Bayesian calibration of numerical models using Gaussian processes

François Bachoc

CEA-Saclay, DEN, DM2S, SFME, LGLS, F-91191 Gif-sur-Yvette, France.
Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VII

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

- Probabilistic modelling of the error between a numerical code (or numerical model) and the physical system.
- Goals: To calibrate the numerical code and to improve its predictions.

Numerical code and reality

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

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

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

$$\begin{array}{lll} Y_{real} & : \mathbb{R}^d & \to \mathbb{R} \\ & x & \to Y_{real}(x) \end{array}$$

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

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

Least square calibration

We dispose of a set of experimental results: $x_1, Y_{obs}(x_1), ..., x_n, Y_{obs}(x_n)$.

Least Square calibration:

- Compute:

$$\hat{\beta}_{LS} \in \arg\min_{\beta} \sum_{i=1}^{n} (f(x_i, \beta) - Y_{obs}(x_i))^2$$

- For new experimental condition x_{new} we predict the quantity of interest by: $f(x_{new}, \hat{\beta}_{LS})$.

Least Square calibration: Case of insufficiency

- In general:

$$\sum_{i=1}^{n} \left(f(x_i, \hat{\beta}_{LS}) - Y_{obs}(x_i) \right)^2 \neq 0$$

- First justification: $Y_{obs}(x_i) = Y_{real}(x_i) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma_{mes}^2)$.
- Problem when σ_{mes} (or an upper-bound) is known and when the errors $f(x_i, \hat{\beta}_{LS}) - Y_{obs}(x_i)$ are still too large. (Statistical tests available to detect).
- In these cases: a model error needs to be taken into account.

Gaussian processes (1/3)

Random processes

A real random process Z on $\mathbb{R}^d$ is an application Z: $\Omega \times \mathbb{R}^d \to \mathbb{R}$, with Ω a probability space, so that for all fixed $x \in \mathbb{R}^d$, $\omega \to Z(\omega, x)$ is a random variable.

- Notion of "random function".

Finite dimensional distributions of a random process

Let us consider n points of $\mathbb{R}^d$: $x_1, ..., x_n$. By definition, the vector $(Z(x_1), ..., Z(x_n))$ is a random vector of $\mathbb{R}^n$. Its distribution is said to be a finite dimensional distribution of Z. The finite dimensional distributions of Z are the set of these distributions with n et $x_1, ..., x_n$ varying.

- In the sequel, we only consider finite dimensional distributions: Classical probabilities on $\mathbb{R}^n$.

Gaussian processes (2/3)

Processus gaussien

A real random process Z on $\mathbb{R}^d$ is Gaussian when its finite dimensional distributions are Gaussian.

- In the sequel, we only consider Gaussian processes.

Mean and covariance functions

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

- Finite dimensional distributions of a Gaussian process are caracterized by its mean and covariance functions.

Stationary Gaussian process

A Gaussian process Z is said to be stationary when its mean function M is constant and when $\forall x_1 x_2$: $C(x_1, x_2) = C(x_1 - x_2)$.

Gaussian processes (3/3)

Examples of covariance functions

Nugget model $C(x-y) = \sigma^2 \delta_{x-y}$

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

Examples of realizations with Gaussian covariance function

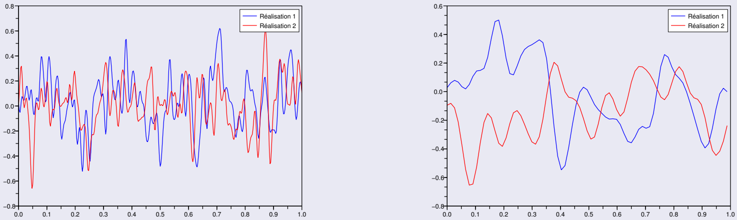


Figure: Left: $\sigma = 0.2$, $l_c = 0.01$. Right: $\sigma = 0.2$, $l_c = 0.05$

Model error

Statistical modelling: The physical phenomenon is one realization among a set of possible realizations. It is modeled as a realization of a random process.

Equation of the statistical model

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

- Equation that holds for a specific parameters vector β. Called "the" parameter of the numerical code.
  - No prior information case: β constant and unknown.
  - Prior information case (Bayesian case): $\beta \sim \mathcal{N}(\beta_{prior}, Q_{prior})$
- Z is (a priori) a centered, stationary, Gaussian process. We denote by C_{mod} the covariance function of Z.

Why a stationary Gaussian process?

