A Bayesian approach to characterize the physical aleas with numerical experiments

… illustrated by a case-study in hydraulics

Mathieu Couplet, Nicolas Bousquet - EDF R&D

Shuai Fu, Gilles Celeux – Université Paris-Sud

GdR Mascot Num - Working meeting
« Validation » - November 13rd, 2013

Contexte

On souhaite caractériser un aléa physique X décrit formellement par :

$$Y \quad = \quad H_r(X, d)$$

où H_r est une fonction de transfert décrivant un phénomène réel (= réponse d'un système industriel, etc.)

- d représente des **conditions environnementales connues/observables** (calables "aisément")
- X représente des **paramètres physiques non-observables**

Exemples d'application

En thermohydraulique système pour la sûreté nucléaire

Modéliser, à partir de données réelles, les paramètres de lois physiques intervenant dans le code CATHARE, utilisé pour l'évaluation de situations accidentelles.

En acoustique

Estimer l'incertitude sur l'impédance des sols environnant les installations d'EDF pour une prise en compte à terme dans les études d'impact sonore (code TYMPAN).

Modélisation et calibration = inversion de X

Concrètement, on décrit le phénomène réel par un modèle :

$$\underbrace{y_i}_{\substack{\text{une mesure}\\ \text{sous des conditions}\\ \text{expérimentales } d_i}} = \underbrace{H_r(x_i, d_i)}_{\substack{\text{réponse } \textit{réelle} \text{ du système}\\ \text{supposée dépendre de la grandeur } X}} + \underbrace{u_i}_{\text{erreur de mesure}}$$

$$= \underbrace{H(x_i, d_i)}_{\substack{\text{réponse } \textit{modélisée}\\ \text{(code de simulation)}}} + \underbrace{H_r(x_i, d_i) - H(x_i, d_i)}_{\text{erreur de modèle}} + u_i$$

X n'a (généralement) pas le même sens ni la même dimension pour H_r et H

- pour H : paramètre de calage lié à la précision, la dimension... peu ou plus de sens physique clair

H = Code déterministe boîte noire, souvent coûteux en temps de calcul

En hydraulique

- Système physique : écoulement à surface libre sur 50 km de Garonne ;

- Éq. de Saint-Venant 1D ou 2D (Mascaret/Telemac) ;

Paramètres physiques (observables) : débit, géométrie
Paramètres de calage (codes 1D/2D) : Strickler (résumant le frottement)
=> physiques mais « non-observables / peu intuitifs »

- Objectif : estimer l'incertitude sur les coefficients de Strickler pour une prise en compte dans les études de risque d'inondation.

Thèse de S. Fu (2012) / Univ. Paris XI

Erreur de modèle (pratique courante actuelle)

Calibration sans biais de modèle : risque de "surajustement" de X

écarter ce risque $\Leftrightarrow$ forte confiance dans la modélisation physique + faibles erreurs numériques

$\Leftarrow$ vérification préalable

It does not really matter if X has no true physical reality, but is simply a tuning parameter : the problem should then be thought of as finding "some type of best-fitting value of X, with the bias defined relative to this". [Bayarri *et al.*, 2007]

En l'absence de modèle d'interpolation pertinent (sur les X sachant d) pour étudier l'erreur (biais) de modèle, on décide pour le moment d'injecter dans la calibration de X une partie de l'incertitude engendrée par l'"oubli" de cette erreur

En outre : codes qualifiés pour un certain domaine de valeurs d'entrée

Aside : about identifiability, regularization and « well-posed and conditioned » problems in inversion

Inverse problems as optimization problems

Inverse problem

Let $y^* \in \mathcal{Y}$ and $H : \mathcal{X} \to \mathcal{Y}^\dagger$, find $x^* \in \mathcal{X}$ such that

$$y^* = H(x^*)$$

That is: find $x^* \in H^{-1}(\{y^*\})$, where $H^{-1}(\{y^*\}) = \{x \in \mathcal{X} \,/\, y^* = H(x^*)\}$.

