Nurisp SP4 Meeting
Gaussian Processes for code validation

François Bachoc
Jean-Marc Martinez

CEA-Saclay, DEN, DM2S, STMF, LGLS, F-91191 Gif-Sur-Yvette, France

April 2012

Introduction

Context

- Phd started in October 2010 in partnership between CEA and Paris VII university.
- CEA supervisor : Jean-Marc Martinez.
- Paris VII supervisor : Josselin Garnier.

Subject

- Context of code validation : Is the code in agreement with a set of reference experiments ?
- Gaussian processes validation : Modelling of the error between the code and the physical system.
- Goals :
  - Calibration of the code
  - Completion of the code by a statistical term based on a set of experiments

Numerical code and physical system

A numerical code, or parametric numerical model, is represented by a function f :

$$\begin{aligned} f \quad &: \mathbb{R}^d \times \mathbb{R}^m \quad \rightarrow \mathbb{R} \\ &(x, \beta) \quad \rightarrow f(x, \beta) \end{aligned}$$

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

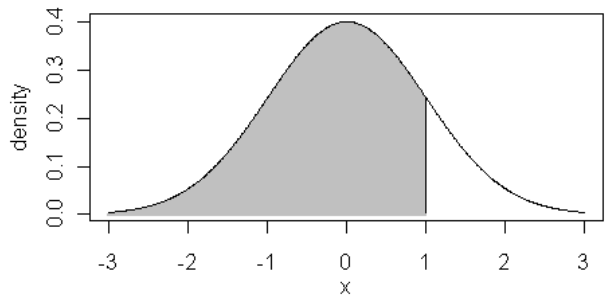
$$\begin{aligned} Y_{real} \quad &: \mathbb{R}^d \quad \rightarrow \mathbb{R} \\ &x \quad \rightarrow Y_{real}(x) \end{aligned}$$

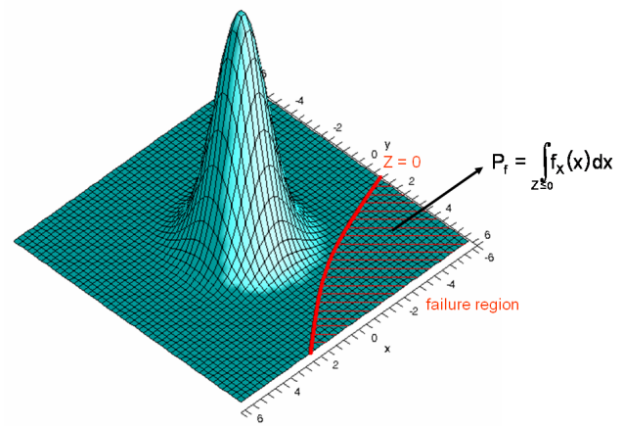
- The inputs x are the experimental conditions.
- The inputs β are the calibration parameters of the numerical code.
- The outputs $f(x, \beta)$ and $Y_{real}(x)$ are the quantity of interest.

A numerical code modelizes (gives an approximation of) a physical system.

Probability notions (1/3)

- A Random Variable X is a random number, defined by a probability law.
- The probability law is defined by a probability density function f with $a \leq X \leq b$ with probability $\int_a^b f(x)dx$
- Similarly a Random Vector $V = (V_1, ..., V_n)$ is a vector of random variable, and is also defined by a probability law.
- The probability law is also defined by a probability density function f with $V \in E$ with probability $\int_E f(v)dv$





Probability notions (2/3)

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

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

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

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

  - High covariance $\longrightarrow$ X and Y have their variations linked.
  - Low covariance $\longrightarrow$ X and Y are almost independant.

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

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

  - High variance $\longrightarrow$ X can be far from its mean $\longrightarrow$ more uncertainty.
  - Low variance $\longrightarrow$ X is close to its mean $\longrightarrow$ less uncertainty.

