Calibration of computer experiments

François Bachoc

Université Paul Sabatier,
Institut de Mathématiques de Toulouse

(Based on joint work with Guillaume Bois, Josselin Garnier and Jean-Marc Martinez)
(With contribution from Guillaume Damblin, Merlin Keller and Guillaume Perrin)

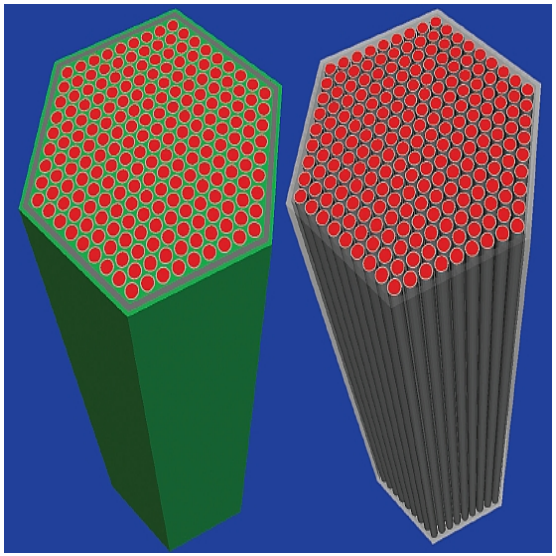
École thématique ETICS - June 2016

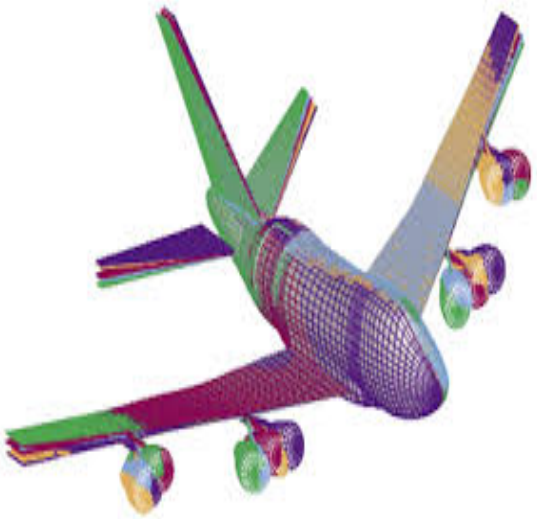
Course outline

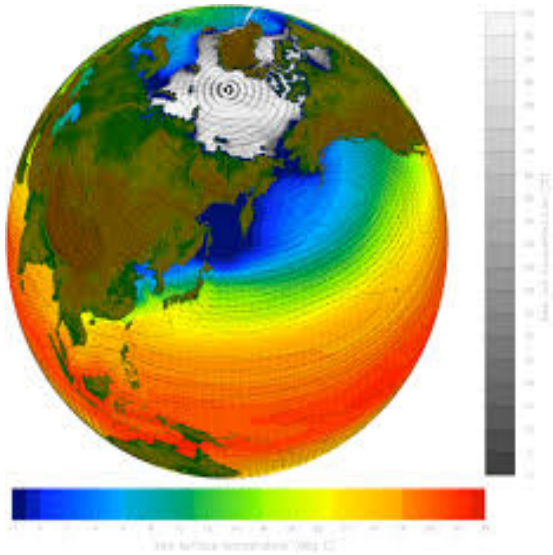
1. General introduction to calibration
2. Introduction to Gaussian processes
3. Gaussian process-based calibration and improved prediction : linear case
4. Gaussian process-based calibration and improved prediction : non-linear case
5. Examples of recent methodological developments (by Guillaume Damblin, Merlin Keller and Guillaume Perrin)

Computer models

Computer models have become essential in science and industry !







For clear reasons : cost reduction, possibility to explore hazardous or extreme scenarios...

Challenges when working with computer models

Challenges

- Computation time can be large
  - We will not try to make the computer model faster
  - We will try to spare computation time. E.g. metamodelling
- There can be numerical errors
  - We do not address this issue
  - It is the verification problem
- The code may require a complex parametrization
  - We will address this issue.
  - It is the calibration problem
  - Also related to sensitivity analysis
- We do not know if the computer model display a reliable picture of “reality”
  - We will address this issue.
  - It is the validation or improved-prediction problem

First notation

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

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

One can also obtain observations from a physical system

$$\boldsymbol{x} \in \mathbb{R}^d \quad \to \boxed{\text{physical system}} \to y \in \mathbb{R}$$

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

A computation code models (gives an approximation of) a physical phenomenon.

Calibration with variable physical system

Basic point :

- For a fixed experimental condition $\boldsymbol{x}$, there does not exist a unique "true" quantity of interest $y \in \mathbb{R}$
- Similarly there does not exist a unique "true" calibration parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^p$

One possible model :

- $\boldsymbol{\beta}$ is a random vector with distribution $\mathcal{L}_{\boldsymbol{\beta}}$
- The computer model characterizes the physical system completely, that is with

$$\boldsymbol{x} \in \mathbb{R}^d \rightarrow \boxed{\text{physical system}} \rightarrow y \in \mathbb{R}$$

we have

$$y = f(\boldsymbol{x}, \boldsymbol{\beta}) + \epsilon$$

with $\epsilon \sim \mathcal{N}(0, \sigma_m^2)$ independently of $\boldsymbol{\beta}$

See e.g.

S. Fu, Inversion probabiliste bayésienne en analyse d'incertitude, *PhD thesis, Université Paris-Sud 11, 2012.* `http://tel.archives-ouvertes.fr/tel-00766341/`.

We do not address this setting in this course

Calibration with fixed physical system

Basic point :

- For a fixed experimental condition $\boldsymbol{x}$, there exists a unique "true" quantity of interest $\phi(\boldsymbol{x}) \in \mathbb{R}$
- Similarly there exists a unique "true" calibration parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^p$

The model for Section 1

- $\boldsymbol{\beta}_0$ is fixed in $\mathbb{R}^p$
- The computer model characterizes the physical system completely, that is with

$$\boldsymbol{x} \in \mathbb{R}^d \rightarrow \boxed{\text{physical system}} \rightarrow y \in \mathbb{R}$$

we have

$$y = f(\boldsymbol{x}, \boldsymbol{\beta}_0) + \epsilon$$

with $\epsilon \sim \mathcal{N}(0, \sigma_m^2)$

$\Longrightarrow$ $f(\boldsymbol{x}, \boldsymbol{\beta}_0)$ is the true and unknown value of the physical system for the experimental condition $\boldsymbol{x}$

Bayesian framework

$$\beta_0 \sim \mathcal{N}(\beta_{prior}, \boldsymbol{Q}_{prior})$$

with

- fixed and known $\beta_{prior} \in \mathbb{R}^p$
- fixed and known positive definite $m \times m$ matrix $\boldsymbol{Q}_{prior} \in \mathbb{R}^p$

Linearization

There exists a fixed and known nominal model parameter β_{nom} so that

$$\forall \boldsymbol{x} : \; f(\boldsymbol{x}, \beta) = f(\boldsymbol{x}, \beta_{nom}) + \sum_{i=1}^{p} h_i(\boldsymbol{x})(\beta_i - \beta_{nom,i})$$

This model is used e.g. in

T. Kawano, K.M. Hanson, S. Frankle, P. Talou, M.B. Chadwick and R.C. Little, Evaluation and propagation of the ^{239}Pu Fission Cross-Section uncertainties using a Monte Carlo technique *Nuclear Science and Engineering (153) 1-7, 2006*.

A simplification

The linearization

$$\forall \boldsymbol{x} : \ f(\boldsymbol{x}, \boldsymbol{\beta}) = f(\boldsymbol{x}, \boldsymbol{\beta}_{nom}) + \sum_{i=1}^{p} h_i(\boldsymbol{x})(\boldsymbol{\beta}_i - \boldsymbol{\beta}_{nom,i})$$

can be simplified by letting

- $\forall \boldsymbol{\beta} \in \mathbb{R}^p : \bar{\boldsymbol{\beta}} = \boldsymbol{\beta} - \boldsymbol{\beta}_{nom}$
- $\bar{\boldsymbol{\beta}}_{prior} = \boldsymbol{\beta}_{prior} - \boldsymbol{\beta}_{nom}$
- $\forall \bar{\boldsymbol{\beta}} \in \mathbb{R}^p : \bar{f}(\boldsymbol{x}, \bar{\boldsymbol{\beta}}) = f(\boldsymbol{x}, \bar{\boldsymbol{\beta}}) - f(\boldsymbol{x}, \boldsymbol{\beta}_{nom})$

This gives us that

$$\bar{\boldsymbol{\beta}}_0 \sim \mathcal{N}(\bar{\boldsymbol{\beta}}_{prior}, \boldsymbol{Q}_{prior})$$

$$\forall \boldsymbol{x} : \ \bar{f}(\boldsymbol{x}, \bar{\boldsymbol{\beta}}) = \sum_{i=1}^{p} h_i(\boldsymbol{x})\bar{\boldsymbol{\beta}}_i$$

No loss of information entailed since $\boldsymbol{\beta}$ and $\bar{\boldsymbol{\beta}}$ are deterministic functions of one another :

- inference on $\bar{\boldsymbol{\beta}} \iff$ inference on $\boldsymbol{\beta}$

In the sequel, unless stated otherwise, we implicitly assume that this transformation has been made

Hence the model is

$$\boldsymbol{\beta}_0 \sim \mathcal{N}(\boldsymbol{\beta}_{prior}, \boldsymbol{Q}_{prior}) \qquad \text{and} \qquad \forall \boldsymbol{x} : \ f(\boldsymbol{x}, \boldsymbol{\beta}) = \sum_{i=1}^{p} h_i(\boldsymbol{x})\boldsymbol{\beta}_i$$

Experimental results

Recall that with

$$\boldsymbol{x} \in \mathbb{R}^d \rightarrow \boxed{\text{physical system}} \rightarrow y \in \mathbb{R}$$

we have

$$y = f(\boldsymbol{x}, \boldsymbol{\beta}_0) + \epsilon$$

Consider that n experiments are carried out at $\boldsymbol{x}_1, ..., \boldsymbol{x}_n \in D \subset \mathbb{R}^d$ with observed quantities of interests $y_1, ..., y_n$
Then we have for $i = 1, ..., n$

$$\begin{aligned} y_i &= f(\boldsymbol{x}_i, \boldsymbol{\beta}_0) + \epsilon_i \\ &= \sum_{j=1}^{n} h_j(\boldsymbol{x}_i)\beta_{0,j} + \epsilon_i, \end{aligned}$$

where $\epsilon_1, ..., \epsilon_n \sim_{iid} \mathcal{N}(0, \sigma_m^2)$.
Define

- the $n \times 1$ vector $\boldsymbol{y} = (y_1, ..., y_n)^t$
- the $n \times 1$ vector $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)^t$
- the $n \times p$ matrix $\boldsymbol{H}$ defined by $H_{i,j} = h_j(\boldsymbol{x}_i)$

Then we have

$$\boxed{\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}}$$

Posterior distribution of β_0

Recall that

$$\boldsymbol{y} = \boldsymbol{H}\beta_0 + \epsilon$$

Let us calculate the joint distribution of $(\boldsymbol{y}, \beta_0)$:

- $\mathbb{E}(\boldsymbol{y}) = \boldsymbol{H}\mathbb{E}(\beta_0) = \boldsymbol{H}\beta_{prior}$
- $cov(\boldsymbol{y}) = \boldsymbol{H}cov(\beta_0)\boldsymbol{H}^t + \sigma_m^2 \boldsymbol{I}_n = \boldsymbol{H}\boldsymbol{Q}_{prior}\boldsymbol{H}^t + \sigma_m^2 \boldsymbol{I}_n$
- $cov(\boldsymbol{y}, \beta_0) = \boldsymbol{H}cov(\beta_0, \beta_0) = \boldsymbol{H}\boldsymbol{Q}_{prior}$ so that

$$\begin{pmatrix} y \\ \beta_0 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{H}\beta_{prior} \\ \beta_{prior} \end{pmatrix}, \begin{pmatrix} \boldsymbol{H}\boldsymbol{Q}_{prior}\boldsymbol{H}^t + \sigma_m^2 \boldsymbol{I}_n & \boldsymbol{H}\boldsymbol{Q}_{prior} \\ \boldsymbol{Q}_{prior}\boldsymbol{H}^t & \boldsymbol{Q}_{prior} \end{pmatrix} \right]$$

From the general results given in Section 3, we thus have

$$\mathcal{L}(\beta_0|\boldsymbol{y}) = \mathcal{N}(\beta_{post}, \boldsymbol{Q}_{post})$$

with

- $\beta_{post} = \beta_{prior} + \frac{1}{\sigma_m^2}(\boldsymbol{Q}_{prior}^{-1} + \frac{1}{\sigma_m^2}\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t(\boldsymbol{y} - \boldsymbol{H}\beta_{prior})$
- $\boldsymbol{Q}_{post} = (\boldsymbol{Q}_{prior}^{-1} + \frac{1}{\sigma_m^2}\boldsymbol{H}^t\boldsymbol{H})^{-1}$

And (see restricted maximum likelihood in Section 3) we can estimate σ_m^2 by

$$\hat{\sigma}_m^2 = \frac{1}{n-p}\boldsymbol{y}^t(\boldsymbol{I}_n - \boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t)\boldsymbol{y}$$

Comments

$$\mathcal{L}(\beta_0|\boldsymbol{y}) = \mathcal{N}(\beta_{post}, \boldsymbol{Q}_{post})$$

with

- $\beta_{post} = \beta_{prior} + \frac{1}{\sigma_m^2}(\boldsymbol{Q}_{prior}^{-1} + \frac{1}{\sigma_m^2}\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t(\boldsymbol{y} - \boldsymbol{H}\beta_{prior})$
- $\boldsymbol{Q}_{post} = (\boldsymbol{Q}_{prior}^{-1} + \frac{1}{\sigma_m^2}\boldsymbol{H}^t\boldsymbol{H})^{-1}$

Comments

- Everything remains linear Gaussian
- Posterior uncertainty $<$ prior uncertainty
- $\beta_{post} \to_{\sigma_m^2 \to +\infty} \beta_{prior}$ and $\boldsymbol{Q}_{post} \to_{\sigma_m^2 \to +\infty} \boldsymbol{Q}_{prior}$
- If $p \leq n$ and $rank(\boldsymbol{H}^t\boldsymbol{H}) = p$ then
  $\beta_{post} = \beta_{prior} + (\sigma_m^2 Q_{prior}^{-1} + \boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t(\boldsymbol{y} - \boldsymbol{H}\beta_{prior}) \to_{\sigma_m^2 \to +0}$
  $\beta_{prior} + (\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t(\boldsymbol{y} - \boldsymbol{H}\beta_{prior}) = (\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{y}$ (least square estimator, ignoring the prior distribution). Also $\boldsymbol{Q}_{post} \to_{\sigma_m^2 \to +0} 0$
- If p is fixed and $n \to \infty$, we generally have $(\boldsymbol{H}^t\boldsymbol{H})^{-1} \to_{n \to \infty} 0$. Hence $\boldsymbol{Q}_{post} \to_{n \to \infty} 0$. Uncertainty on β_0 generally vanishes