---

†In practice, H is a numerical code H^{code}.

Optimization problem

Let $\|.\|$ be a norm on $\mathcal{Y}^\dagger$, find

$$x^* = \arg\min_{x \in \mathcal{X}} \|y^* - H(x)\|.$$

This formulation is more general and allows the use of optimization methods (*e.g.* a simulated annealing algorithm, which is a Metropolis-Hastings one).

---

†especially if $\mathcal{Y}$ is a metric space without norm, another way to generalize the inverse problem is minimizing $\mathrm{d}(y^*, H(x))$ where $\mathrm{d}(.,.)$ is a distance on $\mathcal{Y}$.

The linear euclidean case

Let $y^* \in \mathbb{R}^P$ and H: $\mathbb{R}^Q \to \mathbb{R}^P$, considering the inverse problem $y^* = H(x^*)$ and a euclidean-norm $\|.\|_S$ induced by $S \in \mathcal{S}_P^{++}(\mathbb{R})$, that is $\|y\|_S{}^2 = y^T S\, y$, find

$$\hat{x} = \arg\min_{x \in \mathbb{R}^Q} \|y^* - H(x)\|_S \quad \Leftrightarrow \quad \hat{x} = \arg\min_{x \in \mathbb{R}^q} (y^* - H(x))^T S\, (y^* - H(x))$$

Assumption: $\exists H \in \mathcal{M}_{P,Q}(\mathbb{R}) \quad H(x) = H\,x.$

Then:

- $\hat{x}$ can be defined as a solution to a linear system:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^Q} \|y^* - H(x)\|_S \quad \Leftrightarrow \quad H^T S\, y^* = H^T S\, H\, \hat{x}.$$

- $\hat{x}$ is unique whatever y^* is iif the rank of H is maximal, which implies $P \geq Q$ ("as many pieces of data as many unknowns is needed") and $H^T S\, H \in \mathcal{S}_P^{++}(\mathbb{R})$.

Well-conditioned problems

Well-posed problem in the sense of Hadamard

1. A solution x^* exists: $H^{-1}(y^*) \neq \emptyset$.
2. The solution is unique: $H^{-1}(y^*) = \{x^*\}$.
3. The solution depends continuously on the data, in some reasonnable topology: H^{-1} is continuous around y^*.

Well-conditioned problem: a practical necessity

In practice, it is necessary to go beyond that latter condition and to ensure that "a small error on y^* may not cause a large error on x^*": the *condition number* of H^{-1} around y^* must be small [†].

If not, you cannot expect to compute an "accurate" solution: the problem need to be *regularized*.

[†] Condition number of f at x: $\lim_{\epsilon \to 0^+} \sup_{\|\delta x\| < \epsilon} \left[\frac{\|f(x+\delta x) - f(x)\|}{\|f(x)\|} \right] / \left[\frac{\|\delta x\|}{\|x\|} \right]$.

Tikhonov regularization

This is the most commonly used method of regularization. The linear euclidean Tikhonov-regularized problem is: find

$$\hat{x}_{x_0,\Lambda} = \arg\min_{x\in\mathbb{R}^Q} \|y^* - Hx\|_S^2 + \|\Lambda(x - x_0)\|_2^2 \Leftrightarrow \underbrace{\left[H^T SH + \Lambda^T\Lambda\right]}_{A_{x_0,\Lambda}} \hat{x}_{x_0,\Lambda} = H^T Sy^* + \Lambda^T /$$

with $\Lambda \in \mathcal{M}_{K,Q}(\mathbb{R})^\dagger$ and $x_0 \in \mathbb{R}^Q$.

x_0 appears as some kind of guess of what x^* could be, that is as a prior piece of information! What if one associates a prior distribution $\pi(x)$ on x?