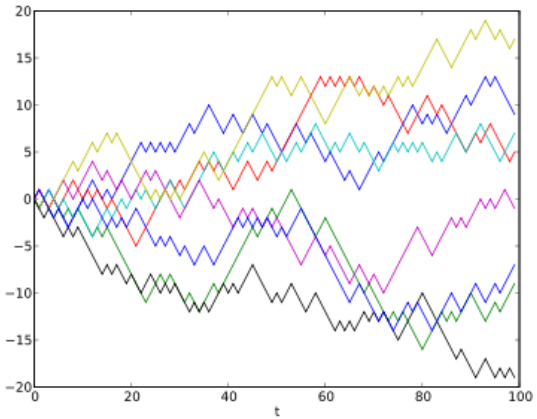
Probability notions (3/3)

Let $V = (V_1, ..., V_n)$ be a random vector. The covariance matrix of V is denoted $cov(V)$ and is 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 links between the components .

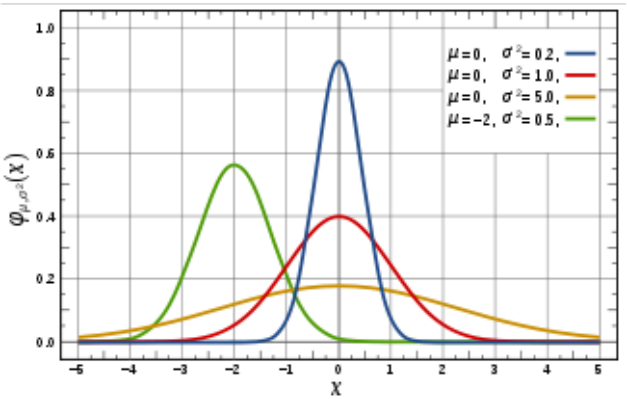
A random function is a function $x \longrightarrow F(x)$ such that $F(x)$ is a random variable. Alternalively a random function is a function that is unknown, or that depends of the hasard.



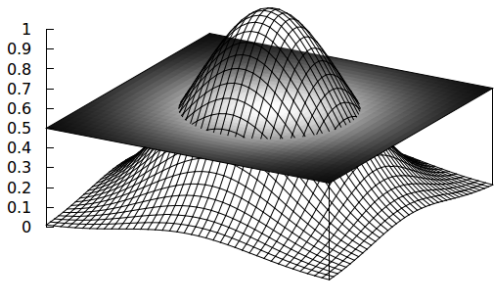
In a random function $F(x)$, x can be multidimensional $\longrightarrow$ it will be the case here

Gaussian variables and vectors

A random variable is a Gaussian variable with mean μ and variance σ^2 when its probability density function is $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$



A n dimensional random vector is a Gaussian vector with mean vector μ and covariance matrix R when its multidimensional probability density function is $f(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det(R)}} \exp\left(-\frac{1}{2}(x-\mu)^t R^{-1}(x-\mu)\right)$



Gaussian processes (1/2)

A random function 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.

In the sequel, we only consider Gaussian processes :

- Gaussian variables : most commonly used to represent errors.
- Gaussian properties make the treatment of the problem simpler.

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

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

Gaussian processes (2/2)

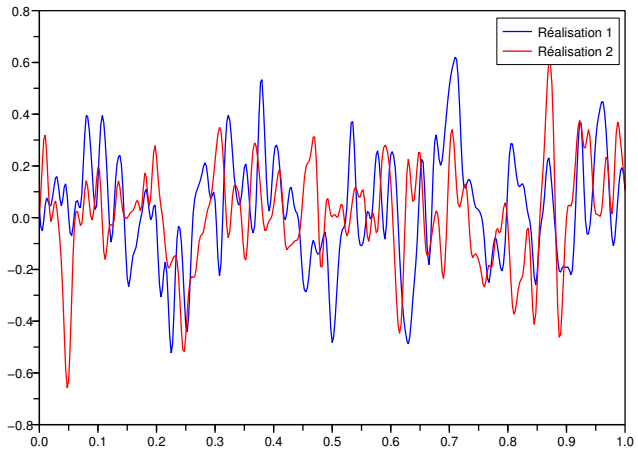
Examples of covariance functions

Nugget covariance function $C(x, y) = \sigma^2 \mathbf{1}_{x=y}$

Gaussian covariance function $C(x, y) = \sigma^2 \exp\left(-\frac{(x-y)^2}{l_c^2}\right)$