Predicting the physical system

Let $\boldsymbol{x}_0 \in D \subset \mathbb{R}^d$ be a new experimental condition.
Let the $p \times 1$ vector $\boldsymbol{h}(\boldsymbol{x}_0) = (h_1(\boldsymbol{x}_0), ..., h_p(\boldsymbol{x}_0))^t$
Then we have

$$\mathcal{L}(f(\boldsymbol{x}_0, \beta_0)|\boldsymbol{y}) = \mathcal{L}(\boldsymbol{h}(\boldsymbol{x}_0)^t \beta_0|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{h}(\boldsymbol{x}_0)^t \beta_{post}, \boldsymbol{h}(\boldsymbol{x}_0)^t \boldsymbol{Q}_{post} \boldsymbol{h}(\boldsymbol{x}_0))$$

Let $y_0 = f(\boldsymbol{x}_0, \beta_0) + \epsilon_0$ be a new potential observation of the physical system. Then we have, since ϵ_0 is independent of $\boldsymbol{y}, \beta_0$,

$$\mathcal{L}(y_0|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{h}(\boldsymbol{x}_0)^t \beta_{post}, \boldsymbol{h}(\boldsymbol{x}_0)^t \boldsymbol{Q}_{post} \boldsymbol{h}(\boldsymbol{x}_0) + \sigma_m^2)$$

We predict future observations by the calibrated code only
Asymptotically ($n \to \infty$, p fixed), the uncertainty on a new observation is only σ_m^2 which is non-reducible

Limitations

Limitations :

- The differences between the correctly calibrated computer model and the observations are only modeled by measure errors
- This assumption can be ruled out statistically if $\hat{\sigma}^2$ is very large
- We predict the physical system by $f(x_0, \beta_{post}) \Longrightarrow$ we are restricted/sub-optimal if the computer model is inappropriate

Introducing the model error function

We would like to extend the statistical model in the following way :
with

$$\boldsymbol{x} \in \mathbb{R}^d \rightarrow \boxed{\text{physical system}} \rightarrow y \in \mathbb{R}$$

we have

$$y = \phi(\boldsymbol{x}) + \epsilon$$

where $\phi : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ is the fixed and unknown physical system function. $\epsilon \sim \mathcal{N}(0, \sigma_m^2)$ is still the measure error

We want to have the following :

$$\forall \boldsymbol{x} \in \mathbb{R}^d;\ \phi(x) = f(x, \beta_0) + e(\boldsymbol{x})$$

where

- β_0 is the fixed and unknown correct model parameter
- $e : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ is a fixed and unknown function called the model error function or model error

Identifiability issues

Recall that we want the following statistical model :

$$\forall \boldsymbol{x} \in \mathbb{R}^d;\ \phi(x) = f(x, \beta_0) + e(\boldsymbol{x}) \tag{1}$$

However, it is not clear how to define β_0 and e w.r.t ϕ in a good way. One can simultaneously change the definition of β_0 and e and satisfy (1) !
This is an important problem which is not fully solved to my knowledge
$\Longrightarrow$ Roughly speaking, this identifiability issue should impact only inference on β_0 not inference on ϕ

One way to obtain a well-defined mathematical model is the Bayesian way where :

- β_0 is a realization of a random vector (as before)
- e is a realization of a random function (of a Gaussian process)

In addition, working with Gaussian processes will enable to work with posterior distributions for the physical system ϕ

$\Longrightarrow$ Hence we now introduce Gaussian processes

Summary

- Code function

$$f(\boldsymbol{x}, \beta)$$

- Physical system

$$\boldsymbol{x} \in \mathbb{R}^d \rightarrow \boxed{\text{physical system}} \rightarrow y \in \mathbb{R}$$

- One setting that we do not address

$$y_i = f(\boldsymbol{x}_i, \beta_i) + \epsilon_i$$

- Our setting

$$y_i = f(\boldsymbol{x}_i, \beta_0) + \epsilon_i$$

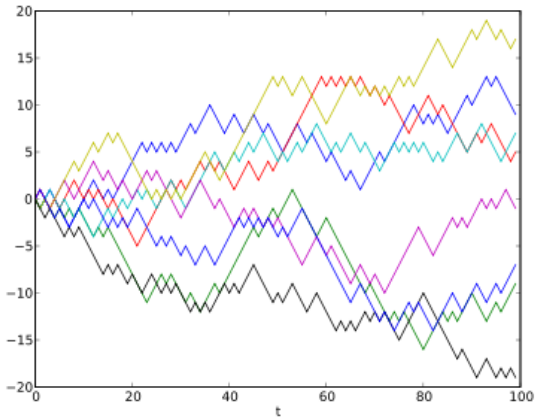
- Bayesian framework and linearization $\Longrightarrow$ calibration and prediction of physical system
  - we stay in the linear Gaussian framework
- We now want to introduce the model error function

$$y_i = f(\boldsymbol{x}_i, \beta_0) + e(\boldsymbol{x}_i) + \epsilon_i$$

  - identifiability issues
  - e as Gaussian process realization

Stochastic processes

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



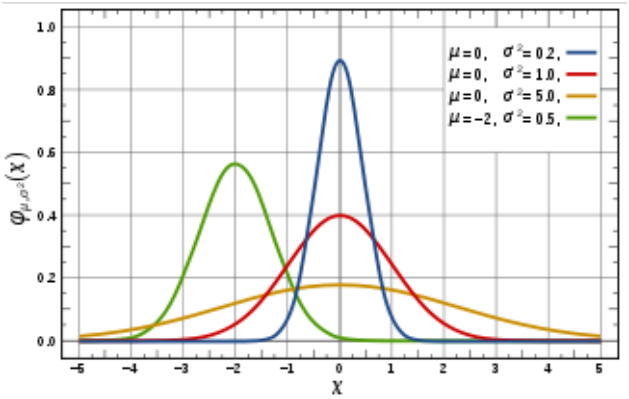
We explicit the randomness of $Z(\boldsymbol{x})$ by writing it $Z(\omega, \boldsymbol{x})$ with ω in a probability space Ω. For a given ω_0, we call the function $x \to Z(\omega_0, \boldsymbol{x})$ a realization of the stochastic process Z.

Mean function $M : \boldsymbol{x} \to M(\boldsymbol{x}) = \mathbb{E}(Z(\boldsymbol{x}))$
Covariance function $C : (\boldsymbol{x}_1, \boldsymbol{x}_2) \to C(\boldsymbol{x}_1, \boldsymbol{x}_2) = cov(Z(\boldsymbol{x}_1), Z(\boldsymbol{x}_2))$

Gaussian variables and vectors

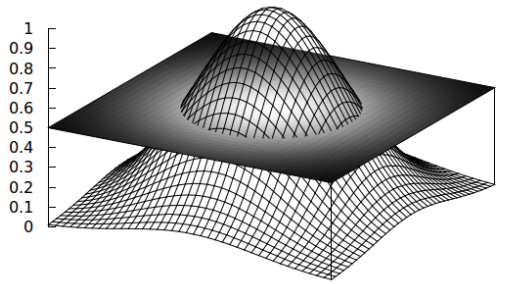
A random variable X is a Gaussian variable with mean μ and variance $\sigma^2 > 0$ when its probability density function is

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



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

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



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 $\int_{-\infty}^{+\infty}(x-\mu)^2 f_{\mu,\sigma^2}(x)dx = \sigma^2$

Let $\boldsymbol{A}$ and $\boldsymbol{m}$ be a fixed matrix and a fixed vector :

If $\boldsymbol{v}$ is a Gaussian vector, then $\boldsymbol{m} + \boldsymbol{A}\boldsymbol{v}$ is also a Gaussian vector

Gaussian processes

A stochastic process Z on $\mathbb{R}^d$ is a Gaussian process when for all $n \in \mathbb{N}$, for all $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$, the random vector $(Z(\boldsymbol{x}_1), ..., Z(\boldsymbol{x}_n))$ 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
- Conditional distributions of components of Gaussian vectors are still Gaussian
- Gaussian processes have been the most studied theoretically

The mean function

- When modeling computer experiments with Gaussian process, it is standard to have a constant or affine mean function :

$$M(\boldsymbol{x}) = a_0$$

or

$$M(\boldsymbol{x}) = a_1 x_1 + \ldots + a_d x_d$$

- In geostatistics (when modeling natural data with Gaussian processes) one may consider more complex mean functions
- In this course, we consider Gaussian processes with zero mean function :

$$\forall \boldsymbol{x} \in \mathbb{R}^d \quad M(\boldsymbol{x}) = 0$$

(Roughly speaking the computer model will play the role of a mean function)

The covariance function

The covariance function

$$C : (\boldsymbol{x}_1, \boldsymbol{x}_2) \to C(\boldsymbol{x}_1, \boldsymbol{x}_2) = cov(Z(\boldsymbol{x}_1), Z(\boldsymbol{x}_2))$$

C must me symmetric non-negative definite

$$\forall n \in \mathbb{N}, \forall \boldsymbol{x}_1, ..., \boldsymbol{x}_n \in \mathbb{R}^d, \forall \lambda_1, ..., \lambda_n \in \mathbb{R} : \quad \sum_{i,j=1}^{n} \lambda_i \lambda_j C(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0$$

$\Longrightarrow$ the covariance matrix $[C(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1,...,n}$ must be non-negative definite

Often, we require the covariance function to be positive definite :

$$\text{if } (\boldsymbol{x}_1, ..., \boldsymbol{x}_n) \text{ are 2-by-2 distinct and } (\lambda_1, ..., \lambda_n) \neq (0, ..., 0) : \quad \sum_{i,j=1}^{n} \lambda_i \lambda_j C(\boldsymbol{x}_i, \boldsymbol{x}_j) > 0$$

$\Longrightarrow$ the covariance matrix $[C(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1,...,n}$ must be positive definite
$\Longrightarrow$ No $Z(\boldsymbol{x})$ can be expressed as a linear combination of $Z(\boldsymbol{x}_1), ..., Z(\boldsymbol{x}_n)$ when $\boldsymbol{x}_1 \neq \boldsymbol{x}, ..., \boldsymbol{x}_n \neq \boldsymbol{x}$
$\Longrightarrow \approx$ the realizations of Z are sufficiently complex

Stationary covariance functions

Stationarity

The covariance function C is stationary when

$$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d : \quad C(\boldsymbol{x}_1, \boldsymbol{x}_2) = C(\boldsymbol{x}_1 - \boldsymbol{x}_2)$$

$\Longrightarrow$ The distribution of the Gaussian process with zero mean function and covariance function C is translation-invariant

Bochner's theorem

Consider a stationary function $C(\boldsymbol{x}_1, \boldsymbol{x}_2) = C(\boldsymbol{x}_1 - \boldsymbol{x}_2)$. Define the Fourier transform $\hat{C} : \mathbb{R}^d \to \mathbb{R}$ of C by

$$\hat{C}(\boldsymbol{f}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} C(\boldsymbol{t}) e^{-\mathrm{i}\boldsymbol{f}\cdot\boldsymbol{t}} d\boldsymbol{t},$$

where $\mathrm{i}^2 = -1$. Assume that

$$C(\boldsymbol{t}) = \int_{\mathbb{R}^d} \hat{C}(\boldsymbol{f}) e^{\mathrm{i}\boldsymbol{f}\cdot\boldsymbol{t}} d\boldsymbol{f}$$

In this case :

$$\boxed{C \text{ is symmetric non-negative definite} \iff \forall \boldsymbol{f} \in \mathbb{R}^d : \quad \hat{C}(\boldsymbol{f}) \geq 0}$$

$$\boxed{C \text{ is symmetric positive definite} \iff \forall \boldsymbol{f} \in \mathbb{R}^d : \quad \hat{C}(\boldsymbol{f}) > 0}$$

Proof of $\Leftarrow$ in Bochner's theorem

Let $n \in \mathbb{N}, x_1, ..., x_n \in \mathbb{R}^d, \alpha_1, ..., \alpha_n \in \mathbb{R}$:

$$\begin{aligned}
\sum_{i,j=1}^{n} \alpha_i \alpha_j C(x_i, x_j) &= \sum_{i,j=1}^{n} \alpha_i \alpha_j \int_{\mathbb{R}^d} \hat{C}(\boldsymbol{f}) e^{i\boldsymbol{f}\cdot(\boldsymbol{x}_i - \boldsymbol{x}_j)} d\boldsymbol{f} \\
&= \int_{\mathbb{R}^d} \hat{C}(\boldsymbol{f}) \left(\sum_{i,j=1}^{n} \alpha_i \alpha_j e^{i\boldsymbol{f}\cdot(\boldsymbol{x}_i - \boldsymbol{x}_j)} \right) d\boldsymbol{f} \\
&= \int_{\mathbb{R}^d} \hat{C}(\boldsymbol{f}) \left(\sum_{i,j=1}^{n} \alpha_i e^{i\boldsymbol{f}\cdot\boldsymbol{x}_i} \overline{\alpha_j e^{i\boldsymbol{f}\cdot\boldsymbol{x}_j}} \right) d\boldsymbol{f} \\
&= \int_{\mathbb{R}^d} \hat{C}(\boldsymbol{f}) \left(\sum_{i=1}^{n} \alpha_i e^{i\boldsymbol{f}\cdot\boldsymbol{x}_i} \right) \overline{\left(\sum_{j=1}^{n} \alpha_j e^{i\boldsymbol{f}\cdot\boldsymbol{x}_j} \right)} d\boldsymbol{f} \\
&= \int_{\mathbb{R}^d} \hat{C}(f) \left| \sum_{i=1}^{n} \alpha_i e^{i\boldsymbol{f}\cdot\boldsymbol{x}_i} \right|^2 d\boldsymbol{f}
\end{aligned}$$

This proves the symmetric non-negative definite part. The symmetric positive definite part is proved by remarking that for $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ 2-by2 distinct, the functions $e^{i\boldsymbol{f}\cdot\boldsymbol{x}_1}, ..., e^{i\boldsymbol{f}\cdot\boldsymbol{x}_n}$ are linearly independent

Mean square continuity

A Gaussian process Z is mean-square continuous on $\mathbb{R}^d$ is for all $\boldsymbol{x}_0 \in \mathbb{R}^d$ we have

$$\mathbb{E}([Z(\boldsymbol{x}) - Z(\boldsymbol{x}_0)]^2) \to_{\boldsymbol{x} \to \boldsymbol{x}_0} 0$$

Mean square differentiability (on $\mathbb{R}$)

A Gaussian process Z is mean-square differentiable on $\mathbb{R}$ if there exists a Gaussian process Z' on $\mathbb{R}$ so that for all $x_0 \in \mathbb{R}$ we have

$$\mathbb{E}\left(\left[\frac{Z(x_0 + h) - Z(x_0)}{h} - Z'(x_0)\right]^2\right) \to_{h \to 0, h > 0} 0$$

$\Longrightarrow$ Z' is called the derivative process of Z

Mean square differentiability of higher order (on $\mathbb{R}$)