Bayesian inference as regularization

We assume the following model:

- Likelihood coming from

$$Y^* = H x^* + U \text{ where } U \sim \mathcal{N}(0, {\sigma_U}^2 S)$$

  with $S = \mathrm{I}_P$ (OLS) and known ${\sigma_U}^{\dagger}$:

$$Y^* \sim \mathcal{N}(Hx^*, {\sigma_U}^2 \mathrm{I}_P);$$

- Prior: $x^* \sim \mathcal{N}(x_0, {\sigma_x}^2 \mathrm{I}_Q)$.

---

†Variability of the measurement "errors" already estimated.

Bayesian inference as regularization

Model: $Y^* \sim \mathcal{N}(Hx^*, {\sigma_U}^2 \mathrm{I}_P)$ and $x^* \sim \mathcal{N}(x_0, {\sigma_x}^2 \mathrm{I}_Q)$.

Therefore,

$$\begin{aligned}
\pi(x|y^*) &\underset{x}{\propto} \mathcal{L}(x; y^*)\, \pi(x) \\
&\underset{x}{\propto} \exp\left(-\tfrac{1}{2}\left[\tfrac{1}{{\sigma_x}^2}(y^* - Hx)^T(y^* - Hx) + \tfrac{1}{{\sigma_U}^2}(x - x_0)^T(x - x_0)\right]\right) \\
&\underset{x}{\propto} \exp\left(-\tfrac{1}{2}\left[x^T\Sigma^{-1}x - 2\mu^T\Sigma^1 x\right]\right) \underset{x}{\propto} \exp\left(-\tfrac{1}{2}(x - \mu)^T\Sigma^{-1}(x - \mu)\right)
\end{aligned}$$

with $\Sigma^{-1} = \frac{1}{{\sigma_x}^2}(H^T H + \frac{{\sigma_x}^2}{{\sigma_U}^2}\mathrm{I}_Q)$ and $\mu^T\Sigma^{-1} = \frac{1}{{\sigma_x}^2}{y^*}^T H + \frac{1}{{\sigma_U}^2}{x_0}^T$.

Thus $\mu = \left(H^T H + \frac{{\sigma_x}^2}{{\sigma_U}^2}\mathrm{I}_Q\right)^{-1}\left({y^*}^T H + \frac{{\sigma_x}^2}{{\sigma_U}^2}{x_0}^T\right)$ and

$$x|y^* \sim \mathcal{N}(\hat{x}_{x_0,\Lambda}, \Sigma) \text{ with } \Lambda = \frac{\sigma_x}{\sigma_U}\mathrm{I}_Q.$$

Hence, bayesian inference appears as a way to regularize ill-conditioned problem. Moreover, here, the bayesian approach gives rationale for the choice of Λ and x_0.

Bayesian inference as regularization against non-identifiability

Bayesian statistical model

Roughly speaking, a *bayesian statistical model* is defined by the "combination" of a parametrized statistical model $(\mathcal{Y}, \mathcal{T}_{\mathcal{Y}}, \{P_\theta \ : \ \theta \in \Theta\})$ with a prior $\pi(\theta)$ which is a probability distribution on $(\Theta, \mathcal{T}_\Theta)$.

Remark

- $(\mathcal{Y}, \mathcal{T}_{\mathcal{Y}}, \{P_\theta \ : \ \theta \in \Theta\})$ is chosen so as to only model the "true" variability/dispersion/randomness which is postulated by the observer;
- whereas the prior $\pi(\theta)$, chosen in coherence with the available knowledge, represents the epistemic uncertainty (lack of knowledge) on θ^*.

Thus, the use of gaussian processes for deterministic numerical code meta-modelling is intrinsically a bayesian modelling.