- Gaussian variables: most commonly used to represent errors. Gaussian property conserved by conditional expectations and linear transforms.
- Stationarity: restrict the number of possible Gaussian processes (statistical bias-variance trade-off). In statistical inference: replace sample repetition (iid case) by spatial repetition.

Goals associated to the modelling

Kinds of work to do:

1. The covariance function of the model error is known: Calibration and Prediction.
2. A covariance function is proposed: Model test.
3. The covariance function is unknown: Model selection.

Classical outline of studies using the modelling

- Step 1: Estimation of the hyper-parameters of the covariance function.
- Step 2: Plug-in of the estimated hyper-parameters to perform calibration and prediction.

Linear code and observations: notations

Linearization of the numerical model around the reference parameter:

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

Observations

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

- $n \times m$ matrix of partial derivatives of the numerical model: H.
- Random vector of observations: y_{obs}.
- Random vector of measure error: ϵ.
- Random vector of model errors: z.
- Covariance matrix of z: R_{mod}.

Matrix equation of the statistical model

The statistical model becomes, for the inputs $x_1, ..., x_n$:

$$y_{obs} = H\beta + z + \epsilon$$

Covariance matrix of $z + \epsilon$

$$R := cov(z + \epsilon) = R_{mod} + K$$

With $K := cov(\epsilon)$. K is diagonal. Most classical case: $K = \sigma^2_{mes} I$.

- No prior information case
  - When $R = \sigma^2 \mathbf{I}_n$: Classical linear regression model.
- Prior information case

$$y_{obs} \sim \mathcal{N}(H\beta_{prior}, R + HQ_{prior}H^T)$$

- Main interest of the correlation: Efficient prediction of the phenomenon when it does not have the same shape as the numerical code.

Calibration (1/2)

Calibration problem $=$ Statistical estimation problem

Estimation of β

- An estimator of β is a function $\hat{\beta}$: $\mathbb{R}^n \to \mathbb{R}^m$.
- $\hat{\beta}(y_{obs})$ is the estimation of β according to the vector of observations y_{obs}.
- Quality measure of an estimator: Mean square error: $\mathbb{E}_{y_{obs},\beta}\left[||\beta - \hat{\beta}(y_{obs})||^2\right]$.

Calibration (2/2)

No prior information case

The estimator $\hat{\beta}$ of β, linear with respect to the vector of observations y_{obs}, unbiased, which minimizes the mean square error is:

$$\hat{\beta} = (H^T R^{-1} H)^{-1} H^T R^{-1} y_{obs}$$

- If $y_{obs} = H\beta$, $\hat{\beta}(y_{obs}) = \beta$

Prior information case

In the prior information case, the conditional law of β, according to the observations y_{obs} is Gaussian with mean β_{post}, where

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

- Best predictor according to the mean square error.
- When $Q_{prior}^{-1} \to 0$ (Uninformative prior) we find the prediction of the no prior information case, even if $\beta_{prior} \neq 0$.

Prediction (1/4)

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

Notations

- Physical phenomenon at x_0: $y_0 := Y_{real}(x_0)$.
- (pseudo) new observation at x_0: $y_{obs,0}$.
- Column vector of partial derivatives of the code: h_0.
- Random variable of the model error: z_0.
- Random variable of the measure error: ϵ_0.
- Column covariance vector r_0: $r_{0,i} := cov((z + \epsilon)_i, z_0 + \epsilon_0)$.

Prediction of y_0

- A predictor of y_0 is a function $\langle y_0 \rangle : \mathbb{R}^n \to \mathbb{R}$.
- $\langle y_0 \rangle (y_{obs})$ is the prediction of y_0 according to the vector of observations y_{obs}.
- Quality measure of a predicor: Mean square error: $\mathbb{E}_{y_{obs}, y_0} \left[|y_0 - \langle y_0 \rangle (y_{obs})|^2 \right]$.

Prediction (2/4): No prior information case

Prediction

The unbiased predictor of $y_{obs,0}$ at x_0, linear with respect to the vector of observations y_{obs}, which minimizes the mean square error (the BLUP) is:

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

with $\hat{\beta}$ the no prior information case estimator of β.

- We do not have access to the best predictor, because its expression makes use of the unknown parameter β.
- The prediction expression is decomposed into a calibration term and a Gaussian inference term of the model error.