Definition by induction : A Gaussian process Z is k times mean-square differentiable on $\mathbb{R}$ with derivative process $Z^{(k)}$ if

- Z is $k-1$ times mean-square differentiable on $\mathbb{R}$ with derivative process $Z^{(k-1)}$
- $Z^{(k-1)}$ is mean-square differentiable on $\mathbb{R}$ with derivative process $Z^{(k)}$

Relationships with covariance function regularity

Proposition (continuity)

Let Z be a Gaussian process on $\mathbb{R}^d$ with mean zero and covariance function C. Then Z is mean-square continuous on $\mathbb{R}^d$ if and only if C is continuous on $\mathbb{R}^d \times \mathbb{R}^d$

Proposition (derivability)

Let Z be a Gaussian process on $\mathbb{R}$ with mean zero and covariance function C. Then, if C is $2k$ times continuously differentiable on $\mathbb{R} \times \mathbb{R}$, Z is k times mean square differentiable on $\mathbb{R}$

Proposition (link with Fourier transforms)

Let C be a stationary covariance function with Fourier transform $\hat{C}$. Then, if

$$\int_{\mathbb{R}} \hat{C}(\boldsymbol{f})|\boldsymbol{f}|^l d\boldsymbol{f} < +\infty$$

C is l times continuously differentiable on $\mathbb{R} \times \mathbb{R}$

$$\int_{\mathbb{R}} \hat{C}(\boldsymbol{f})\boldsymbol{f}^{2k} d\boldsymbol{f} < +\infty \Longrightarrow C \text{ is } C^{2k} \Longrightarrow Z \text{ is } k \text{ times mean square differentiable}$$

Matérn model

A covariance function C on $\mathbb{R}$ is Matérn with covariace parameters $\phi > 0$, $\nu > 0$ and $\alpha > 0$ if C is stationary and its Fourier transform is

$$\hat{C}(f) = \phi \frac{1}{(\alpha^2 + f^2)^{\frac{1}{2}+\nu}}$$

- $\hat{C}(f) > 0 \Longrightarrow C$ is symmetric positive definite
- ν : smoothness parameter.
  $\nu > k \Longleftrightarrow \int_{\mathbb{R}} \hat{C}(f) f^{2k} df < +\infty \Longrightarrow C$ is $2k$ times continuously differentiable $\Longrightarrow Z$ is k times mean square differentiable.

Parameterization of the Matérn model

Alternative parameterization by $\sigma^2 > 0$, $\ell > 0$, $\nu > 0$:

$$C(x) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{2\sqrt{\nu}x}{\ell}\right)^{\nu} K_\nu \left(\frac{2\sqrt{\nu}x}{\ell}\right)$$

Interpretation of the parameters

- $\sigma^2 = K(0)$ is the variance $\rightarrow$ order of magnitude of the Gaussian Process
- ℓ is the correlation length $\rightarrow$ scale of variation of the Gaussian Process
- ν is the smoothness parameter $\rightarrow$ smoothness of the Gaussian Process

Particular cases

- $\nu = \frac{1}{2}$: exponential covariance function

$$C(x) = \sigma^2 e^{-\sqrt{2}\frac{|x|}{\ell}}$$

$\Longrightarrow$ mean square continuous

- $\nu = \frac{3}{2}$: Matérn $3/2$ covariance function

$$C(x) = \sigma^2 \left(1 + \sqrt{6}\frac{|x|}{\ell}\right) e^{-\sqrt{6}\frac{|x|}{\ell}}$$

$\Longrightarrow$ mean square differentiable

- $\nu = \frac{5}{2}$: Matérn $5/2$ covariance function

$$C(x) = \sigma^2 \left(1 + \sqrt{10}\frac{|x|}{\ell} + \frac{10}{3}\frac{|x|^2}{\ell^2}\right) e^{-\sqrt{10}\frac{|x|}{\ell}}$$

$\Longrightarrow$ twice mean square differentiable

- $\nu \to +\infty$: Gaussian covariance function

$$C(x) = \sigma^2 e^{-\frac{x^2}{\ell^2}}$$

$\Longrightarrow$ infinitely mean square differentiable

Plot of some Matérn covariance functions on $\mathbb{R}$

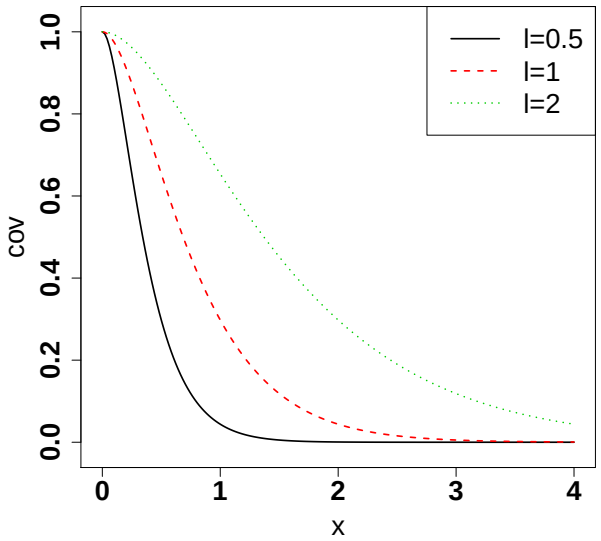
The Matérn 3/2 covariance function

$$C_{\sigma^2,\ell}(x_1, x_2) = \sigma^2 \left(1 + \sqrt{6}\frac{|x_1 - x_2|}{\ell}\right) e^{-\sqrt{6}\frac{|x_1 - x_2|}{\ell}}$$

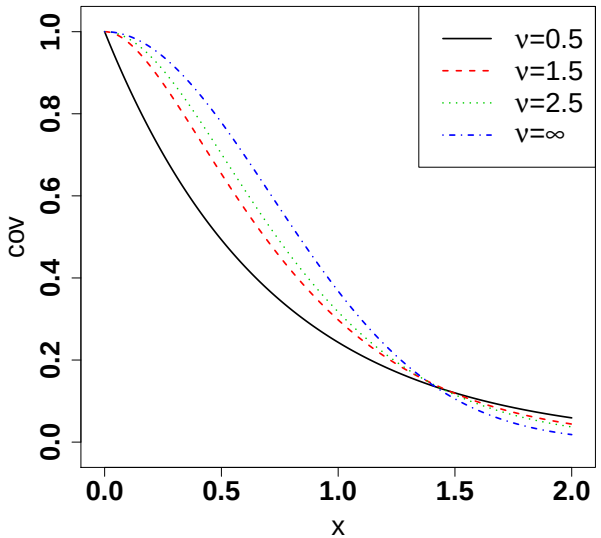
$\Longrightarrow$ two times differentiable at 0 but not three times :

$$(1 + |t|)e^{-|t|} = 1 - \frac{t^2}{2} + \frac{|t|^3}{3} + O(t^4)$$

$\Longrightarrow$ decreases with $|t|$



Plot of the Matérn covariance function with $\sigma^2 = 1$, $\ell = 1$ and $\nu = \frac{1}{2}$, $\nu = \frac{3}{2}$, $\nu = \frac{5}{2}$ and $\nu = \infty$.
$\Longrightarrow$ The value of ℓ has the same interpretation regardless of ν



Impact of σ^2 on Gaussian process realizations

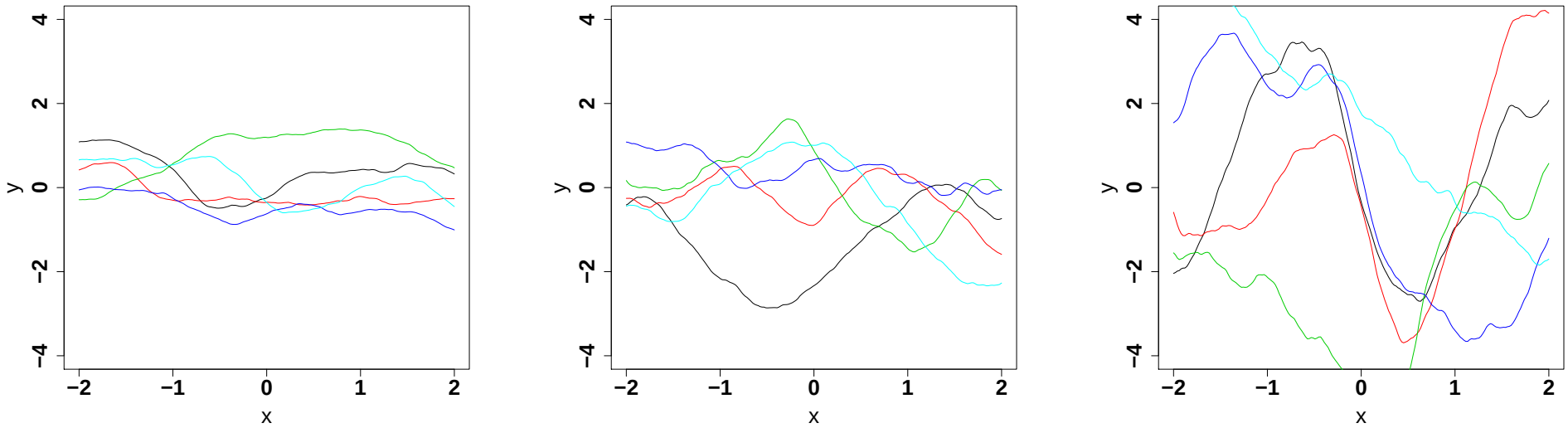


FIGURE: Influence of the variance parameter σ^2 for the Matérn covariance function. Plot of trajectories of Gaussian processes with the Matérn covariance function with correlation length $\ell = 1$, smoothness parameter $\nu = \frac{3}{2}$ and variance $\sigma^2 = \frac{1}{2}, 1, 2$ from left to right.

Impact of ℓ on Gaussian process realizations

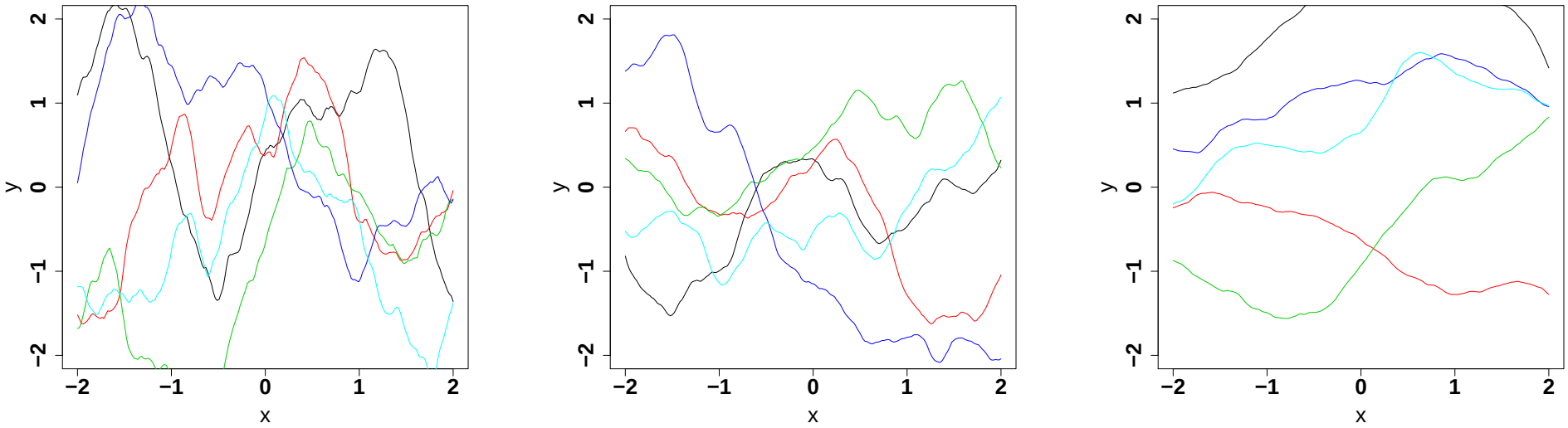


FIGURE: Influence of the correlation length ℓ for the Matérn covariance function. Plot of trajectories of Gaussian processes with the Matérn covariance function with variance $\sigma^2 = 1$, smoothness parameter $\nu = \frac{3}{2}$ and correlation length $\ell = \frac{1}{2}, 1, 2$ from left to right.

Impact of ν on Gaussian process realizations

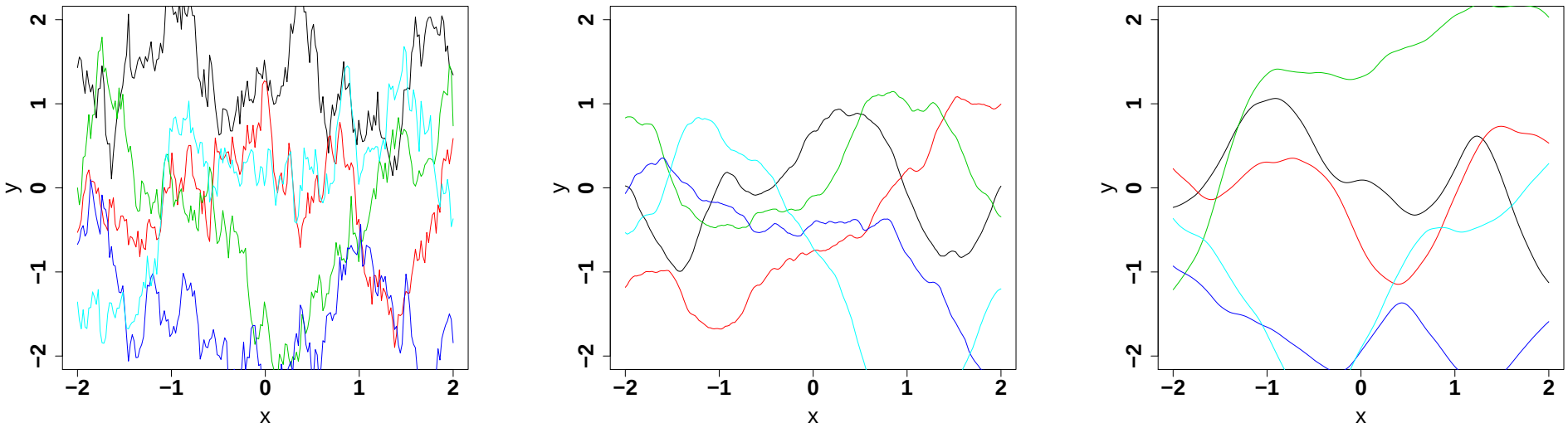


FIGURE: Influence of the smoothness parameter ν for the Matérn covariance function. Plot of trajectories of Gaussian processes with the Matérn covariance function with variance $\sigma^2 = 1$, correlation length $\ell = 1$ and smoothness parameter $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$ from left to right.