Bayesian inference as regularization against non-identifiability

About the well-poseness of the bayesian inference problem

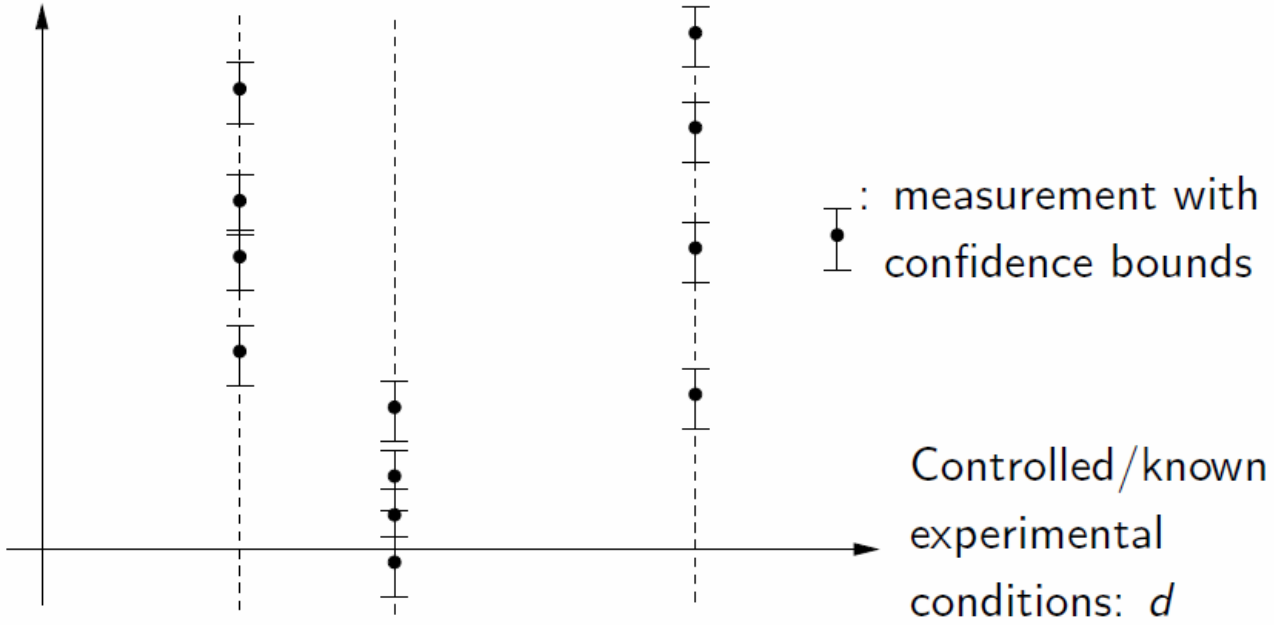
The bayesian inference approach is completely based on the determination of the posterior $\pi(\theta|y^*)$ whose existence and uniqueness are always garanteed (unless the data are incoherent with your model or in some cases of improper prior).

Moreover, most of the Bayes rule (minimizing the posterior expected loss) lead to a unique Bayes estimator[†]: the existence and uniqueness of a Bayes estimate $\hat{\theta}$ of θ^* is generally ensured in practice.

[†] Posterior expectation or quantile, typically.

Back to the problem : the data point of view

Response of the physical system: $y = H^{\text{real}}(x, d)$



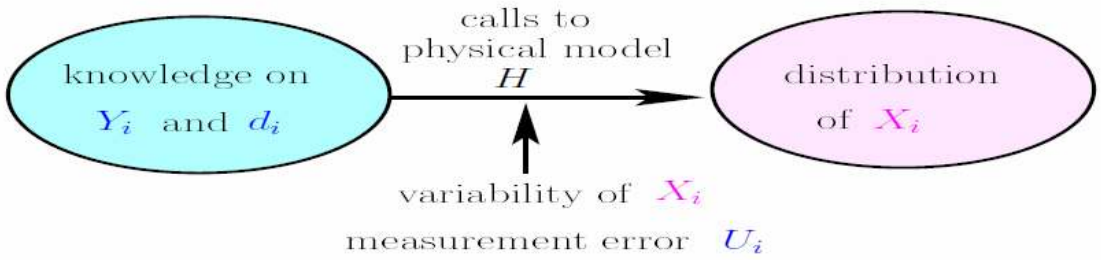
For a given d, the variability of the measurements is significant: the uncertain parameter x cannot be reasonably be thought as steady. x has to be consider as an irreductible hazard X: its characterization is naturally a statistical problem and not a "classical" inverse problem.