Predictive variance

The mean square error of the BLUP is:

$$\hat{\sigma}^2_{x_0} = \mathbb{E}((z_0 + \epsilon_0)^2) - \begin{pmatrix} h_0 \\ r_0 \end{pmatrix}^t \begin{pmatrix} 0 & H^t \\ H & R \end{pmatrix}^{-1} \begin{pmatrix} h_0 \\ r_0 \end{pmatrix}$$

Confidence intervals available

Prediction (3/4): Prior information case

Prediction

The conditional law of $y_{obs,0}$ according to the observations y_{obs} is Gaussian with mean $\langle y_{obs,0} \rangle$, with:

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

- Best predictor.

Predictive variance

Conditionally to y_{obs} the variance of $y_{obs,0}$ is :

$$\hat{\sigma}^2_{x_0} = \mathbb{E}((z_0 + \epsilon_0)^2) - \begin{pmatrix} h_0 \\ r_0 \end{pmatrix}^t \begin{pmatrix} -Q_{prior}^{-1} & H^t \\ H & R \end{pmatrix}^{-1} \begin{pmatrix} h_0 \\ r_0 \end{pmatrix}$$

- When $Q_{prior}^{-1} \to 0$ (uninformative prior) we find the no prior information case.

Prediction (4/4): from $y_{obs,0}$ to y_0

No prior information case

The BLUP of the observation equals the BLUP of the physical phenomenon:

$$\forall \lambda \in \mathbb{R}^n : \quad \mathbb{E}\left((\lambda^t y_{obs} - y_{obs,0})^2 \right) = \mathbb{E}\left((\lambda^t y_{obs} - y_0)^2 \right) + \mathbb{E}\left((\epsilon_0)^2 \right)$$

Prior information case

The conditional mean are the same and the conditional variance are the same up to the measure error:

- $\mathbb{E}(y_0 | y_{obs}) = \mathbb{E}(y_{obs,0} | y_{obs})$
- $var(y_{obs,0} | y_{obs}) = var(y_0 | y_{obs}) + \mathbb{E}\left((\epsilon_0)^2 \right)$

$\boxed{\longrightarrow}$ In both cases, we keep the same prediction, and remove $\mathbb{E}\left((\epsilon_0)^2 \right)$ to the predictive variance.

Illustration of calibration (1/3)

- Observation of the physical phenomenon: $Y_{obs}(x) = x^2 + \epsilon$. $\epsilon \sim \mathcal{N}(0, \sigma_{mes}^2 = 0.1^2)$
- Numerical code: $f(x, \beta) = \beta_0 + \beta_1 x$.
- Model error as a realization of a Gaussian process with covariance function: $C_{mod}(x - y) = \sigma^2 \exp\left(-\frac{|x-y|^2}{l_c^2}\right)$. $\sigma = 0.3$, $l_c = 0.5$ (known).
- Bayesian case with :

$$\beta_{prior} = \begin{pmatrix} 0.2 \\ 1 \end{pmatrix}, Q_{prior} = \begin{pmatrix} 0.09 & 0 \\ 0 & 0.09 \end{pmatrix}$$

- Observations: $x_1 = 0.2$, $x_2 = 0.4$, $x_3 = 0.6$ and $x_4 = 0.8$.

Illustration of calibration (2/3) (unnoised case)

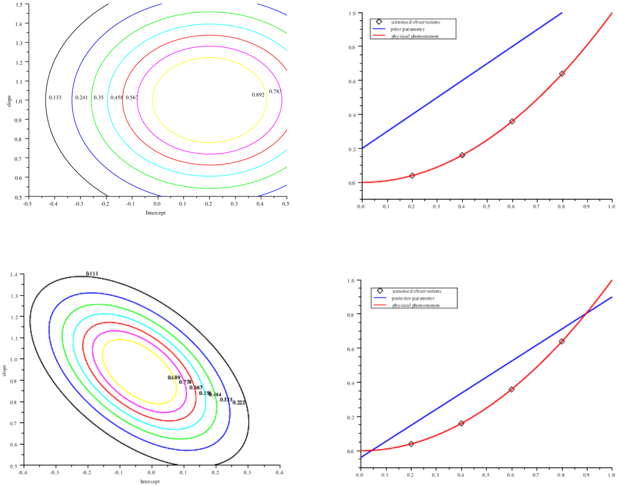


Figure: Up-left: Prior distribution of the parameter β. Down-left: Posterior distribution of the parameter β. Right: plot of the code response corresponding to prior and posterior mean of the code parameter.