Geometric anisotropic Matérn covariance function

Parameterized by $\sigma^2 > 0$, $\ell_1 > 0$, ...,$\ell_d > 0$, $\nu > 0$
Defined by, with

$$|\boldsymbol{x}|_\ell = \sqrt{\sum_{i=1}^{d} \frac{x_i^2}{\ell_i^2}},$$

and with $C_{1,\nu}$ the Matérn covariance function in dimension one with $\sigma^2 = 1$ and $\ell = 1$,

$$C(\boldsymbol{x}) = \sigma^2 C_{1,\nu}(|x|_\ell)$$

(symmetric positive definite, see Stein 99, Rasmussen 06)

Tensor product Matérn covariance function

Parameterized by $\sigma^2 > 0$, $\ell_1 > 0$, ...,$\ell_d > 0$, $\nu > 0$
Defined by

$$C(\boldsymbol{x}) = \sigma^2 \prod_{i=1}^{d} C_{1,\nu}\left(\frac{x_i}{\ell_i}\right)$$

$\rightarrow \ell_i$ is the i-th correlation length and is the scale of variation corresponding to x_i
$\rightarrow \ell_i$ very large $\Longleftrightarrow$ Gaussian process realizations practically do not depend on x_i

Remark on tensor product covariance functions

General properties :

- If $C_1, ..., C_d$ are symmetric non-negative definite functions on $\mathbb{R} \times \mathbb{R}$, then

$$C(\boldsymbol{x}) = C_1(x_1)...C_d(x_d)$$

  is a symmetric non-negative definite functions on $\mathbb{R} \times \mathbb{R}$

- If furthermore $C_1, ..., C_d$ are stationary covariance functions on $\mathbb{R}$, with Fourier transforms $\hat{C}_1, ..., \hat{C}_d$, and if for $i = 1, ..., n$

$$C_i(t) = \int_{\mathbb{R}} \hat{C}_i(f) e^{\mathrm{i}ft} df$$

  then we have

$$\hat{C}(\boldsymbol{f}) = \hat{C}_1(f_1), ..., \hat{C}_d(f_d)$$

  and

$$C(\boldsymbol{t}) = \int_{\mathbb{R}^d} \hat{C}(\boldsymbol{f}) e^{\mathrm{i}\boldsymbol{f}\dot{\boldsymbol{t}}} d\boldsymbol{f}$$

  $\Longrightarrow$ The tensor product Matérn covariance function is symmetric positive definite
  $\Longrightarrow$ Tensorization always enables to build multidimensional covariance functions from monodimensional ones

Main tool for inference on Gaussian processes : Gaussian conditioning theorem

Theorem

Let $(\boldsymbol{Y}_1, \boldsymbol{Y}_2)^t$ be a $(n_1 + n_2) \times 1$ Gaussian vector with mean vector $(\boldsymbol{m}_1^t, \boldsymbol{m}_2)^t$ and covariance matrix

$$\begin{pmatrix} \boldsymbol{R}_1 & \boldsymbol{R}_{1,2} \\ \boldsymbol{R}_{1,2}^t & \boldsymbol{R}_2 \end{pmatrix}$$

Then, conditionally on $\boldsymbol{Y}_1 = \boldsymbol{y}_1$, $\boldsymbol{Y}_2$ is a Gaussian vector with mean

$$\mathbb{E}(\boldsymbol{Y}_2 | \boldsymbol{Y}_1 = \boldsymbol{y}_1) = \boldsymbol{m}_2 + \boldsymbol{R}_{1,2}^t \boldsymbol{R}_1^{-1} (\boldsymbol{y}_1 - \boldsymbol{m}_1)$$

and variance

$$var(\boldsymbol{Y}_2 | \boldsymbol{Y}_1 = \boldsymbol{y}_1) = \boldsymbol{R}_2 - \boldsymbol{R}_{1,2}^t \boldsymbol{R}_1^{-1} \boldsymbol{R}_{1,2}$$

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

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

Summary

- Gaussian processes are a model for random functions
- Characterized by mean and covariance functions
- Covariance function must be symmetric non-negative definite (many possibilities on $\mathbb{R}^d \times \mathbb{R}^d$)
- Matérn covariance function can carry information on
  - smoothness
  - scale of variations
  - order of magnitude
- For inference on Gaussian processes : Gaussian conditioning theorem
  - conditional distributions are Gaussian
  - explicit matricial formulas

The stochastic model

With, for $i = 1, ..., n$

$$\boldsymbol{x}_i \in \mathbb{R}^d \rightarrow \boxed{\text{physical system}} \rightarrow y_i \in \mathbb{R}$$

our stochastic model is

$$y_i = \sum_{j=1}^{p} h_j(\boldsymbol{x}_i)\beta_{0,j} + Z(\boldsymbol{x}_i) + \epsilon_i$$

where

- $\boldsymbol{\beta}_0 = (\beta_{0,1}, ..., \beta_{0,p})^t \sim \mathbb{N}(\boldsymbol{\beta}_{prior}, \boldsymbol{Q}_{prior})$, with $\boldsymbol{\beta}_{prior}$ and $\boldsymbol{Q}_{prior}$ known
- $\boldsymbol{x}_i \in D$ is fixed and observed
- The functions $h_1, ..., h_p : D \rightarrow \mathbb{R}$ can be evaluated (partial derivatives of code function f)
- $\phi(\boldsymbol{x}) = \sum_{j=1}^{p} h_j(\boldsymbol{x})\beta_{0,j} + Z(\boldsymbol{x})$ is the physical system function (code+model error)
- Z is a centered Gaussian process with unknown covariance function C
- $\epsilon_1, ..., \epsilon_n \sim_{iid} \mathcal{N}(0, \sigma_m^2)$
- y_i is thus an observed Gaussian variable

Comments :

- The randomness of $\epsilon_1, ..., \epsilon_n$ is “frequentist”
- The randomness of β_0 and Z is “Bayesian”
- The model is identifiable when $h_1, ..., h_p$ are linearly independent

Matrix notations

Let

- the $n \times 1$ vector $\boldsymbol{y} = (y_1, ..., y_n)^t$
- the $n \times 1$ vector $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)^t$
- the $n \times 1$ vector $\boldsymbol{z} = (Z(\boldsymbol{x}_1), ..., Z(\boldsymbol{x}_n))^t$
- the $n \times p$ matrix $\boldsymbol{H}$ defined by $H_{i,j} = h_j(\boldsymbol{x}_i)$

Then we have

$$\boxed{\boldsymbol{y} = \boldsymbol{H}\beta_0 + \boldsymbol{z} + \boldsymbol{\epsilon}}$$

Let

- $\boldsymbol{\Sigma}$ be the $n \times n$ matrix defined by $\Sigma_{i,j} = cov(z_i, z_j) = C(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Hence $\boldsymbol{\Sigma} = cov(\boldsymbol{z})$
- $\boldsymbol{R}$ be the $n \times n$ matrix defined by $\boldsymbol{R} = \boldsymbol{\Sigma} + \sigma_m^2 \boldsymbol{I}_n$. Hence $\boldsymbol{R} = cov(\boldsymbol{z} + \boldsymbol{\epsilon})$

Then we have

$$\boxed{\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{H}\beta_{prior}, \boldsymbol{H}\boldsymbol{Q}_{prior}\boldsymbol{H}^t + \boldsymbol{R}\right)}$$

Estimation of C and σ_m^2 : parametrization

To estimate $\boldsymbol{R}$ we need to estimate C and σ_m^2 :

- We select C in the parametric set $\{\sigma^2 C_{\boldsymbol{\theta}}, \sigma^2 \in S \subset (0, \infty), \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q\}$ (e.g. Matérn model)
- σ_m^2 is selected in the set S_m

Let for $(\sigma^2, \boldsymbol{\theta}, \sigma_m^2) \in S \times \Theta \times S_m$

$$\boldsymbol{R}_{\sigma^2, \boldsymbol{\theta}, \sigma_m^2} = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\theta}} + \sigma_m^2 \boldsymbol{I}_n$$

with $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ the $n \times n$ matrix defined by $\Sigma_{\boldsymbol{\theta},i,j} = C_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{x}_j)$

Remark : If the measure error variance σ_m^2 is known (expert knowledge), the following remains valid with $S_m = \{\sigma_m^2\}$

Estimation of C and σ_m^2 : restricted maximum likelihood

Under covariance parameters $(\sigma^2, \boldsymbol{\theta}, \sigma_m^2)$ we have :

$$\boxed{\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{H}\beta_{prior}, \boldsymbol{H}\boldsymbol{Q}_{prior}\boldsymbol{H}^t + \boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\right)}$$

Let

- $\boldsymbol{W}$ be a $(n-p) \times n$ matrix, with full rank, so that $\boldsymbol{W}\boldsymbol{H} = 0$
- $\boldsymbol{w} = \boldsymbol{W}\boldsymbol{y}$

Then $\boldsymbol{w}$ is a Gaussian vector and we have

$$\mathbb{E}(\boldsymbol{w}) = \boldsymbol{W}\boldsymbol{H}\beta_{prior} = 0$$

and

$$cov(\boldsymbol{w}) = \boldsymbol{W}\boldsymbol{H}\boldsymbol{Q}_{prior}\boldsymbol{H}^t\boldsymbol{W}^t + \boldsymbol{W}\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{W}^t = \boldsymbol{W}\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{W}^t$$

Then restricted maximum likelihood on $\boldsymbol{y}$ is maximum likelihood on $\boldsymbol{w}$:

$$\boxed{(\hat{\sigma}^2, \hat{\boldsymbol{\theta}}, \hat{\sigma}_m^2) \in \underset{\sigma^2,\boldsymbol{\theta},\sigma_m^2 \in S \times \Theta \times S_m}{\operatorname{argmin}} \ln(|\boldsymbol{W}\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{W}^t|) + \boldsymbol{w}^t(\boldsymbol{W}\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{W}^t)^{-1}\boldsymbol{w}}$$

$\Longrightarrow$ We are not impacted by β_{prior} and $\boldsymbol{Q}_{prior}$

Proposition (Harville, 74)

i) Changing $\boldsymbol{W}$ only multiplies the restricted likelihood by a constant not depending on $\sigma^2, \boldsymbol{\theta}, \sigma_m^2$

ii) For $\boldsymbol{W}$ so that $\boldsymbol{W}\boldsymbol{W}^t = \boldsymbol{I}_{n-m}$ and $\boldsymbol{W}^t\boldsymbol{W} = \boldsymbol{I}_n - \boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t$ we have

$$\begin{aligned}&\ln(|\boldsymbol{W}\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{W}^t|) + \boldsymbol{w}^t(\boldsymbol{W}\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{W}^t)^{-1}\boldsymbol{w}\\ &= -\ln|\boldsymbol{H}^t\boldsymbol{H}| + \ln|\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}| + \ln|\boldsymbol{H}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{H}| + \boldsymbol{y}^t\boldsymbol{\Pi}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y},\end{aligned}$$

with

$$\boldsymbol{\Pi}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2} = \boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2} - \boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}$$

iii) Let $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^t$ with $n \times p$ matrix $\boldsymbol{U}$ so that $\boldsymbol{U}^t\boldsymbol{U} = \boldsymbol{I}_{p,p}$, $p \times p$ diagonal matrix $\boldsymbol{S}$ with non-negative components, and orthogonal $p \times p$ matrix $\boldsymbol{V}$. Then

$$\begin{aligned}(\hat{\sigma}^2, \hat{\boldsymbol{\theta}}, \hat{\sigma}_m^2) \quad \in \quad & \underset{\sigma^2,\boldsymbol{\theta},\sigma_m^2 \in S\times\Theta\times S_m}{\operatorname{argmin}} \ln\left|\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U}\right| + \ln\left|\boldsymbol{R}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\right| + \boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y}\\ & -\boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U}(\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U})^{-1}\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y}\end{aligned}$$

- The estimator $(\hat{\sigma}^2, \hat{\boldsymbol{\theta}}, \hat{\sigma}_m^2)$ is not impacted by the choice of $\boldsymbol{W}$
- We can avoid $n \times n$ matrix products. The condition number of $\boldsymbol{H}$ is not impacting the estimation

Proof of i)

Let $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ so that, for $i = 1, 2$ $\boldsymbol{W}_i$ is $(n-p) \times n$, with full rank, so that $\boldsymbol{W}_i \boldsymbol{H} = 0$.

Let E be the linear space orthogonal to the column space of $\boldsymbol{H}$. Then $E \subset \mathbb{R}^n$ has dimension $n-p$. Let P_E be the orthogonal projection from $\mathbb{R}^n$ to E. Let, for $i = 1, 2$, S_i be the linear mapping from E to $\mathbb{R}^p$ defined by $S_i \boldsymbol{v} = \boldsymbol{W}_i^t \boldsymbol{v}$. Then S_i is invertible with inverse $S_i^{-1} : \mathbb{R}^p \to E$.

We have

$$\begin{aligned} \boldsymbol{W}_2^t \boldsymbol{y} &= \boldsymbol{W}_2^t (P_E \boldsymbol{y}) \\ &= S_2 P_E \boldsymbol{y} \\ &= S_2 S_1^{-1} S_1 P_E \boldsymbol{y} \\ &= S_2 S_1^{-1} \boldsymbol{W}_1^t (P_E \boldsymbol{y}) \\ &= S_2 S_1^{-1} (\boldsymbol{W}_1^t \boldsymbol{y}) \end{aligned}$$

Hence, there exists an invertible $p \times p$ matrix $\boldsymbol{F}$ so that $\boldsymbol{W}_2^t \boldsymbol{y} = \boldsymbol{F} \boldsymbol{W}_1^t \boldsymbol{y}$. Hence, with $g_{\sigma^2, \boldsymbol{\theta}, \sigma_m^2}$ the pdf of $\boldsymbol{W}_1^t \boldsymbol{y}$,

$$\mathbb{E}(f(\boldsymbol{W}_2^t \boldsymbol{y})) = \mathbb{E}(f(\boldsymbol{F} \boldsymbol{W}_1^t \boldsymbol{y})) = \int_{\mathbb{R}^p} f(\boldsymbol{F}\boldsymbol{z}) g_{\sigma^2, \boldsymbol{\theta}, \sigma_m^2}(\boldsymbol{z}) d\boldsymbol{z} = \int_{\mathbb{R}^p} f(\boldsymbol{z}) g_{\sigma^2, \boldsymbol{\theta}, \sigma_m^2}(\boldsymbol{F}^{-1}\boldsymbol{z}) |\boldsymbol{F}^{-1}| d\boldsymbol{z}$$

so that

$$\text{(likelihood of } \boldsymbol{W}_2^t \boldsymbol{y}) = g_{\sigma^2, \boldsymbol{\theta}, \sigma_m^2}(\boldsymbol{F}^{-1} \boldsymbol{W}_2^t \boldsymbol{y}) |\boldsymbol{F}^{-1}| = g_{\sigma^2, \boldsymbol{\theta}, \sigma_m^2}(\boldsymbol{W}_1^t \boldsymbol{y}) |\boldsymbol{F}^{-1}| = |\boldsymbol{F}^{-1}| \text{(likelihood of } \boldsymbol{W}_1^t \boldsymbol{y})$$

Proof of ii) and iii)