Exponential covariance function $C(x, y) = \sigma^2 \exp\left(-\frac{|x-y|}{l_c}\right)$

Examples of realizations with Gaussian covariance function



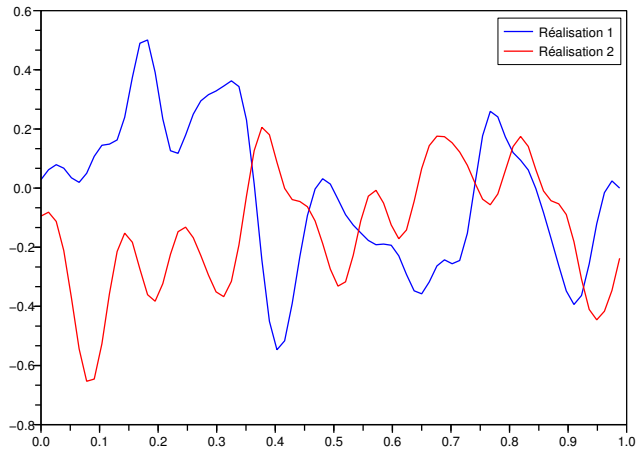


FIG.: Left : $\sigma = 0.2$, $l_c = 0.01$. Right : $\sigma = 0.2$, $l_c = 0.05$

Model error

Statistical modelling : The physical system is unknown $\longrightarrow$ It is one realization among a set of possible realizations $\longrightarrow$ It is modeled as a random function.

Equation of the statistical model

$$Y_{real}(x) = f(x, \beta) + Z(x)$$

- Equation that holds for a specific parameters vector β. Called "the" parameter of the numerical code. We study the Bayesian case in which β is a Gaussian random vector. Its mean vector and covariance matrix are set by the user. The Bayesian framework allows the user to make use of an expert knowledge.
- Z is a Gaussian process. Z has mean 0. We denote by C_{mod} the covariance function of Z.

Steps of Gaussian processes code validation

- Step 1 : Estimation of the covariance function for the model error.
  - Important. Will not be detailed here.
- Step 2 : With a given covariance function : calibration and prediction.
  - Calibration : gives a posterior mean value for the code parameter β and a posterior variance.
  - Prediction : for a new experimental condition x_{new}, gives a posterior mean value for $Y_{real}(x_{new})$ and a posterior variance.

Linear approximation for the code

Linearization of the numerical model around the reference parameter :

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

The approximation is correct when

- The code is approximatively linear with respect to the paramters

or

- The uncertainty of the parameters is small.

Introduction

Assume we have fixed the covariance function C_{mod} of the model error.

- The statistical model is a linear regression model with a Gaussian process error.
- It is the same as the Kriging Model, well known e.g in Geostatistic and in analysis of computer experiments.
- We have closed form formulas for the calibration and the prediction.

Experimental results

We observe the physical system $Y_{real}(x)$ for n inputs $x_1, ..., x_n$.
We keep :

- The vector of observations : $y_{obs} = (Y_{obs}(x_1), ..., Y_{obs}(x_n))$.
- The $n \times m$ matrix of partial derivatives of the code at $x_1, ..., x_n$: H.
- The covariance matrix of $z + \epsilon$ at $x_1, ..., x_n$: $R := R_{mod} + K$
  - R_{mod} is the covariance matrix of the model error process Z. Comes from C_{mod}.
  - K is the covariance matrix of the measure error. Can be set by the user

Calibration

Recall the a priori probability law of β is normal with mean vector β_{prior} and covariance matrix Q_{prior}. The posterior mean of β is

$$\beta_{post} = \beta_{prior} + (Q_{prior}^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} (y_{obs} - H\beta_{prior}).$$

The posterior covariance matrix of β is

$$Q_{post} = (Q_{prior}^{-1} + H^T R^{-1} H)^{-1}$$

- When $Q_{prior} \longrightarrow 0$ $\beta_{post} \longrightarrow \beta_{prior}$.
- When $Q_{prior}^{-1} \longrightarrow 0$ the value of β_{prior} is unused. $\longrightarrow$ uninformative prior.