Illustration of calibration (3/3) (noised case)

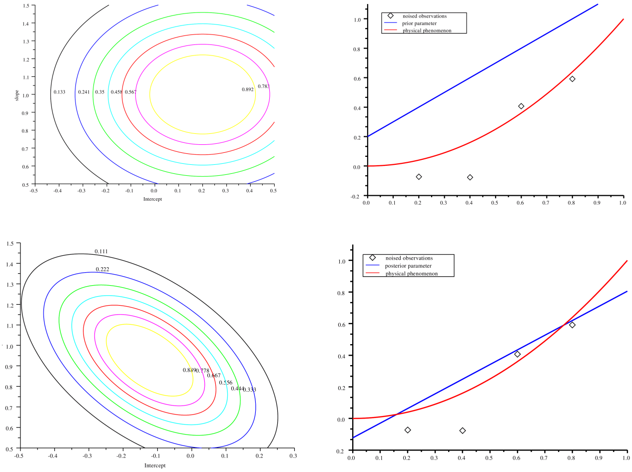
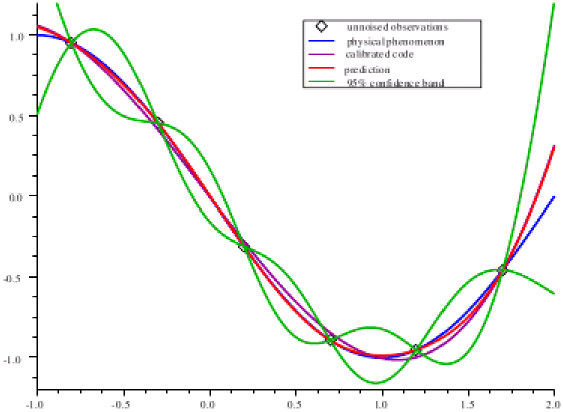


Figure: Up-left: Prior distribution of the parameter β. Down-left: Posterior distribution of the parameter β. Right: plot of the code response corresponding to prior and posterior mean of the code parameter.

Illustration of prediction (1/3)

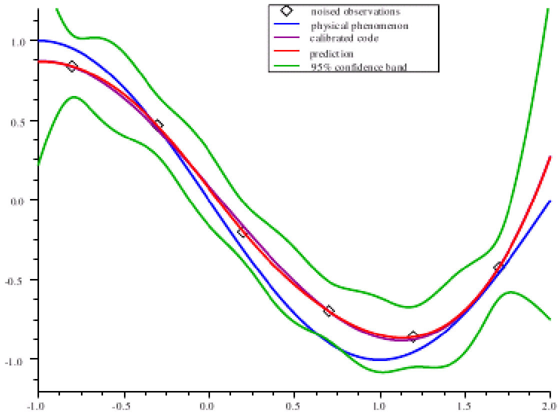
- Observation of the physical phenomenon: $Y_{obs}(x) = -sin(\frac{\pi x}{2}) + \epsilon$. $\epsilon \sim \mathcal{N}(0, \sigma_{mes}^2 = 0.1^2)$
- Numerical code: $f(x, \beta) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.
- Model error as a realization of a Gaussian process with covariance function: $C_{mod}(x - y) = \sigma^2 \exp\left(-\frac{|x-y|^2}{l_c^2}\right)$. $\sigma = 0.3$, $l_c = 0.5$ (known).
- No prior information case.
- 6 observations regularly sampled between -0.8 and 1.7.

Illustration of prediction (2/3) (unnoised case)



- The use of the model error improves the prediction given by the numerical code.

Illustration of prediction (3/3) (noised case)



- The measure error deteriorates the quality of the predictions.
- The confidence intervals are however still reliable.

Framework

- The calibration and prediction methods presented above give good results because we used a reasonable covariance function.
- The model selection is a statistical parameter estimation problem.

In our case, the covariance function of the measure error process ϵ is known for physical expertise. We want to take C_{mod} in a parametric set:

$$\left\{ \sigma^2 C_{mod,\theta} \right\}$$

with $C_{mod,\theta}$ a correlation function.
Hence, with variance matrix $R_{\sigma,\theta} = \sigma^2 R_{mod,\theta} + K$, we have $(z + \epsilon) \sim \mathcal{N}(0, R_{\sigma,\theta})$ and we want to estimate σ and θ.
We present 2 methods for model selection: Restricted Maximum Likelihood and Leave One Out.

Restricted Maximum Likelihood

Principle: Estimate σ and θ independently of β (hence, same method with or without prior information).
Let C a $(n-m \times n)$ matrix of maximal rank such that $CH = 0$. Then we have:

$$w := Cy_{obs} \sim \mathcal{N}(0, CR_{\sigma,\theta}C')$$

We do maximum likelihood on the vector w.
The likelihood writes itself:

$$\ell_{\sigma,\theta}(w) \propto \frac{1}{det(CR_{\sigma,\theta}C^t)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} w^t (CR_{\sigma,\theta}C^t)^{-1} w\right)$$

We maximize it:

$$\hat{\sigma}, \hat{\theta} \in \underset{\sigma,\theta}{\arg\max}\, \ell_{\sigma,\theta}(w).$$

Hence we estimate σ and θ to make the vector w the most probable.