For ii), see Harville, 74

For iii)

$$
\begin{aligned}
& -\ln|\boldsymbol{H}^t\boldsymbol{H}| + \ln|\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}| + \ln|\boldsymbol{H}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{H}| + \boldsymbol{y}^t\boldsymbol{\Pi}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y} \\
= \quad & -\ln|\boldsymbol{H}^t\boldsymbol{H}| + \ln|\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}| + \ln|\boldsymbol{H}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{H}| + \boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y} \\
& -\boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{H}(\boldsymbol{H}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y} \\
= \quad & -\ln|\boldsymbol{H}^t\boldsymbol{H}| + \ln|\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}| + \ln|\boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^t| + \boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y} \\
& -\boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^t(\boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^t)^{-1}\boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y} \\
= \quad & -\ln|\boldsymbol{H}^t\boldsymbol{H}| + \ln|\boldsymbol{R}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}| + 2\ln|\boldsymbol{V}\boldsymbol{S}| + \ln|\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U}| + \boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y} \\
& -\boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^t(\boldsymbol{S}\boldsymbol{V}^t)^{-1}(\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U})^{-1}(\boldsymbol{V}\boldsymbol{S})^{-1}\boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma^2,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y} \\
= \quad & -\ln|\boldsymbol{H}^t\boldsymbol{H}| + 2\ln|\boldsymbol{V}\boldsymbol{S}| + \ln\left|\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U}\right| + \ln\left|\boldsymbol{R}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\right| + \boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y} \\
& -\boldsymbol{y}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U}(\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{U})^{-1}\boldsymbol{U}^t\boldsymbol{R}^{-1}_{\sigma,\boldsymbol{\theta},\sigma_m^2}\boldsymbol{y}
\end{aligned}
$$

After estimation : plug-in

Plug-in

- C is estimated by $C_{\hat{\sigma}^2,\hat{\boldsymbol{\theta}}}$
- σ_m^2 is estimated by $\hat{\sigma}_m^2$
- We make the approximation that $C = C_{\hat{\sigma}^2,\hat{\boldsymbol{\theta}}}$ and that $\sigma_m^2 = \hat{\sigma}_m^2$ to compute all subsequent conditional distributions

$\Longrightarrow$ convenient simplification
$\Longrightarrow$ one alternative : Bayesian framework on $\sigma^2, \boldsymbol{\theta}, \sigma_m^2$

In the rest of this section we consider that C and σ_m^2 are known

Proposition

We have

$$\mathcal{L}(\beta_0|\boldsymbol{y}) = \mathcal{N}(\beta_{post}, \boldsymbol{Q}_{post})$$

with

- $\beta_{post} = \beta_{prior} + (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}^t\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{H}\beta_{prior})$
- $\boldsymbol{Q}_{post} = (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}^t\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}$

Comments

- Everything remains linear Gaussian
- Posterior uncertainty $<$ prior uncertainty
- $\beta_{post} \rightarrow_{\boldsymbol{R}^{-1}\rightarrow 0} \beta_{prior}$ and $\boldsymbol{Q}_{post} \rightarrow_{\boldsymbol{R}^{-1}\rightarrow 0} \boldsymbol{Q}_{prior}$
- If p is fixed and $n \rightarrow \infty$, we do not necessarily have $\boldsymbol{Q}_{post} \rightarrow_{n\rightarrow\infty} 0$
  $\Longrightarrow$ Properties of fixed domain asymptotics (book, Stein 99)
  $\Longrightarrow$ Related to the identifiability issues for the realizations of Z and β_0
  $\Longrightarrow$ cf exercise session

Proof (1/2)

The Gaussian pdf with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$ is of the form

$$\ln(p(\boldsymbol{z})) = Cst. - \frac{1}{2}\boldsymbol{z}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{z} + \boldsymbol{z}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{m},$$

where $Cst.$ does not depend on $\boldsymbol{z}$.

We have, where $Cst.$ does not depend on β

$$\begin{aligned}
\ln(p_{\beta_0|\boldsymbol{y}}(\beta|\boldsymbol{y})) &= \ln\left(\frac{p_{\boldsymbol{y}|\beta_0}(\boldsymbol{y}|\beta)p_{\beta_0}(\beta)}{p_{\boldsymbol{y}}(\boldsymbol{y})}\right) \\
&= Cst. - \frac{1}{2}(\beta-\beta_{prior})^t\boldsymbol{Q}_{prior}^{-1}(\beta-\beta_{prior}) - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{H}\beta)^t\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{H}\beta) \\
&= Cst. - \frac{1}{2}\beta^t\boldsymbol{Q}_{prior}^{-1}\beta + \beta^t\boldsymbol{Q}_{prior}^{-1}\beta_{prior} - \frac{1}{2}\beta^t\boldsymbol{H}^t\boldsymbol{R}^{-1}\boldsymbol{H}\beta + \beta^t\boldsymbol{H}^t\boldsymbol{R}^{-1}\boldsymbol{y} \\
&= Cst. - \frac{1}{2}\beta^t(\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}^t\boldsymbol{R}^{-1}\boldsymbol{H})\beta + \beta^t(\boldsymbol{Q}_{prior}^{-1}\beta_{prior} + \boldsymbol{H}^t\boldsymbol{R}^{-1}\boldsymbol{y})
\end{aligned}$$

Hence,

$$\boldsymbol{Q}_{post} = (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}^t\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}$$

and

$$\beta_{post} = (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}^t\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}(\boldsymbol{Q}_{prior}^{-1}\beta_{prior} + \boldsymbol{H}^t\boldsymbol{R}^{-1}\boldsymbol{y})$$

Proof (2/2)

$$
\begin{aligned}
\beta_{post} &= (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}^t \boldsymbol{R}^{-1} \boldsymbol{H})^{-1} (\boldsymbol{Q}_{prior}^{-1} \beta_{prior} + \boldsymbol{H}^t \boldsymbol{R}^{-1} \boldsymbol{y}) \\
&= (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}^t \boldsymbol{R}^{-1} \boldsymbol{H})^{-1} (\boldsymbol{Q}_{prior}^{-1} \beta_{prior} + \boldsymbol{H}^t \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{H}\beta_{prior}) + \boldsymbol{H}^t \boldsymbol{R}^{-1} \boldsymbol{H} \beta_{prior}) \\
&= \beta_{prior} + (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}^t \boldsymbol{R}^{-1} \boldsymbol{H})^{-1} \boldsymbol{H}^t \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{H}\beta_{prior})
\end{aligned}
$$

Illustration of calibration (1/3)

- Observation of the physical system : for $i = 1, ..., n$ $y_i = x_i^2 + \epsilon_i$. $\epsilon_1, ..., \epsilon_n$ are *iid* $\mathcal{N}(0, \sigma_m^2)$ with $\sigma_m^2 = 0$ (without measure errors) or $\sigma_m^2 = 0.1^2$ (with measure errors)
- Computer model : $f(x, \boldsymbol{\beta}) = \beta_1 + \beta_2 x$
- Model error as a realization of a Gaussian process with covariance function : $C(x, y) = \sigma^2 \exp\left(-\frac{|x-y|^2}{\ell^2}\right)$. $\sigma = 0.3$, $\ell = 0.5$ (known)
- $$\boldsymbol{\beta}_{prior} = \begin{pmatrix} 0.2 \\ 1 \end{pmatrix}, Q_{prior} = \begin{pmatrix} 0.09 & 0 \\ 0 & 0.09 \end{pmatrix}$$
- Observations : $x_1 = 0.2$, $x_2 = 0.4$, $x_3 = 0.6$ and $x_4 = 0.8$ on $D = [0, 1]$

Illustration of calibration (2/3) (without measure errors)

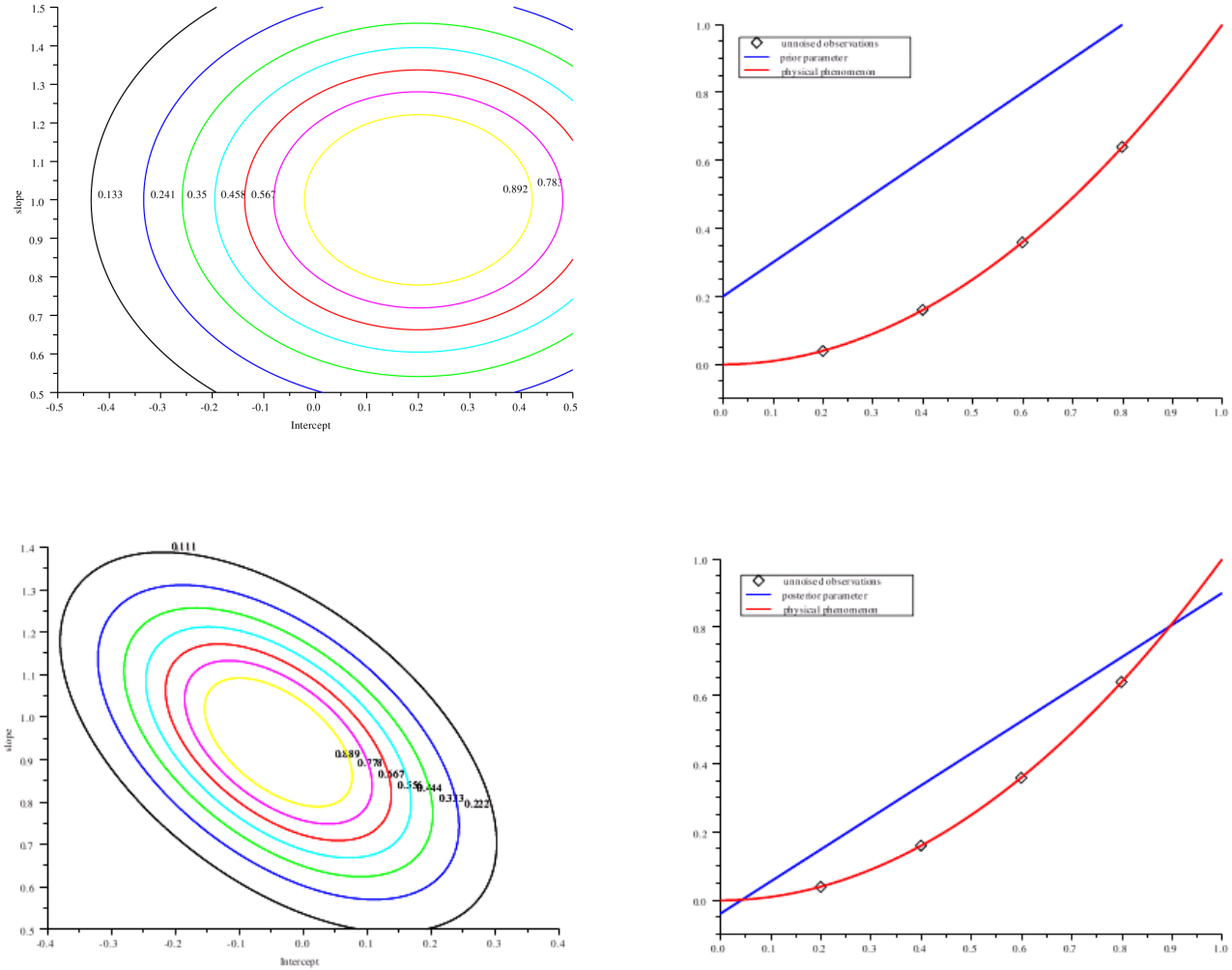


FIGURE: Up-left : prior probability density function of the parameter β_0. Down-left : posterior probability density function of the parameter β_0. Right : plot of the code response corresponding to prior and posterior mean of the code parameter.

Illustration of calibration (3/3) (with measure errors)

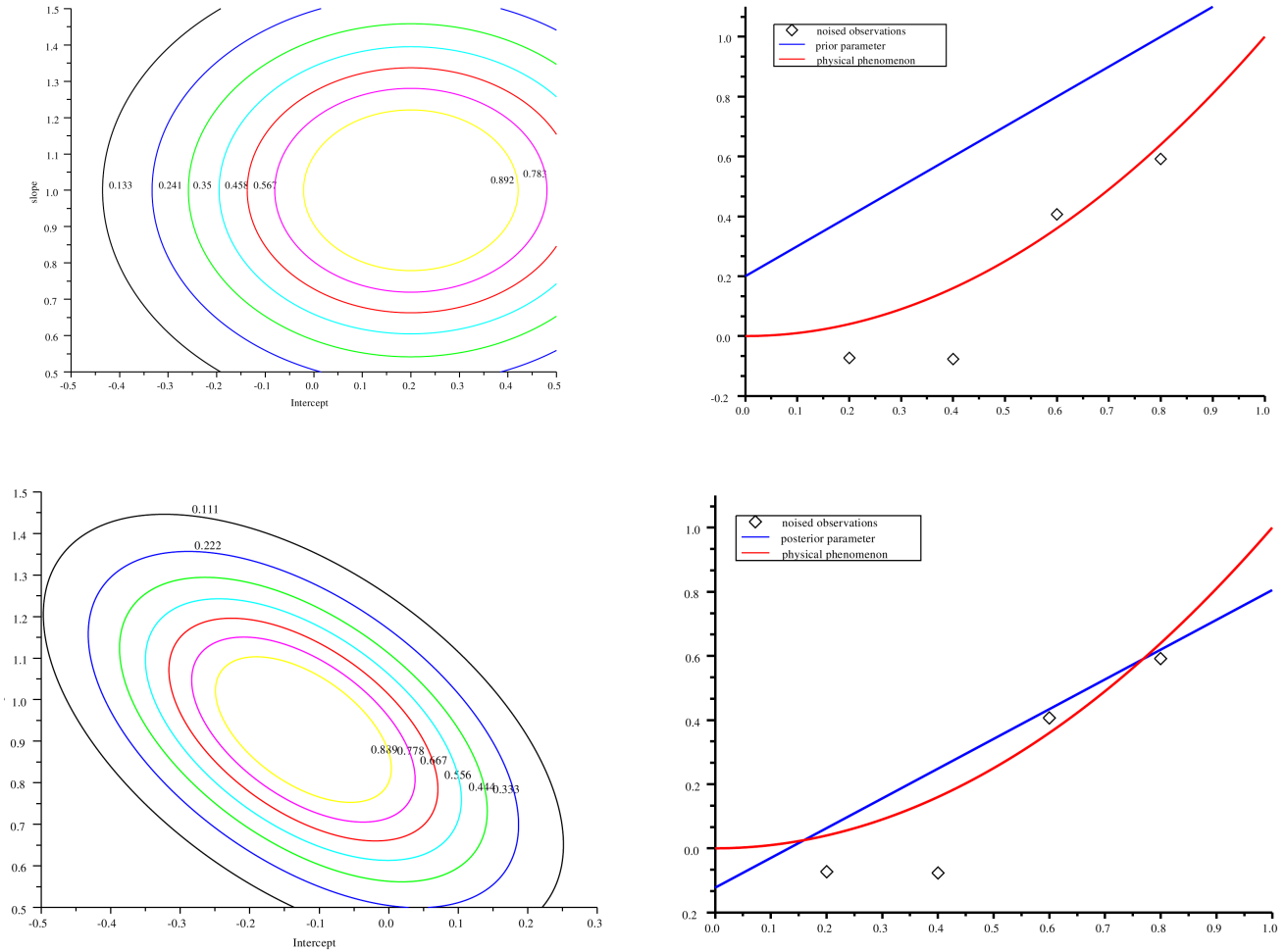


FIGURE: Up-left : Prior probability density function of the parameter β_0. Down-left : Posterior probability density function of the parameter β_0. Right : plot of the code response corresponding to prior and posterior mean of the code parameter.