Calibration of a « physical tuning » parameter by inversion of a costly nonlinear computer code

From the Ph.D work by S. Fu (2012), supervized by G. Celeux

Formulation : $Y_i = H(X_i, d_i) + U_i,\ 1 \leq i \leq n,$ with

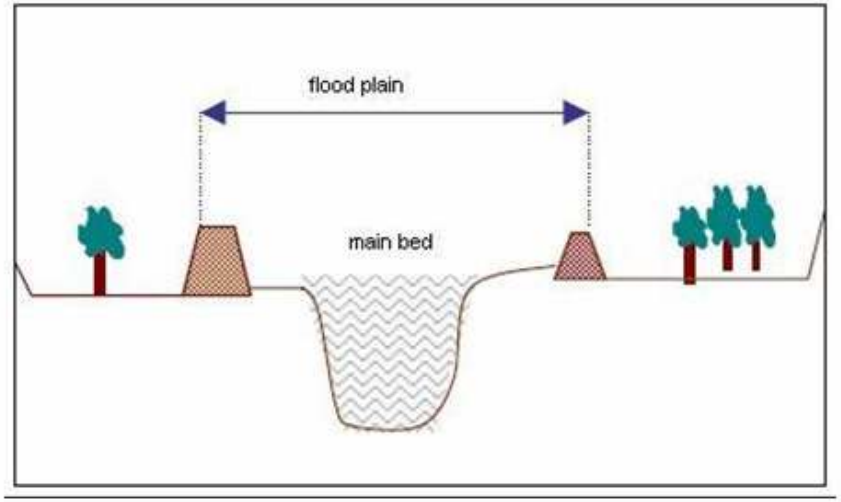
- $Y_i \in \mathbb{R}^p$: measured/observed vector data ;
- H : known physical model, highly expensive ;
- $X_i \in \mathbb{R}^q$: non observed random data, assumed i.i.d. ;
- $d_i \in \mathbb{R}^{q_2}$: observed variables related to experimental conditions ;
- $U_i \in \mathbb{R}^p$: measurement errors, assumed i.i.d., independent from X_i.



Objective : to estimate the distribution of X_i from the data Y_i and d_i.

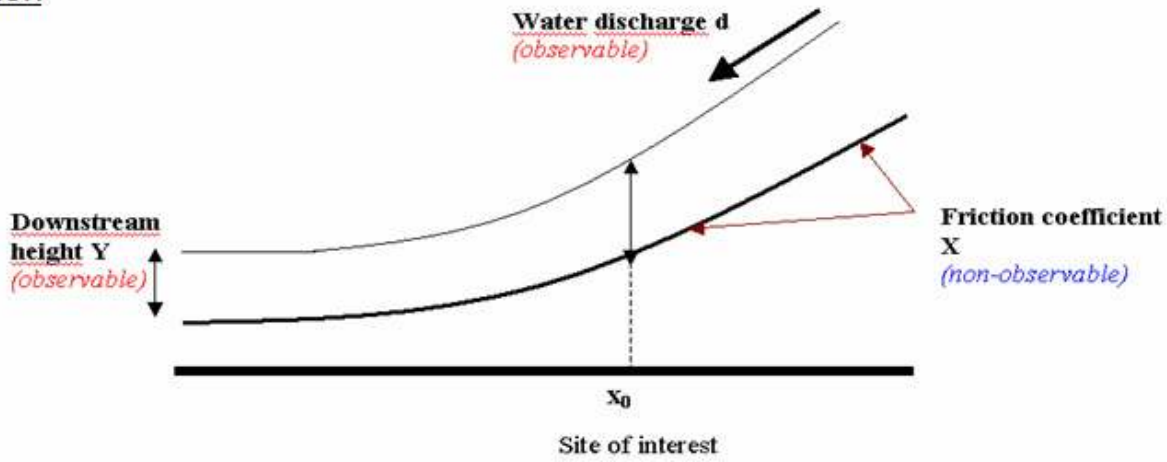
Motivating application : hydraulical case-study

- A river portion is considered
- *H* is a **hydraulical code** predicting the downstream water level *Y* at various sites of interest



- **Known inputs** (*d*) : measures of upstream water discharge

- **Unknown stochastic inputs to inverse** (*X*) : the independent friction coefficients characterizing :
  - main bed
  - flood plain



How to estimate the missing-data X_i ?