Leave One Out (1/4)

We have seen that the prediction procedure (Bayesian or non-Bayesian framework) leads to a simple stochastic metamodel: $x_0 \rightarrow \mathcal{N}\left(\langle y_{obs,0}\rangle, \hat{\sigma}^2_{x_0}\right)$. This metamodel depends on σ and θ.

- It is built according to the observations ($\approx$ learning set).

Leave One Out

- Given a vector of hyper-parameters (σ, θ).
- For i from 1 to n we learn $x_0 \rightarrow \mathcal{N}\left(\langle y_{obs,0}\rangle, \hat{\sigma}^2_{x_0}\right)$ with the reduced observations vector $\{(x_1, y_{obs,1}), ..., (x_{i-1}, y_{obs,i-1}), (x_{i+1}, y_{obs,i+1}), ..., (x_n, y_{obs,n})\}$
- we compute the LOO errors by:

$$\epsilon_{LOO,i}(\sigma, \theta) = y_{obs,i} - \langle y_{obs,i}\rangle(y_{obs,-i}).$$

- we compute the LOO predictive variance by:

$$\hat{\sigma}^2_{LOO,i}(\sigma, \theta) = \hat{\sigma}^2_{x_i}(y_{obs,-i})$$

General utility of the Leave One Out:

- See how large the errors are.
- Check that the predictive variance are of the right size.

Leave One Out (2/4): closed form formulas

No prior information case

With:

$$Q^{-}(\sigma,\theta) = \left(R_{\sigma,\theta}^{-1} - R_{\sigma,\theta}^{-1}H(H^T R_{\sigma,\theta}^{-1}H)^{-1}H^T R_{\sigma,\theta}^{-1}\right)$$

We have:

$$\epsilon_{LOO}(\sigma,\theta) = (diag(Q^{-}))^{-1}Q^{-}y_{obs} \quad \text{and} \quad \hat{\sigma}^2_{LOO,i}(\sigma,\theta) = \frac{1}{(Q^{-})_{i,i}}$$

Prior information case

With:

$$Q = R_{\sigma,\theta} + HQ_{prior}H^t$$

We have:

$$\epsilon_{LOO}(\sigma,\theta) = (diag(Q^{-1}))^{-1}Q^{-1}y_{obs} \quad \text{and} \quad \hat{\sigma}^2_{LOO,i}(\sigma,\theta) = \frac{1}{(Q^{-1})_{i,i}}$$

Leave One Out (3/4): closed form formulas

- The no prior information case is the limit of the prior information case when $Q_{prior}^{-1} \to 0$.
- From a computational point of view: computing the LOO errors and predictive variance has the same order of complexity than REML and Maximum Likelihood.
- $\boxed{\longrightarrow}$ Can be use as an alternative of Maximum Likelihood techniques.

Leave One Out (4/4): A model selection method

General principle, optimize a quality criterion based on $\epsilon_{LOO}(\sigma, \theta)$ and the $\hat{\sigma}^2_{LOO,i}$. For instance:

- Minimize norm of LOO errors.
- Set number of valid LOO p-confidence intervals close to p.
- Set $\frac{1}{n}\sum_{i=1}^{n} \frac{\epsilon^2_{LOO,i}(\sigma,\theta)}{\hat{\sigma}^2_{LOO,i}(\sigma,\theta)}$ close to 1

When the covariance matrix K of the measure error is null and no prior information case, we have $R_{\sigma,\theta} = \sigma^2 R_{mod,\theta}$, hence:

- $\epsilon_{LOO}(\sigma, \theta)$ independent of σ
- $\hat{\sigma}^2_{LOO}(\sigma, \theta) = \sigma^2 \hat{\sigma}^2_{LOO}(\theta)$

Hence a classical method is:

$$\hat{\theta} \in \arg\min_{\theta} ||\epsilon_{LOO}(\theta)||^2 \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{\epsilon^2_{LOO,i}(\hat{\theta})}{\hat{\sigma}^2_{LOO,i}(\hat{\theta})}$$

When $K \neq 0$ or prior information case: no classical method.

Experiment and model

The experiment consists in the measure of a pressure drop between the two ends of a cylinder crossed by pressurized water and possibly heat. (Representation of the heart of a nuclear reactor).