Prediction (1/2)

Goal : To compute $\mathcal{L}(\phi(\boldsymbol{x}_0)|\boldsymbol{y})$ at a new experimental condition x_0

Notations

- y_0 : potential observation at $\boldsymbol{x}_0$. $y_0 = \phi(\boldsymbol{x}_0) + \epsilon_0$ with $\epsilon_0 \sim \mathcal{N}(0, \sigma_m^2)$ (independent of the rest)
- $\boldsymbol{h}(\boldsymbol{x}_0)$: $p \times 1$ vector defined by $[\boldsymbol{h}(\boldsymbol{x}_0)]_j = h_j(\boldsymbol{x}_0)$
- $\boldsymbol{r}(\boldsymbol{x}_0)$: $n \times 1$ vector defined by $[\boldsymbol{r}(\boldsymbol{x}_0)]_i = cov(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_0)) = C(\boldsymbol{x}_i, \boldsymbol{x}_0)$

Prediction (2/2)

Proposition

$$\mathcal{L}(\phi(\boldsymbol{x}_0)|\boldsymbol{y}) = \mathcal{N}(\hat{\phi}(\boldsymbol{x}_0), \hat{\sigma}^2(x_0))$$

$$\mathcal{L}(y_0|\boldsymbol{y}) = \mathcal{N}(\hat{\phi}(\boldsymbol{x}_0), \hat{\sigma}^2(x_0) + \sigma_m^2)$$

with

- $\hat{\phi}(\boldsymbol{x}_0) = (\boldsymbol{h}(\boldsymbol{x}_0))^t \beta_{post} + (\boldsymbol{r}(\boldsymbol{x}_0))^t \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{H}\beta_{post})$
- $\hat{\sigma}^2(\boldsymbol{x}_0) = C(\boldsymbol{x}_0, \boldsymbol{x}_0) - \boldsymbol{r}(\boldsymbol{x}_0)^t \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}_0) + (\boldsymbol{h}(\boldsymbol{x}_0) - \boldsymbol{H}^t \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}_0))^t (\boldsymbol{H}^t \boldsymbol{R}^{-1} \boldsymbol{H} + Q_{prior}^{-1})^{-1} (\boldsymbol{h}(\boldsymbol{x}_0) - \boldsymbol{H}^t \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}_0))$

- 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.
- $\mathcal{L}(y_0|\boldsymbol{y})$ can be used in cross validation of the stochastic model

Proof (1/2)

Law of total expectation :

$$\mathbb{E}(\phi(\boldsymbol{x}_0)|\boldsymbol{y}) \quad = \quad \mathbb{E}[\mathbb{E}(\phi(\boldsymbol{x}_0)|\boldsymbol{y},\boldsymbol{\beta}_0)|\boldsymbol{y}]$$

Then, since

$$p(\phi(\boldsymbol{x}_0)|\boldsymbol{y},\boldsymbol{\beta}_0) = \frac{p(\phi(\boldsymbol{x}_0),\boldsymbol{y},\boldsymbol{\beta}_0)}{p(\boldsymbol{y},\boldsymbol{\beta}_0)} = \frac{p(\phi(\boldsymbol{x}_0),\boldsymbol{y},\boldsymbol{\beta}_0)p(\boldsymbol{\beta}_0)}{p(\boldsymbol{y},\boldsymbol{\beta}_0)p(\boldsymbol{\beta}_0)} = \frac{p(\phi(\boldsymbol{x}_0),\boldsymbol{y}|\boldsymbol{\beta}_0)}{p(\boldsymbol{y}|\boldsymbol{\beta}_0)} = p_{|\boldsymbol{\beta}_0}(\phi(\boldsymbol{x}_0)|\boldsymbol{y}),$$

we have

$$\mathbb{E}(\phi(\boldsymbol{x}_0)|\boldsymbol{y},\boldsymbol{\beta}_0) \quad = \quad \mathbb{E}_{|\boldsymbol{\beta}_0}(\phi(\boldsymbol{x}_0)|\boldsymbol{y})$$

Now, conditionally to $\boldsymbol{\beta}_0$,

$$\begin{pmatrix} y_0 \\ \boldsymbol{y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{h}(\boldsymbol{x}_0)^t\boldsymbol{\beta}_0 \\ \boldsymbol{H}\boldsymbol{\beta}_0 \end{pmatrix}, \begin{pmatrix} C(\boldsymbol{x}_0,\boldsymbol{x}_0) & \boldsymbol{r}(\boldsymbol{x}_0)^t \\ \boldsymbol{r}(\boldsymbol{x}_0) & \boldsymbol{R} \end{pmatrix}\right)$$

Hence, from the Gaussian conditioning theorem

$$\mathbb{E}_{|\boldsymbol{\beta}_0}(\phi(\boldsymbol{x}_0)|\boldsymbol{y}) = \boldsymbol{h}(\boldsymbol{x}_0)^t\boldsymbol{\beta}_0 + \boldsymbol{r}(\boldsymbol{x}_0)^t\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{\beta}_0).$$

So

$$\begin{aligned} \mathbb{E}(\phi(\boldsymbol{x}_0)|\boldsymbol{y}) \quad &= \quad \mathbb{E}[\boldsymbol{h}(\boldsymbol{x}_0)^t\boldsymbol{\beta}_0 + \boldsymbol{r}(\boldsymbol{x}_0)^t\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{\beta}_0)|\boldsymbol{y}] \\ &= \quad (\boldsymbol{h}(\boldsymbol{x}_0))^t\boldsymbol{\beta}_{post} + (\boldsymbol{r}(\boldsymbol{x}_0))^t\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{\beta}_{post}) \end{aligned}$$

Proof (2/2)

Law of total variance :

$$var(\phi(\boldsymbol{x}_0)|\boldsymbol{y}) = \mathbb{E}[var(\phi(\boldsymbol{x}_0)|\boldsymbol{y}, \beta_0)|\boldsymbol{y}] + var[\mathbb{E}(\phi(\boldsymbol{x}_0)|\boldsymbol{y}, \beta_0)|\boldsymbol{y}]$$

We have, from the previous Gaussian conditioning theorem,

$$\begin{aligned} \mathbb{E}[var(\phi(\boldsymbol{x}_0)|\boldsymbol{y}, \beta_0)|\boldsymbol{y}] &= \mathbb{E}[C(\boldsymbol{x}_0, \boldsymbol{x}_0) - \boldsymbol{r}(\boldsymbol{x}_0)^t \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}_0)|\boldsymbol{y}] \\ &= C(\boldsymbol{x}_0, \boldsymbol{x}_0) - \boldsymbol{r}(\boldsymbol{x}_0)^t \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}_0) \end{aligned}$$

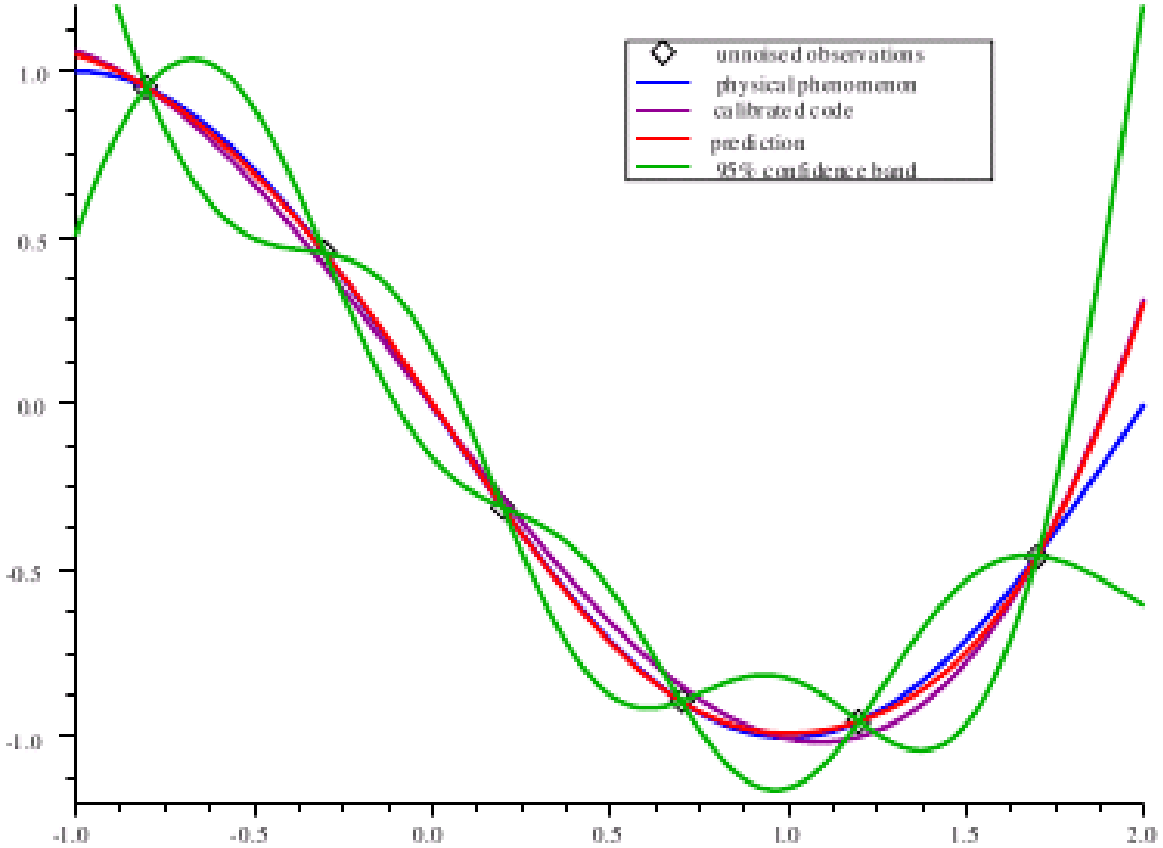
and

$$\begin{aligned} var[\mathbb{E}(\phi(\boldsymbol{x}_0)|\boldsymbol{y}, \beta_0)|\boldsymbol{y}] &= var[\boldsymbol{h}(\boldsymbol{x}_0)^t \beta_0 + \boldsymbol{r}(\boldsymbol{x}_0)^t \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{H}\beta_0)|\boldsymbol{y}] \\ &= var[(\boldsymbol{h}(\boldsymbol{x}_0)^t - \boldsymbol{r}(\boldsymbol{x}_0)^t \boldsymbol{R}^{-1} \boldsymbol{H})\beta_0|\boldsymbol{y}] \\ &= var[(\boldsymbol{h}(\boldsymbol{x}_0) - \boldsymbol{H}^t \boldsymbol{R}^{-1} r(\boldsymbol{x}_0))^t \beta_0|\boldsymbol{y}] \\ &= (\boldsymbol{h}(\boldsymbol{x}_0) - \boldsymbol{H}^t \boldsymbol{R}^{-1} r(\boldsymbol{x}_0))^t \boldsymbol{Q}_{post} (\boldsymbol{h}(\boldsymbol{x}_0) - \boldsymbol{H}^t \boldsymbol{R}^{-1} r(\boldsymbol{x}_0)) \\ &= (\boldsymbol{h}(\boldsymbol{x}_0) - \boldsymbol{H}^t \boldsymbol{R}^{-1} r(\boldsymbol{x}_0))^t (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}^t \boldsymbol{R}^{-1} \boldsymbol{H})^{-1} (\boldsymbol{h}(\boldsymbol{x}_0) - \boldsymbol{H}^t \boldsymbol{R}^{-1} r(\boldsymbol{x}_0)) \end{aligned}$$

Illustration of prediction (1/3)

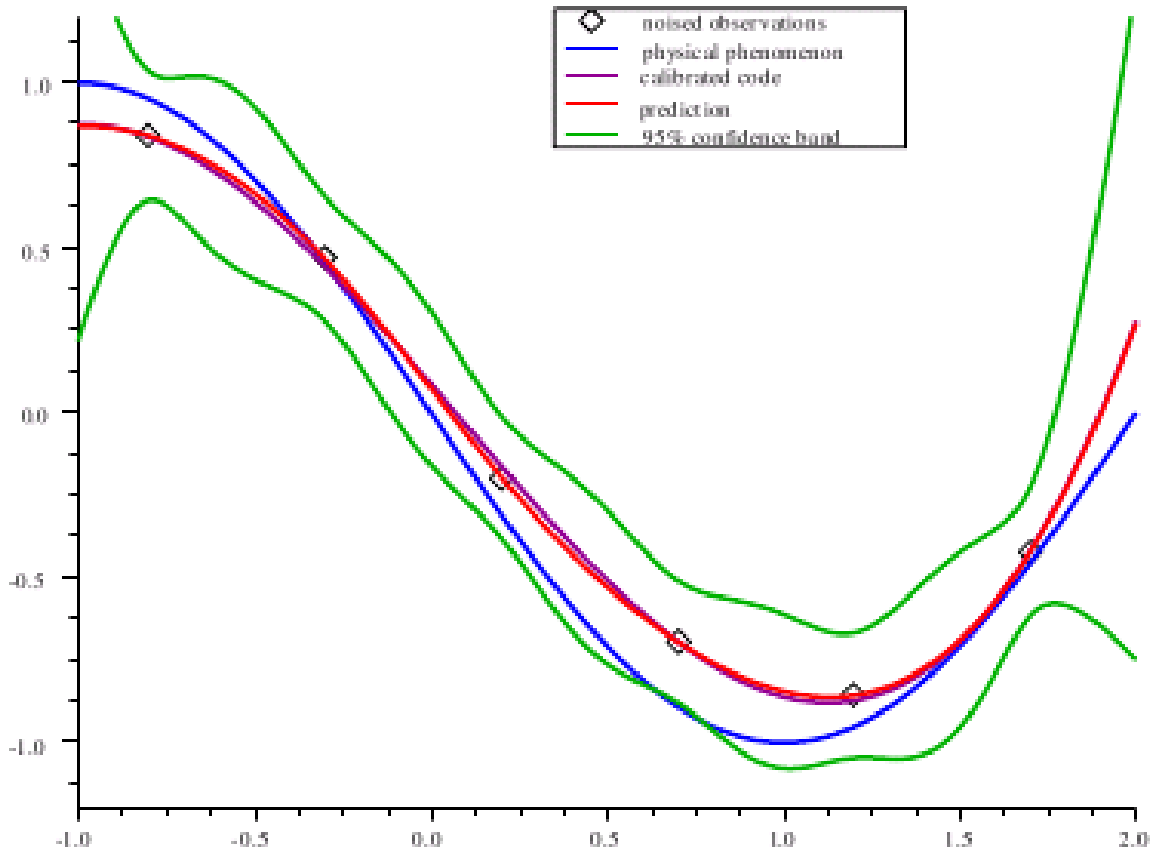
- Observation of the physical system : for $i = 1, ..., n$, $y_i = -sin(\frac{\pi x}{2}) + \epsilon_i$ $\epsilon_1, ..., \epsilon_n$ are iid $\mathcal{N}(0, \sigma_m^2)$ with $\sigma_m^2 = 0$ (without measure errors) or $\sigma_m^2 = 0.1^2$ (with measure errors)
- Computer model : $f(x, \beta) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3$
- Model error as a realization of a Gaussian process with covariance function : $C(x, y) = \sigma^2 \exp\left(-\frac{|x-y|^2}{\ell^2}\right)$. $\sigma = 0.3$, $\ell = 0.5$ (known)
- $\boldsymbol{Q}_{prior}$ very large
- $D = [-1, 2]$: 6 observations regularly sampled between -0.8 and 1.7