Model assumptions :

$$\begin{aligned} X_i|m, C &\sim \mathcal{N}_q(m, C) \\ U_i &\sim \mathcal{N}_p(0, R),\ R \text{ assumed to be known} \end{aligned}$$

$\Rightarrow$ estimate $\theta = (m, C)$ from observations $\mathbf{y} = \{y_1, \ldots, y_n\}$ and $\mathbf{d} = \{d_1, \ldots, d_n\}$.

- Frequentist inference :

$$\text{e.g. (MLE)} \qquad \widehat{\theta} = \arg\max_{\theta \in \Theta} \log \int \mathcal{L}(\theta;\, \mathbf{X}, \mathbf{y}, \mathbf{d})\, d\mathbf{X},$$

with $\mathbf{X} = \{X_1, \ldots, X_n\}$ and $\mathcal{L}(\theta; \ldots)$ the likelihood of θ.

$\Rightarrow$ the ECME (Celeux et *al.*, 2010), SEM (Barbillon, 2010), MCEM, SAEM (Delyon et *al.*, 1999, Kuhn, 2003) algorithms

- Bayesian inference :

$\Rightarrow$ the key viewpoint in this thesis : θ is considered as a random variable. But why Bayesian ? And how ?

The Bayesian inference is chosen because it ...

- adds the available expert knowledge (often the case in engineering);

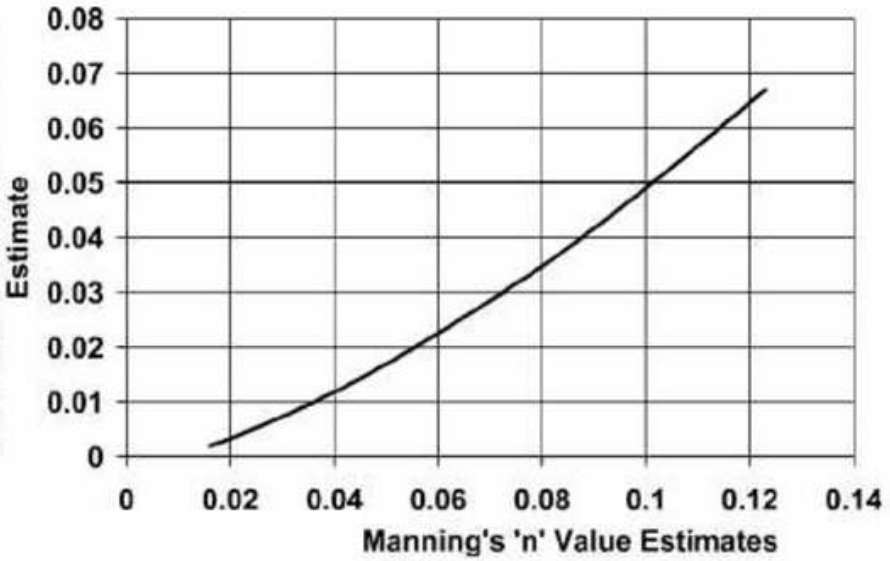
Contextual values of Manning's coefficient

[US Geological Survey 1989, Viollet et al. 1998]

Surface Description	n^a
Smooth surfaces (concrete, asphalt, gravel, or bare soil)	0.011
Fallow (no residue)	0.05
Cultivated soils	
Residue cover (less than 20%)	0.06
Residue cover (greater than 20%)	0.17
Grass	
Short grass prairie	0.15
Dense grasses[b]	0.24
Bermuda grass	0.41
Range (natural)	0.13
Woods[c]	
Light underbrush	0.40
Dense underbrush	0.80

Link between mean and relative error

[Liu 2009]



- is favorable in small sample size setting, with few data y_i and d_i available;

  Example : during several years, there are only dozen measurements of the discharge are available;

- improves the robustness with regard to identifiability problems;

Prior constraint on Cov(X) (1/3)

Solving the inversion problem (assessing the distribution of X from information on Y) gets sense only if the uncertainty on Y is mainly explained by the uncertainty on X.