Quantity of interest: The part of the pressure drop due to friction: ΔP_{fro}

Experimental conditions we consider:

- Hydraulic diameter D_h
- Friction height H_f
- Density ρ
- Viscosity μ
- Flow rate G
- Reynolds coefficient R_e

Model in the isotherm turbulent physical domain is parameterized by a_t, b_t:

$$\text{with} \quad R_e = \frac{G D_h}{\mu}, \qquad \Delta P_{fro}^{mod} = \frac{H_f G^2}{2 \rho D_h} \times a_t R_e^{-b_t}$$

Previous studies of calibration: $a_t = 0.22$, $b_t = 0.21$

We dispose of 85 experimental results in this domain. Hence $n = 85$, $m = 2$.

Need for a model error?

Statistical test

Let C a (83×85) matrix of full rank such that $CH = 0$. Recall K is the covariance matrix of the measure error process and assume there is no model error. Then

$$t_{test} := (Cy_{obs})^t (CKC^t)^{-1} Cy_{obs} \sim \mathcal{X}^2(83).$$

Measure error on experimental conditions

We have nominal measure error variance on 3 experimental conditions:

- $\sigma_{nom}(H_f)$
- $\sigma_{nom}(D_h)$
- $\sigma_{nom}(G)$

They can be formally taken into account in the measure error covariance matrix K. The statistical test is still correct.

Test result:

$\sigma_{mes}(Pa)$	$\alpha(H_f) \times \sigma_{nom}(H_f)$	$\alpha(D_h) \times \sigma_{nom}(D_h)$	$\alpha(G) \times \sigma_{nom}(G)$	t_{test}
100.	$0 \times \sigma_{nom}(H_f)$	$0 \times \sigma_{nom}(D_h)$	$0 \times \sigma_{nom}(G)$	4334.1393
100.	$1 \times \sigma_{nom}(H_f)$	$1 \times \sigma_{nom}(D_h)$	$1 \times \sigma_{nom}(G)$	489.22775
200.	$2 \times \sigma_{nom}(H_f)$	$2 \times \sigma_{nom}(D_h)$	$2 \times \sigma_{nom}(G)$	122.30694

$\boxed{\longrightarrow}$ Need for a model error ($q_{0.95}(\mathcal{X}^2(83)) \approx 105$).

Cross Validation for comparison with Least Square (1/2)

We want to compare the quality of prediction of the LS method and the GP modelling method.
Idea: They both are metamodels of the physical phenomenon.

metamodel

Let f: $\mathcal{X} \to \mathbb{R}$. A metamodel is a function $\hat{f}_{\mathcal{X}_n,\mathcal{Y}_n}$, built from a procedure $\hat{f}$:

$$\hat{f} : [\mathcal{X}_n, \mathcal{Y}_n] = [x_1 ..., x_n, y_1, ..., y_n] \longrightarrow \hat{f}_{\mathcal{X}_n,\mathcal{Y}_n}.$$

With $\hat{f}_{\mathcal{X}_n,\mathcal{Y}_n} : \mathcal{X} \to \mathbb{R}$ an approximation of f.

Evaluation of a metamodel

Quality criterion:

$$C = \frac{1}{Vol(\mathcal{X})} \int_{\mathcal{X}} \left(f(x) - \hat{f}_{\mathcal{X}_n,\mathcal{Y}_n}(x) \right)^2 dx,$$

Ideal case: estimation of this criterion on a new test sample $\mathcal{X}_{test}, \mathcal{Y}_{test}$:

$$C \approx \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left(f(x_{test,i}) - \hat{f}_{\mathcal{X}_n,\mathcal{Y}_n}(x_{test,i}) \right)^2.$$

Cross Validation for comparison with Least Square (2/2)

Evaluation of a metamodel

More realistic case: split the learning data: $\mathcal{X}_n = (\mathcal{X}_{app}, \mathcal{X}_{test})$ $(n = n_{app} + n_{test})$:

$$C \approx \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left(f(x_{test,i}) - \hat{f}_{\mathcal{X}_{app},\mathcal{Y}_{app}}(x_{test,i}) \right)^2 .$$

K Fold Cross validation is an iteration of this principle. Divide the data: $\mathcal{X}_n = (\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_K)$, and use:

$$C \approx \frac{1}{n} \sum_{k=1}^{K} \sum_{x_i \in \mathcal{X}_k} \left(f(x_i) - \hat{f}_{\mathcal{X}_{-k},\mathcal{Y}_{-k}}(x_i) \right)^2$$

Hence, each $f(x_i)$ is predicted one time, with a learning sample that does not contain it.