Prediction (1/2)

Goal : to complete the prediction of the code $f(x_0, \hat{\beta})$ at a new experimental condition x_0.

Notations

- Physical system at x_0 : $y_0 := Y_{real}(x_0)$.
- Column vector of partial derivatives of the code at x_0 : h_0.
- Variance of y_0 : σ^2_{mod}
- Column covariance vector r_0 : $r_{0,i} := cov(Z(x_i), Z(x_0))$.

Prediction (2/2)

The posterior mean of y_0 is :

$$\langle y_{obs,0} \rangle = (h_0)^T \beta_{post} + (r_0)^T R^{-1} (y_{obs} - H\beta_{post})$$

with β_{post} the posterior mean of β.

- 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_0 uses almost only the calibrated code.

The posterior variance of y_0 is

$$\hat{\sigma}^2_{x_0} = \sigma^2_{mod} - r_0^t R^{-1} r_0 + (h_0 - H^t R^{-1} r_0)^t (H^t R^{-1} H + Q^{-1}_{prior})^{-1} (h_0 - H^t R^{-1} r_0)$$

- Confidence intervals available

The experimental results for the friction model

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 dispose of 253 experimental results. 115 are in the isothermal domain and 138 in the monophasic (non-isothermal) domain.
Important : Among the 253 experimental results, only 8 different system parameters $\longrightarrow$ Not enough to use the Gaussian processes model for prediction for new system parameters $\longrightarrow$ We predict for new environment variables only.

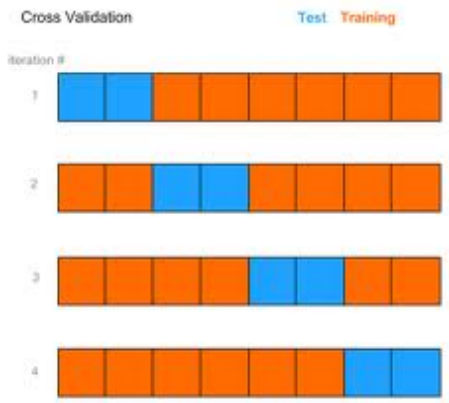
The Flica IV code for the friction model

Parameterized a_t and b_t.
Prior information (coming from previous studies) :

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

Results

We compare predictions to observations using Cross Validation



We dispose of :

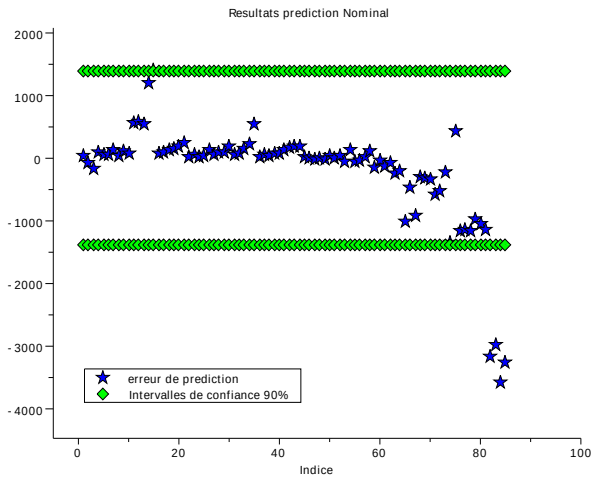
- The vector of posterior mean $\Delta\hat{P}_{fro}$ of size n.
- The vector of posterior variance σ^2_{pred} of size n.

