\right)^2}$
- Confidence Interval Reliability : proportion of observations that fall in the posterior 90% confidence interval.

Results with the thermal-hydraulic code Flica IV (2/2)

	RMSE	90% Confidence Interval Reliability
Nominal code	$661Pa$	$234/253 \approx 0.925$
Gaussian Processes	$189Pa$	$235/253 \approx 0.93$





Conclusion

- We can improve the predictions of a computer code by completing it with a Kriging model built with the experimental results
- The number of experimental results needs to be sufficient. No extrapolation

For more details

Bachoc F, Bois G, Garnier J and Martinez J.M, Calibration and improved prediction of computer models by universal Kriging, *Nuclear Science and Engineering 176(1) (2014) 81-97*.

General conclusion

Standard Kriging framework

- Versatile and easy-to-use statistical model
- We can incorporate a priori knowledge in the choice of the covariance function family
- After this choice, the standard method is rather automatic
- We associate confidence intervals to the predictions
- The Gaussian framework brings numerical criteria for the quality of the obtained model

Extensions

- Kriging model can be goal-oriented : optimization, code validation, estimation of failure regions, global sensitivity analysis...
- Standard Kriging method can be computationally costly for large $n \Rightarrow$ approximate Kriging prediction and covariance function estimation is a current research domain

Some references (1/2)

General books


C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, *The MIT Press, Cambridge (2006)*.


Santner, T.J, Williams, B.J and Notz, W.I, The Design and Analysis of Computer Experiments *Springer, New York (2003)*.


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

On covariance function estimation


F. Bachoc, Cross Validation and Maximum Likelihood Estimations of Hyper-parameters of Gaussian Processes with Model Mispecification, *Computational Statistics and Data Analysis 66 (2013) 55-69*.


F. Bachoc, Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes, *Journal of Multivariate Analysis 125 (2014) 1-35*.

Kriging for optimization


D. Ginsbourger and V. Picheny, Noisy kriging-based optimization methods : a unified implementation within the DiceOptim package, *Computational Statistics and Data Analysis 71 (2013) 1035-1053*.

Some references (2/2)

Kriging for failure-domain estimation

C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, and Y. Richet, Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set, *Technometrics, (2014)*.

Kriging with multifidelity computer codes

L. Le Gratiet, Bayesian Analysis of Hierarchichal Multifidelity Codes, *SIAM/ASA J. Uncertainty Quantification (2013)*.

L. Le Gratiet and J. Garnier, Recursive co-kriging model for Design of Computer experiments with multiple levels of fidelity, *International Journal for Uncertainty Quantification*.

Kriging for global sensitivity analysis

A. Marrel, B. Iooss, S. da Veiga and M. Ribatet, Global sensitivity analysis of stochastic computer models with joint metamodels, *Statistics and Computing, 22 (2012) 833-847*.

Kriging for code validation

S. Fu, Inversion probabiliste bayésienne en analyse d'incertitude, *PhD thesis, Université Paris-Sud 11*.

Thank you for your attention !