In our case we will take $K = 10$, and:

$$\begin{aligned} \mathcal{X}_1 &= (x_1, x_{11}, x_{21}, x_{31}, x_{41}, x_{51}, x_{61}, x_{71}, x_{81}) \\ \dots & \end{aligned}$$

As experiments are grouped, we have heterogeneity in each test sample, and reproducibility of the Cross validation.

Details of LS and GP (1/2)

We trust that $\Delta P_{fro} \propto H_f$. Hence we do both LS and GP on the pseudo-measure $\frac{\Delta P_{fro}}{H_f}$.
For evaluation of predictions we go back to the ΔP_{fro} quantity.

Least Square

- Prediction formulas with an iid model error and prior information case.
- $\beta_{prior} = (0.22, 0.21)^t$, Q_{prior} diagonal with standard-deviation at 50% of β_{prior}.
- Model error variance estimated by REML.
- We are similar to LS in prediction because the prediction only uses the calibrated code.
- We have predictive variance and hence confidence intervals.

Details of LS and GP (2/2)

Gaussian Process Modelling

- Choice of the covariance function:
  Use of the Matern stationary covariance function:
  $C_{mod}(x) = \sigma^2 \prod_{i=1}^{d} Matern(2\sqrt{\nu}\frac{x_i}{l_{c,i}})$ with:

$$Matern(x) = \frac{1}{\Gamma(\nu)2^{\nu-1}} x^{\nu} K_{\nu}(x)$$

  with K_ν the modified Bessel function of order ν. Hyper-parameters are: σ (Variance), $l_{c,1}, ..., l_{c,d}$ (correlation lengths) and ν (regularity). We enforce $\nu = \frac{3}{2}$.
- Choice of the experimental conditions:
  - Interest of division by H_f: simplification of the correlation function $\longrightarrow$ less hyper-parameters to estimate. (very few values for H_f)
  - ρ and μ are physically linked, we merge them into a pseudo-experimental condition $X_{\rho,\mu}$.
  - 4 hyper-parameters to estimate: σ, l_G, l_{D_h}, $l_{\rho,\mu}$.
- Estimation:
  - REML estimation.
  - We linearly transform all experimental condition to put them in [0, 1].

Results

With the Cross Validation we use, each experiment is predicted one time. Hence we dispose of:

- The vector of predictions $\Delta \hat{P}_{fro}^{exp}$ of size 85.
- The vector of predictive variance σ_{pred}^2 of size 85.

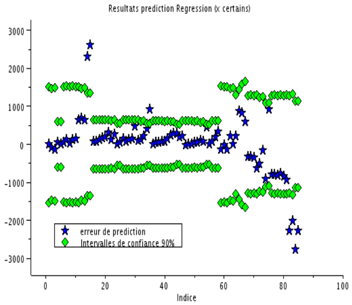
2 quantitative criteria:

- RMSE: $\sqrt{\frac{1}{85} \sum_{i=1}^{85} \left(\Delta P_{fro}^{exp} - \Delta \hat{P}_{fro}^{exp} \right)^2}$
- Confidence Intervals: $\frac{1}{85} card \left\{ i | 1 \leq i \leq 85, |\Delta \hat{P}_{fro,i}^{exp} - \Delta P_{fro,i}^{exp}| \leq 1.64 \sigma_{pred,i} \right\}$ (should be around 0.9)

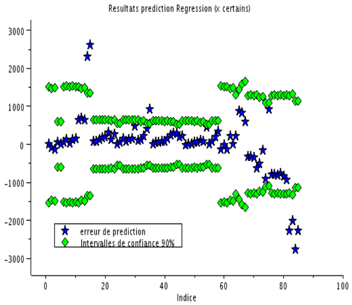
We do 2 different cases:

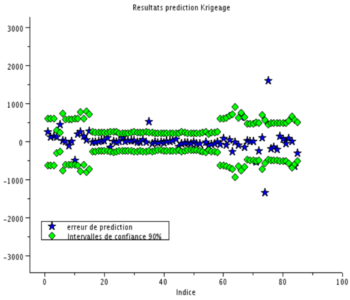
- **Case 1** We do not take measure error on H_f, G and D_h into account. We enforce $\sigma_{mes} = 200^2$. Hence $K = \sigma_{mes}^2 I_n$.
- **Case 2** We take the measure error on H_f, G and D_h into account (nominal values of the statistical test part). We enforce $\sigma_{mes}^2 = 100^2$.