Illustration of prediction (2/3) (without measure errors)



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

Illustration of prediction (3/3) (with measure errors)



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

Case study : FLICA IV

We now illustrate the implementation of this Gaussian process model with the FLICA IV case study

Based on

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

FLICA IV : experimental results

$$\boldsymbol{x} \in \mathbb{R}^d \quad \rightarrow \boxed{\text{physical system}} \rightarrow y \in \mathbb{R}$$

The experiment/physical system

- Pressurized and possibly heated water flowing through a cylinder
- We measure the pressure drop between the two ends of the cylinder
- y : the part of the pressure drop due to friction

Two kinds of experimental conditions

x is partitioned into

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

Experimental results

There are $n = 253$ experimental results $\boldsymbol{x}_1, y_1, ..., \boldsymbol{x}_n, y_n$

- Only 8 configurations of system parameters are explored ! (8 campaigns where a system is built and tested under different environments)

FLICA IV : computer model

FLICA IV modelling of the experiment is based on the local relations

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

$$f_{iso} = \begin{cases} \frac{a_l}{Re} & \text{if } Re < Re_l \\ \frac{a_t}{Re^{b_t}} & \text{if } Re_t < Re \\ \frac{a_l}{Re}\frac{Re_t - Re}{Re_t - Re_l} + \frac{a_t}{Re^{b_t}}\frac{Re - Re_l}{Re_t - Re_l} & \text{if } Re_l < Re < Re_t \end{cases}$$

and

$$f_h = 1 - \frac{P_h}{P_w}\frac{C_f(T_w - T_b)}{1 + d\left(\frac{T_w + T_b}{2T_0}\right)^n},$$

where a_l, a_t, b_t, C_f, n, d are the model parameters

- The local relations above are integrated numerically
- We focus on the parameters a_t, b_t and fix the parameters a_l, C_f, n, d to nominal values. Hence $\beta_0 = (a_t, b_t)^t$. We let $\beta_{prior} = (0.22, 0.21)^t$ and $\boldsymbol{Q}_{prior} = diag(0.11^2, 0.105^2)$
- When $f_h = 1$: Isothermal regime. Else anisothermal regime. Among the 253 experiments, 115 are in the isothermal regime
- Run time $\approx$ 1 min

FLICA IV : covariance functions

We experimented the following settings for the choice of C and σ_m^2

- Let $\boldsymbol{x} = (\boldsymbol{x}_s, \boldsymbol{x}_e)$ (decomposition between system parameters and environment variables). Then
  - either $C(\boldsymbol{x}_1, \boldsymbol{x}_2) = \bar{C}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ (dependence between different system parameters)
  - or $C(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathbf{1}\{\boldsymbol{x}_{s,1} = \boldsymbol{x}_{s,2}\}\bar{C}(\boldsymbol{x}_{e,1}, \boldsymbol{x}_{e,2})$ (independence between different system parameters)
- Exponential, Matérn 3/2, Matérn 5/2 and Gaussian for $\bar{C}$
- σ_m^2 either provided by experimentalists or estimated by restricted maximum likelihood

$\Longrightarrow$ the prediction differences were small
$\Longrightarrow$ In the sequel we present the results with $C(\boldsymbol{x}_1, \boldsymbol{x}_2) = \bar{C}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and σ_m^2 provided by experimentalists

FLICA IV : cross validation

10-fold cross validation :

- We partition the 253 experimental results into 10 subsets.
- For each y_i, we compute $\tilde{\phi}(\boldsymbol{x}_i)$ and $\tilde{\sigma}^2(\boldsymbol{x}_i) + \sigma_m^2$ as $\hat{\phi}(\boldsymbol{x}_i)$ and $\hat{\sigma}^2(\boldsymbol{x}_i) + \sigma_m^2$ but when only the 9 subsets not containing y_i are used
- We re-estimate the β_0, C and σ_m^2 for each left out subset (10 estimations)
- In each subset all of the 8 system parameter configurations are present
  $\Longrightarrow$ We predict for known system parameters

Prediction criteria :

- root mean square error (RMSE) ; should be minimal

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{\phi}(\boldsymbol{x}_i))^2$$

- 90% confidence intervals ; should be close to 0.9

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{|y_i - \tilde{\phi}(\boldsymbol{x}_i)| \leq 1.645\sqrt{\tilde{\sigma}^2(\boldsymbol{x}_i) + \sigma_m^2}\right\}$$

FLICA IV : calibration results

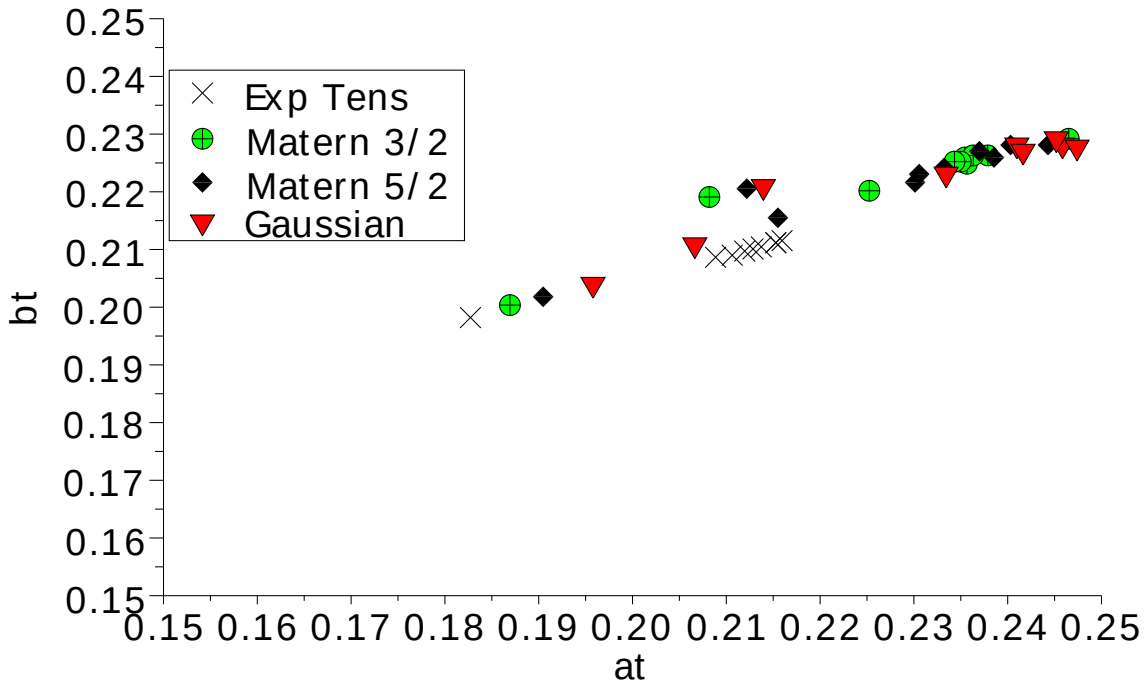


FIGURE: For each choice of covariance function for C : the 10 values of $\beta_{post} = (a_{t,post}, b_{t,post})^t$ for the 10 left out subsets.

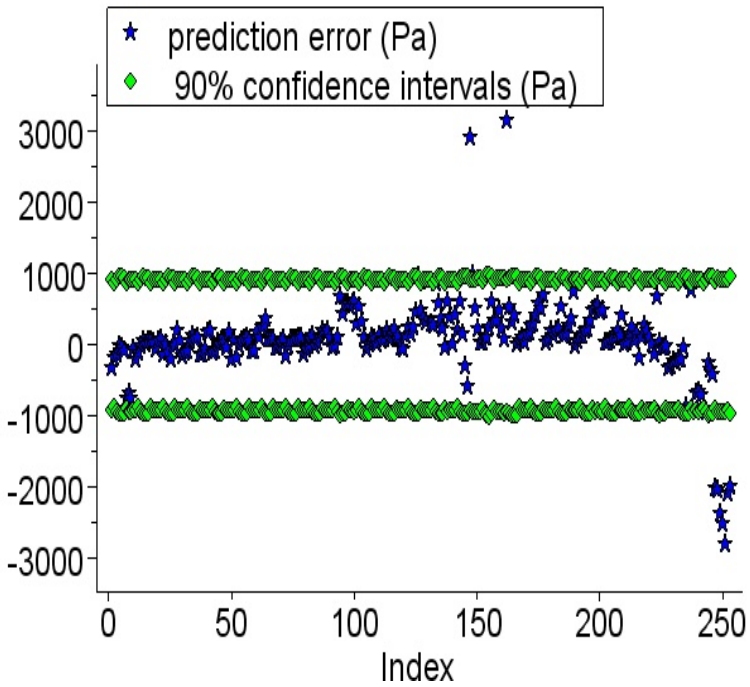
The empirical correlation between a_t and b_t comes from the relation $a_t R_e^{-b_t}$

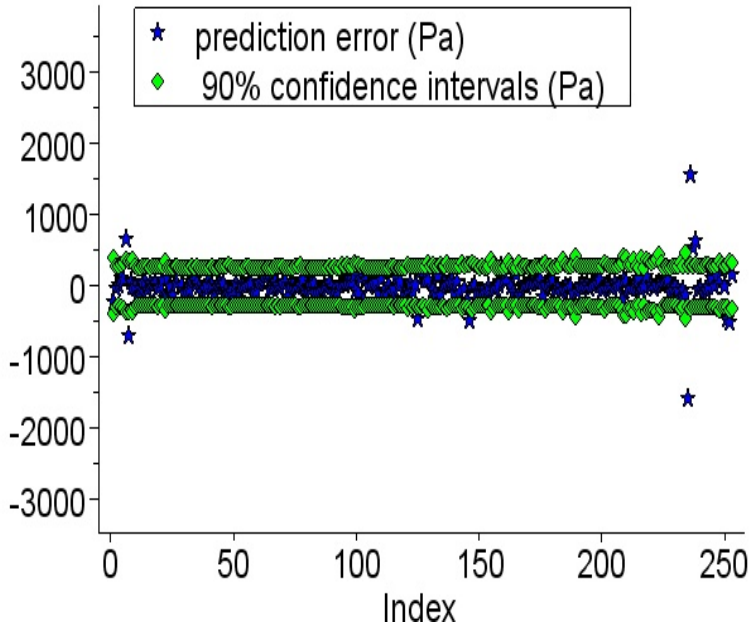
FLICA IV : prediction results

Prediction results with 10-fold cross validation of the 253 experimental results :

	RMSE	90% Confidence Intervals
Calibrated code	$567Pa$	$241/253 \approx 0.95$
Gaussian Processes	$196Pa$	$241/253 \approx 0.95$

Left : calibrated code. Right : Gaussian processes





Summary

- Our setting
$$y_i = f(\boldsymbol{x}_i, \beta_0) + Z(\boldsymbol{x}_i) + \epsilon_i$$
- Physical system (quantity of interest)
$$\phi(\boldsymbol{x}_i) = f(\boldsymbol{x}_i, \beta_0) + Z(\boldsymbol{x}_i)$$
- model error Z : Gaussian process
- Bayesian framework for β_0
- Linearization of $f(\boldsymbol{x}, \beta_0) \Longrightarrow$ calibration and prediction
  - we stay in the linear Gaussian framework
  - explicit matrix vector formulas
- identifiability issue solved in the stochastic framework (different code functions $\Longrightarrow$ different distributions of the Gaussian process Z)
- identifiability issue is likely to remain in practice

$\Longrightarrow$ we now investigate the limitations of the linear approximation of the code

Notations

The stochastic model is, for $i = 1, ..., n$

$$y_i = f(\boldsymbol{x}_i, \beta_0) + Z(\boldsymbol{x}_i) + \epsilon_i$$

Let for $\beta \in \mathbb{R}^p$

- $\boldsymbol{f}_\beta$ be the $n \times 1$ vector defined by
$$[\boldsymbol{f}_\beta]_i = f(\boldsymbol{x}_i, \beta)$$
- $\boldsymbol{H}_\beta$ be the $n \times p$ matrix defined by
$$[\boldsymbol{H}_\beta]_{i,j} = \frac{\partial f(\boldsymbol{x}_i, \beta)}{\partial \beta_j}$$

Then, a linear approximation around $\beta_{nom} \in \mathbb{R}^p$ yields

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{f}_{\beta_0} + \boldsymbol{z} + \boldsymbol{\epsilon} \\ &\approx \boldsymbol{f}_{\beta_{nom}} + \boldsymbol{H}_{\beta_{nom}}(\beta_0 - \beta_{nom}) + \boldsymbol{z} + \boldsymbol{\epsilon} \end{aligned}$$

$\Longrightarrow$ this is our model from Section 3 when $\beta_{nom} = 0$ and $\boldsymbol{f}_{\beta_{nom}} = 0$. $\Longrightarrow$ We can always recover this setting upon changing the definition of β and $\boldsymbol{y}$

Calibration

From the calibration proposition in Section 3, if the relation

$$\boldsymbol{y} = \boldsymbol{f}_{\beta_{nom}} + \boldsymbol{H}_{\beta_{nom}}(\beta_0 - \beta_{nom}) + \boldsymbol{z} + \epsilon$$

is exact, then we have

$$\boxed{\mathbb{E}(\beta_0|\boldsymbol{y}) = \beta_{prior} + (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}_{\beta_{nom}}^t \boldsymbol{R}^{-1} \boldsymbol{H}_{\beta_{nom}})^{-1} \boldsymbol{H}_{\beta_{nom}}^t \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{f}_{\beta_{nom}} - \boldsymbol{H}_{\beta_{nom}}(\beta_{prior} - \beta_{nom}))}$$

and

$$\boxed{cov(\beta_0|\boldsymbol{y}) = (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}_{\beta_{nom}}^t \boldsymbol{R}^{-1} \boldsymbol{H}_{\beta_{nom}})}$$