How expressing this idea ?

1 - **Variance analysis** : assume a linear or a linearized model

$$Y_i \quad = \quad H_i(X_i - x_{0,i}) + \epsilon_i \qquad i = 1, \ldots, n$$

where H_i is the jacobian of $H(., d_i)$ at $x_{0,i}$. Simplify in

$$Y_i \quad = \quad H_i X_i + \epsilon_i$$

and rewrite

$$\mathbf{Y} \quad = \mathbf{\Gamma X} + \boldsymbol{\epsilon}$$

with

- $\mathbf{Y}^T = [Y_1, \ldots, Y_n]$, $X^T = [X_1, \ldots, X_n]$ and $\boldsymbol{\epsilon}^T = [\epsilon_1, \ldots, \epsilon_n]$
- $\mathbf{\Gamma} = \begin{bmatrix} H_1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & H_2 & 0 & \ldots & 0 & 0 \\ \ldots & & & & & \\ 0 & \ldots & & & 0 & H_n \end{bmatrix}$
- and we define $\mathbf{H}^T = [H_1, \ldots, H_n]$

Prior constraint on Cov(X) (2/3)

It is proposed to consider the following condition of "relevance" or "identifiability" of the inversion problem

$$|\mathrm{Cov}(\mathbf{\Gamma X})| \quad > \quad |\mathrm{Cov}(\epsilon)| \tag{1}$$

Such a constraint can be a part of prior specification

Ex : Gaussian model

$$\begin{aligned} X|\theta = (m, C) &\sim \mathcal{N}_q(m, C) \\ m|C &\sim \mathcal{N}_q(\mu, C/a) \\ C &\sim \pi(C) \end{aligned}$$

Then

$$|\mathrm{Cov}(\mathbf{\Gamma X})| \quad = \quad |\mathbf{\Gamma C \Gamma}^T| \; = \; |C|^n |\mathbf{\Gamma}^T \mathbf{\Gamma}|$$

$\pi(C)$ must be such that (1) is true :

$$\boxed{|C| \quad > \quad |R|/(|\mathbf{\Gamma}^T \mathbf{\Gamma}|)^{1/n}}$$

- in linear cases originally
- it seems reasonnable to extend this rule to nonlinear cases

Prior constraint on Cov(X) (3/4)

2 - A more intrinsic approach

$$\mathcal{E}(H(X,d)) \quad \geq \quad \mathcal{E}(\epsilon) \tag{2}$$

where $\mathcal{E}(Z)$ is the entropy of the random variable $Z \sim f$, ie. $\mathcal{H}(Z) = -\int f \log f$

The same result can be found

Under the condition that $|\mathbf{\Gamma}^T \mathbf{\Gamma}| > 0$, C lives in

$$\Omega_C \quad = \quad \{C \in \mathcal{S}_q^+, \ |C| \geq |C_0|\}$$

Consequently, any noninformative prior $\pi(C) \propto |C|^{-1-k}$ with $k > 0$ is proper on Ω_C

Prior constraint on Cov(X) (4/4)

This condition is not satisfactory, since

$$|\mathbf{\Gamma}^T \mathbf{\Gamma}| \quad = \quad \prod_{i=1}^{n} |H_i^T H_i|$$

If at least one H_i is not of rank q (non-injective), then $H_i^T H_i = 0$

This occurs when $p < q$ (so frequently)