Results: Case 1

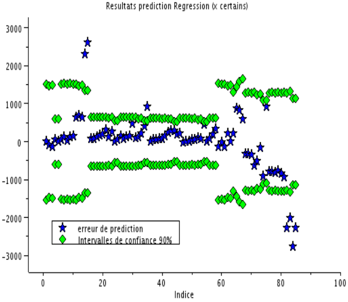


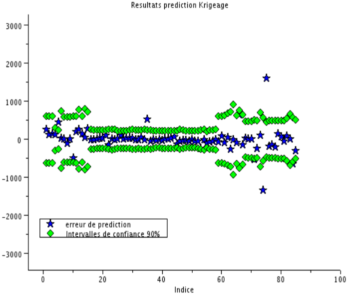
Results: Case 1



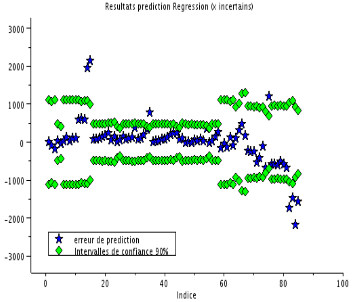


	RMSE	Confidence Intervals
LS	741.72591	0.9176471
GP	289.49389	0.9294118

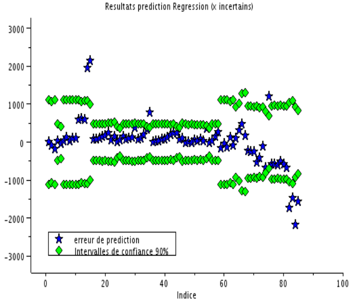


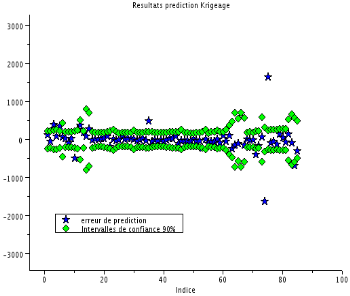


Results: Case 2



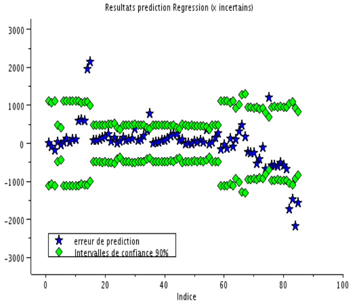
Results: Case 2

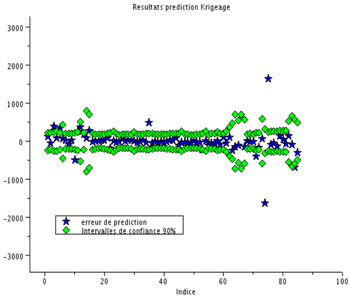




Results: Case 2

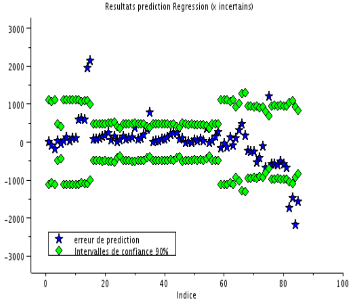
	RMSE	Confidence Intervals
LS	581.35775	0.9058824
GP	307.76398	0.8823529

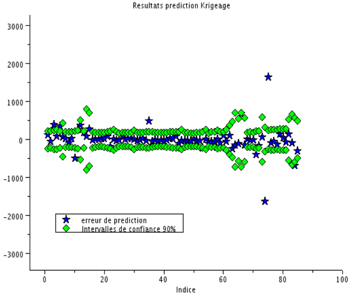




Results: Case 2

	RMSE	Confidence Intervals
LS	581.35775	0.9058824
GP	307.76398	0.8823529





LS is improved because prediction take into account the correlation between measure errors on geometric conditions.

Conclusion and prospects

Conclusion

- We can improve the prediction capability of the numerical model by completing it with a statistical model based on the observations.
- In the application case: Cross Validation estimation of the 'performances' (for both LS and GP) are accurate when we want to predict at an experimented geometry.
- The hyper-parameter estimation step is important.
- Computationally expensive when the number of experiments is large (But there exists state of the art methods).

Prospects

- Hyper-parameter estimation by Leave One Out or Cross Validation.
- Application on the Friction model in more general physical domains $\longrightarrow$ more physical models to calibrate with more experiments

Some references

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.

E. Vazquez.
Modélisation comportementale de systèmes non-linéaires multivariables par méthodes à noyaux et applications.
PhD thesis, Université Paris XI Orsay, 2005.

Thank you for your attention.