Proof : We have

$$\boldsymbol{y} - \boldsymbol{f}_{\beta_{nom}} = \boldsymbol{H}_{\beta_{nom}}(\beta_0 - \beta_{nom}) + \boldsymbol{z} + \epsilon$$

and $\beta_0 - \beta_{nom} \sim \mathcal{N}(\beta_{prior} - \beta_{nom}, \boldsymbol{Q}_{prior})$, so we can apply the proposition in Section 3 which gives

$$\mathbb{E}(\beta_0 - \beta_{nom}|\boldsymbol{y} - \boldsymbol{f}_{\beta_{nom}}) =$$

$$\beta_{prior} - \beta_{nom} + (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}_{\beta_{nom}}^t \boldsymbol{R}^{-1} \boldsymbol{H}_{\beta_{nom}})^{-1} \boldsymbol{H}_{\beta_{nom}}^t \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{f}_{\beta_{nom}} - \boldsymbol{H}_{\beta_{nom}}(\beta_{prior} - \beta_{nom}))$$

and

$$var(\beta_0 - \beta_{nom}|\boldsymbol{y} - \boldsymbol{f}_{\beta_{nom}}) = (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}_{\beta_{nom}}^t \boldsymbol{R}^{-1} \boldsymbol{H}_{\beta_{nom}})$$

Algorithm

Start with $i = 1$ and $\boldsymbol{\beta}_{nom,1} = \boldsymbol{\beta}_{nom}$
Until $\boldsymbol{\beta}_{nom,i+1} \approx \boldsymbol{\beta}_{nom,i}$ (or after a computation budget is reached) do

1. Compute $\boldsymbol{f}_{\boldsymbol{\beta}_{nom,i}}$ and $\boldsymbol{H}_{\boldsymbol{\beta}_{nom,i}}$ (calls to the computer model)
2. Let

$$\boldsymbol{\beta}_{nom,i+1} = \boldsymbol{\beta}_{prior} + (\boldsymbol{Q}_{prior}^{-1} + \boldsymbol{H}_{\boldsymbol{\beta}_{nom,i}}^{t} \boldsymbol{R}^{-1} \boldsymbol{H}_{\boldsymbol{\beta}_{nom,i}})^{-1} \boldsymbol{H}_{\boldsymbol{\beta}_{nom,i}}^{t} \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{f}_{\boldsymbol{\beta}_{nom,i}} - \boldsymbol{H}_{\boldsymbol{\beta}_{nom,i}} (\boldsymbol{\beta}_{prior} - \boldsymbol{\beta}_{nom,i}))$$

3. Let $i = i + 1$

At the end of the algorithm, one can use $\boldsymbol{\beta}_{nom} = \boldsymbol{\beta}_{nom,i}$ for the linear approximation, compute $\boldsymbol{f}_{\boldsymbol{\beta}_{nom,i}}$ and $\boldsymbol{H}_{\boldsymbol{\beta}_{nom,i}}$, and use the calibration and prediction formulas of Section 3
$\Longrightarrow$ We aim at choosing the linearization point at the maximum a posteriori, so that the linear approximation is most accurate where the conditional distribution concentrates

Prediction : general formulas

Let for $\boldsymbol{x}_0 \in D$, $\boldsymbol{h}_{\beta_{nom}}(\boldsymbol{x}_0)$ be the $p \times 1$ vector defined by

$$[\boldsymbol{h}_{\beta_{nom}}(\boldsymbol{x}_0)]_j = \frac{\partial f(\boldsymbol{x}_0, \beta_{nom})}{\partial \beta_j}$$

If the two relations

$$\boldsymbol{y} = \boldsymbol{f}_{\beta_{nom}} + \boldsymbol{H}_{\beta_{nom}}(\beta_0 - \beta_{nom}) + \boldsymbol{z} + \boldsymbol{\epsilon}$$

$$y_0 = f(\boldsymbol{x}_0, \beta_{nom}) + \boldsymbol{h}_{\beta_{nom}}(\boldsymbol{x}_0)^t(\beta_0 - \beta_{nom}) + Z(\boldsymbol{x}_0) + \epsilon_0$$

are exact, then we have

$$\mathcal{L}(\phi(\boldsymbol{x}_0)|\boldsymbol{y}) = \mathcal{N}(\hat{\phi}(\boldsymbol{x}_0), \hat{\sigma}^2(x_0)) \qquad \text{and} \qquad \mathcal{L}(y_0|\boldsymbol{y}) = \mathcal{N}(\hat{\phi}(\boldsymbol{x}_0), \hat{\sigma}^2(x_0) + \sigma_m^2)$$

with

$$\hat{\phi}(\boldsymbol{x}_0) =$$
$$f(\boldsymbol{x}_0, \beta_{nom}) + (\boldsymbol{h}_{\beta_{nom}}(\boldsymbol{x}_0))^t(\beta_{post} - \beta_{nom}) + (\boldsymbol{r}(\boldsymbol{x}_0))^t \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{f}_{\beta_{nom}} - \boldsymbol{H}_{\beta_{nom}}(\beta_{post} - \beta_{nom}))$$

and

$$\hat{\sigma}^2(\boldsymbol{x}_0) = C(\boldsymbol{x}_0, \boldsymbol{x}_0) - \boldsymbol{r}(\boldsymbol{x}_0)^t \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}_0)$$
$$+ (\boldsymbol{h}_{\beta_{nom}}(\boldsymbol{x}_0) - \boldsymbol{H}_{\beta_{nom}}^t \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}_0))^t (\boldsymbol{H}_{\beta_{nom}}^t \boldsymbol{R}^{-1} \boldsymbol{H}_{\beta_{nom}} + Q_{prior}^{-1})^{-1} (\boldsymbol{h}_{\beta_{nom}}(\boldsymbol{x}_0) - \boldsymbol{H}_{\beta_{nom}}^t \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}_0))$$

Proof : same as for calibration

Exact non-linear calibration

We have

$$\boldsymbol{y} = \boldsymbol{f}_{\beta_0} + \boldsymbol{z} + \epsilon,$$

and

$$\phi(x_0) = f(\boldsymbol{x}_0, \beta_0) + Z(\boldsymbol{x}_0)$$

Hence

$$\mathbb{E}(\beta_0|\boldsymbol{y}) = \frac{\int_{\beta\in\mathbb{R}^d} \beta p(\boldsymbol{y}|\beta)p(\beta)d\beta}{\int_{\beta\in\mathbb{R}^d} p(\boldsymbol{y}|\beta)p(\beta)d\beta}$$

and

$$cov(\beta_0|\boldsymbol{y}) = \frac{\int_{\beta\in\mathbb{R}^d} \beta\beta^t p(\boldsymbol{y}|\beta)p(\beta)d\beta}{\int_{\beta\in\mathbb{R}^d} p(\boldsymbol{y}|\beta)p(\beta)d\beta} - \mathbb{E}(\beta_0|\boldsymbol{y})\mathbb{E}(\beta_0|\boldsymbol{y})^t$$

with

$$p(\boldsymbol{y}|\beta) = \frac{1}{(2\pi)^{n/2}\sqrt{|\boldsymbol{R}|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{f}_\beta)^t \boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{f}_\beta)\right)$$

and

$$p(\beta) = \frac{1}{(2\pi)^{p/2}\sqrt{|\boldsymbol{Q}_{prior}|}} \exp\left(-\frac{1}{2}(\beta-\beta_{prior})^t \boldsymbol{Q}_{prior}^{-1}(\beta-\beta_{prior})\right)$$

One approximate non-linear calibration

We approximate the integrals with sums :
Let $\beta_1, ..., \beta_N \in \mathbb{R}^p$ be suitable to approximate integrals
$\Longrightarrow$ e.g. taken by determining a compact $B \subset \mathbb{R}^p$ containing most of the mass of $p(\beta)$, and then by random or regular sampling
$\Longrightarrow$ possibility to use integration quadratures as well

Then one can use

$$\tilde{\mathbb{E}}(\beta_0|\boldsymbol{y}) = \frac{\sum_{i=1}^N \beta_i p(\boldsymbol{y}|\beta_i)p(\beta_i)}{\sum_{i=1}^N p(\boldsymbol{y}|\beta_i)p(\beta_i)}$$

and

$$\tilde{cov}(\beta_0|\boldsymbol{y}) = \frac{\sum_{i=1}^N \beta_i \beta_i^t p(\boldsymbol{y}|\beta_i)p(\beta_i)}{\sum_{i=1}^N p(\boldsymbol{y}|\beta_i)p(\beta_i)} - \tilde{\mathbb{E}}(\beta_0|\boldsymbol{y})\tilde{\mathbb{E}}(\beta_0|\boldsymbol{y})^t$$

$\Longrightarrow$ Necessitates $N.n$ calls to the computer model
$\Longrightarrow$ Estimating C and σ_m^2 becomes complex. References with Bayesian framework on C and σ_m^2, Higdon et al 2004, Kennedy and O'Hagan 2001
$\Longrightarrow$ Here we use the estimates of C and σ_m^2 obtained from the linear approximation

Exact non-linear prediction

We have, using the Gaussian conditioning theorem

$$\mathbb{E}(\phi(x_0)|\boldsymbol{y},\beta_0) = f(\boldsymbol{x}_0,\beta_0) + \boldsymbol{r}(\boldsymbol{x}_0)^t \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{f}_{\beta_0})$$

Hence,

$$\begin{aligned}\mathbb{E}(\phi(x_0)|\boldsymbol{y}) &= \mathbb{E}(\mathbb{E}(\phi(x_0)|\boldsymbol{y},\beta_0)|\boldsymbol{y}) \\ &= \frac{\int_{\boldsymbol{\beta}\in\mathbb{R}^d} \mathbb{E}(\phi(x_0)|\boldsymbol{y},\boldsymbol{\beta}) p(\boldsymbol{y}|\boldsymbol{\beta}) p(\boldsymbol{\beta}) d\boldsymbol{\beta}}{\int_{\boldsymbol{\beta}\in\mathbb{R}^d} p(\boldsymbol{y}|\boldsymbol{\beta}) p(\boldsymbol{\beta}) d\boldsymbol{\beta}}\end{aligned}$$

Similarly, from the Gaussian conditioning theorem

$$\mathbb{E}(\phi(x_0)^2|\boldsymbol{y},\beta_0) = \mathbb{E}(\phi(x_0)|\boldsymbol{y},\beta_0)^2 + var(\phi(x_0)|\boldsymbol{y},\beta_0)$$

with

$$var(\phi(x_0)|\boldsymbol{y},\beta_0) = C(\boldsymbol{x}_0,\boldsymbol{x}_0) - \boldsymbol{r}(\boldsymbol{x}_0)^t \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}_0)$$

Hence,

$$\begin{aligned}var(\phi(x_0)|\boldsymbol{y}) &= \mathbb{E}(\phi(x_0)^2|\boldsymbol{y}) - \mathbb{E}(\phi(x_0)|\boldsymbol{y})^2 \\ &= \mathbb{E}(\mathbb{E}(\phi(x_0)^2|\boldsymbol{y},\beta_0)|\boldsymbol{y}) - \mathbb{E}(\phi(x_0)|\boldsymbol{y})^2 \\ &= \frac{\int_{\boldsymbol{\beta}\in\mathbb{R}^d} \mathbb{E}(\phi(x_0)^2|\boldsymbol{y},\boldsymbol{\beta}) p(\boldsymbol{y}|\boldsymbol{\beta}) p(\boldsymbol{\beta}) d\boldsymbol{\beta}}{\int_{\boldsymbol{\beta}\in\mathbb{R}^d} p(\boldsymbol{y}|\boldsymbol{\beta}) p(\boldsymbol{\beta}) d\boldsymbol{\beta}} - \mathbb{E}(\phi(x_0)|\boldsymbol{y})^2\end{aligned}$$

One approximate non-linear prediction

Same as for calibration :

$$\tilde{\mathbb{E}}(\phi(x_0)|\boldsymbol{y}) = \frac{\sum_{i=1}^{N} \mathbb{E}(\phi(x_0)|\boldsymbol{y}, \boldsymbol{\beta}_i) p(\boldsymbol{y}|\boldsymbol{\beta}_i) p(\boldsymbol{\beta}_i)}{\sum_{i=1}^{N} p(\boldsymbol{y}|\boldsymbol{\beta}_i) p(\boldsymbol{\beta}_i)}$$

and

$$\tilde{var}(\phi(x_0)|\boldsymbol{y}, \boldsymbol{\beta}_0) = \frac{\sum_{i=1}^{N} \mathbb{E}(\phi(x_0)^2|\boldsymbol{y}, \boldsymbol{\beta}_i) p(\boldsymbol{y}|\boldsymbol{\beta}_i) p(\boldsymbol{\beta}_i)}{\sum_{i=1}^{N} p(\boldsymbol{y}|\boldsymbol{\beta}_i) p(\boldsymbol{\beta}_i)} - \tilde{\mathbb{E}}(\phi(x_0)|\boldsymbol{y})^2$$

$\Longrightarrow$ No need for additional calls to computer model after calibration

Note that we still have

$$\mathbb{E}(y_0|\boldsymbol{y}) = \mathbb{E}(\phi(x_0)|\boldsymbol{y})$$

and

$$var(y_0|\boldsymbol{y}) = var(\phi(x_0)|\boldsymbol{y}) + \sigma_m^2$$

Comparison of linear and non-linear calibration and prediction on the FLICA IV case study

With the same FLICA IV data, we compare

- Approximate non-linear calibration and prediction with 25 grid points for integration over β
- Approximate linear calibration and prediction with the same grid (we use $f(\boldsymbol{x}, \beta) = f(\boldsymbol{x}, \beta_{nom}) + \boldsymbol{h}_{\beta_{nom}}(\boldsymbol{x})^t(\beta - \beta_{nom})$)

We obtain

- RMSE non-linear = 196.9
- RMSE linear = 197.8
- L^2 norm between linear and non-linear conditional means of β_0 = .025 (each of them has L^2 norm around 0.3)

$\Longrightarrow$ more impact of linear approximation on calibration than on prediction
$\Longrightarrow$ The model error Gaussian process can take into account the difference between the linearized computer model and the exact computer model

Summary

- Iterative linear approximations
- In the non-linear case, we can write the conditional distributions, but the integrals are intractable, since they involve the computer model
- One possibility is to integrate numerically
- There are other methods which we did not talk about, e.g. Gaussian process model of the computer model as well (Higdon et al 2004, Kennedy and O'Hagan 2001)
- We believe that the linear approximation generally has more impact on calibration than on prediction

References


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


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


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


S. Fu, Inversion probabiliste bayésienne en analyse d'incertitude, *PhD thesis, Université Paris-Sud 11, 2012*. `http://tel.archives-ouvertes.fr/tel-00766341/`.


D. Higdon, M. Kennedy, J.C. Cavendish, J.A. Cafeo and R.D. Ryne, Combining Field Data and Computer Simulations for Calibration and Prediction, *SIAM Journal on Scientific Computing (26) 448-466, 2004*.


M. Kennedy and A. O'Hagan, Bayesian Calibration of Computer Models, *Journal of the Royal Statistical Society : Series B (Statistical Methodology) (63) 425-464, 2001*.


D.A. Harville, Bayesian inference for variant components using only error contrasts, *Biometrika (61) 383-385, 1974*.

Thank you for your attention !