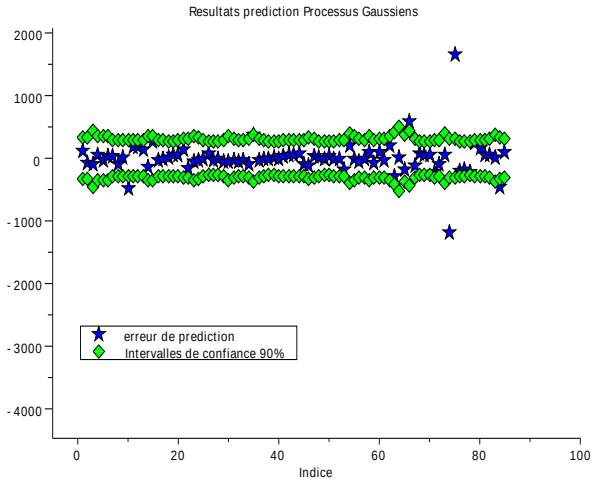
2 quantitative criteria :

- RMSE : $\sqrt{\frac{1}{n}\sum_{i=1}^{85}\left(\Delta P_{fro,i} - \Delta\hat{P}_{fro,i}\right)^2}$
- Confidence Intervals : proportion of observations that fall in the posterior 90% confidence interval.

Results : isothermal domain

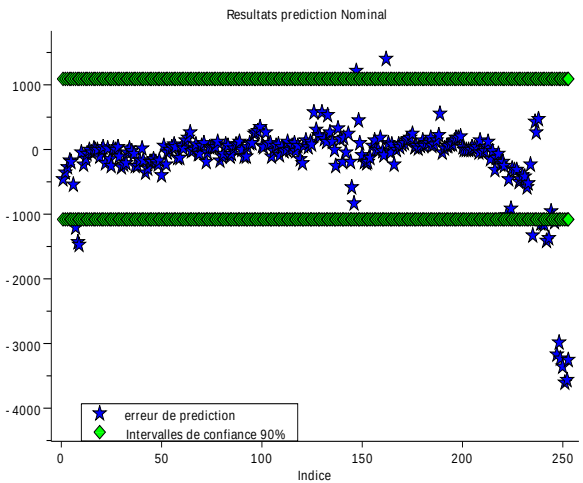
	RMSE	Confidence Intervals
Nominal code	$840Pa$	$80/85 \approx 0.94$
Gaussian Processes	$265Pa$	$79/85 \approx 0.93$

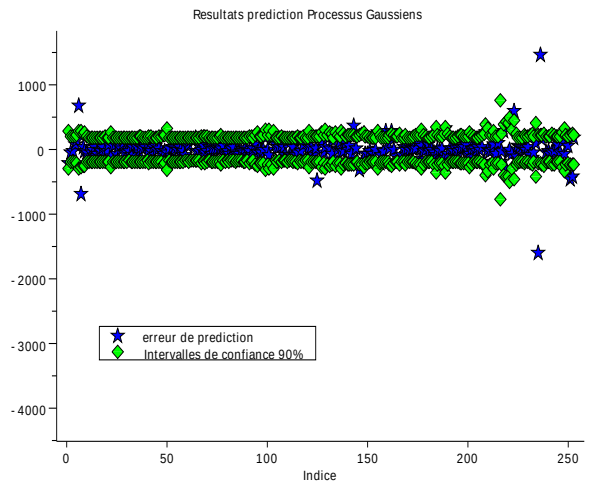




Results : isothermal and monophasic domain

	RMSE	Confidence Intervals
Nominal code	$661 Pa$	$234/253 \approx 0.925$
Gaussian Processes	$189 Pa$	$235/253 \approx 0.93$





Conclusion

- We can improve the prediction capability of the code by completing it with a statistical model based on the experimental results.
- Number of experimental results needs to be sufficient. No extrapolation.
- The choice of the covariance function is important.

Increasing use of probabilistic methods for numerical simulation : Kriging and Gaussian processes methods for surrogate models and code calibration and validation.

Some references

F Bachoc, G Bois, and J.M Martinez.
Contribution à la validation des codes de calcul par processus Gaussiens. Application à la calibration du modéle de frottement pariétal de flica 4.
Technical report, Commissariat à l'Energie Atomique et aux Energies Alternatives DEN/DANS/DM2S/STMF, 2012.

M.L Stein.
Interpolation of Spatial Data Some Theory for Kriging.
Springer, 1999.

T.J Santner, B.J Williams, and W.I Notz.
The Design and Analysis of Computer Experiments.
Springer, 2003.