The condition $|\mathbf{\Gamma}^T \mathbf{\Gamma}| > 0$ is much stronger than the identifiability condition (rank($\mathbf{H}$) $= q$)

Inference via Gibbs algorithm

Given $(m^{[r]}, C^{[r]}, \mathbf{X}^{[r]})$ for $r = 0, 1, 2, \ldots,$ generate:

1. $C^{[r+1]}|\ldots \sim$
   $\mathcal{IW}\Big(\Lambda + \sum_{i=1}^{n}(m^{[r]} - X_i^{[r]})(m^{[r]} - X_i^{[r]})' + a(m^{[r]} - \mu)(m^{[r]} - \mu)', \nu + n + 1\Big)$,
2. $m^{[r+1]}|\ldots \sim \mathcal{N}\Big(\frac{a}{n+a}\mu + \frac{n}{n+a}\overline{\mathbf{X}_n^{[r]}}, \frac{C^{[r+1]}}{n+a}\Big)$ where $\overline{\mathbf{X}_n^{[r]}} = \frac{1}{n}\sum_{i=1}^{n} X_i^{[r]}$,
3. $X^{[r+1]}|\ldots \propto \exp\Big[-\frac{1}{2} - \sum_{i=1}^{n}(m - X_i)^T C^{-1}(m - X_i) - \frac{1}{2}\sum_{i=1}^{n}\Big(Y_i - H(X_i, d_i)\Big)^T R^{-1}\Big(Y_i - H(X_i, d_i)\Big)\Big]$.

★ unknown density form !

⇒ numerical method (ex. Metropolis-Hastings) required

⇒ high number of calls to H, necessary to replace H by cheaper $\widehat{H}$

Use of kriging techniques

Adaptation to code kriging emulation

$$Y_i = H(Z_i) + U_i, \text{ with } Z_i = (X_i, d_i), i = 1, \ldots, n$$

New uncertainty $\downarrow$ Y_i is the realization of a GP $\mathcal{Y}_i$

New dependence $\downarrow$ between sample points $\mathcal{H}(Z_k)$ and $\mathcal{H}(Z_l)$

$$\mathcal{Y} \,|\, \mathbf{Z}, \mathcal{H}_{D_N} = \mathbf{H}_{D_N} \sim \mathcal{N}\left[\widehat{H}(\mathbf{Z}), \mathbf{R} + \mathrm{MSE}(\mathbf{Z})\right],$$

where

$$\mathrm{MSE}(\mathbf{Z}) = \begin{pmatrix} \mathrm{MSE}_1(\mathbf{Z}) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathrm{MSE}_p(\mathbf{Z}) \end{pmatrix} \begin{matrix} \} \ n \text{ lines} \\ \\ \} \ n \text{ lines} \end{matrix}$$

and

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{R}_p \end{pmatrix} \begin{matrix} \} \ n \text{ lines} \\ \\ \} \ n \text{ lines} \end{matrix}, \text{ with } \mathbf{R}_i = \begin{pmatrix} R_{ii} & & 0 \\ & \ddots & \\ 0 & & R_{ii} \end{pmatrix}$$

Prior calibration and choice of domain Ω

1. Elicitation for conjugate priors :

$$\begin{array}{rcl} X_i|m,C & \sim & \mathcal{N}_q(m,C) \\ m|C & \sim & \mathcal{N}_q(\mu, C/a) \\ C & \sim & \mathcal{IW}_q(\Lambda,\nu) \end{array} \quad \Longrightarrow \quad \left\{ \begin{array}{l} \mu = \text{prior mean}, \\ a = \text{size of virtual sample}, \\ t = a+1, \\ \nu = t+q+1, \\ \Lambda = t \cdot \widetilde{C}_{\text{Exp}}. \end{array} \right.$$

2. We choose the domain Ω as a confidence domain according to the prior predictive distribution

$$X_i \sim \text{St}_q\left(\mu, \frac{(1+\frac{1}{a})\cdot\Lambda}{\nu+1-q}, \nu+1-q\right)$$

Application to the Garonne river (France)

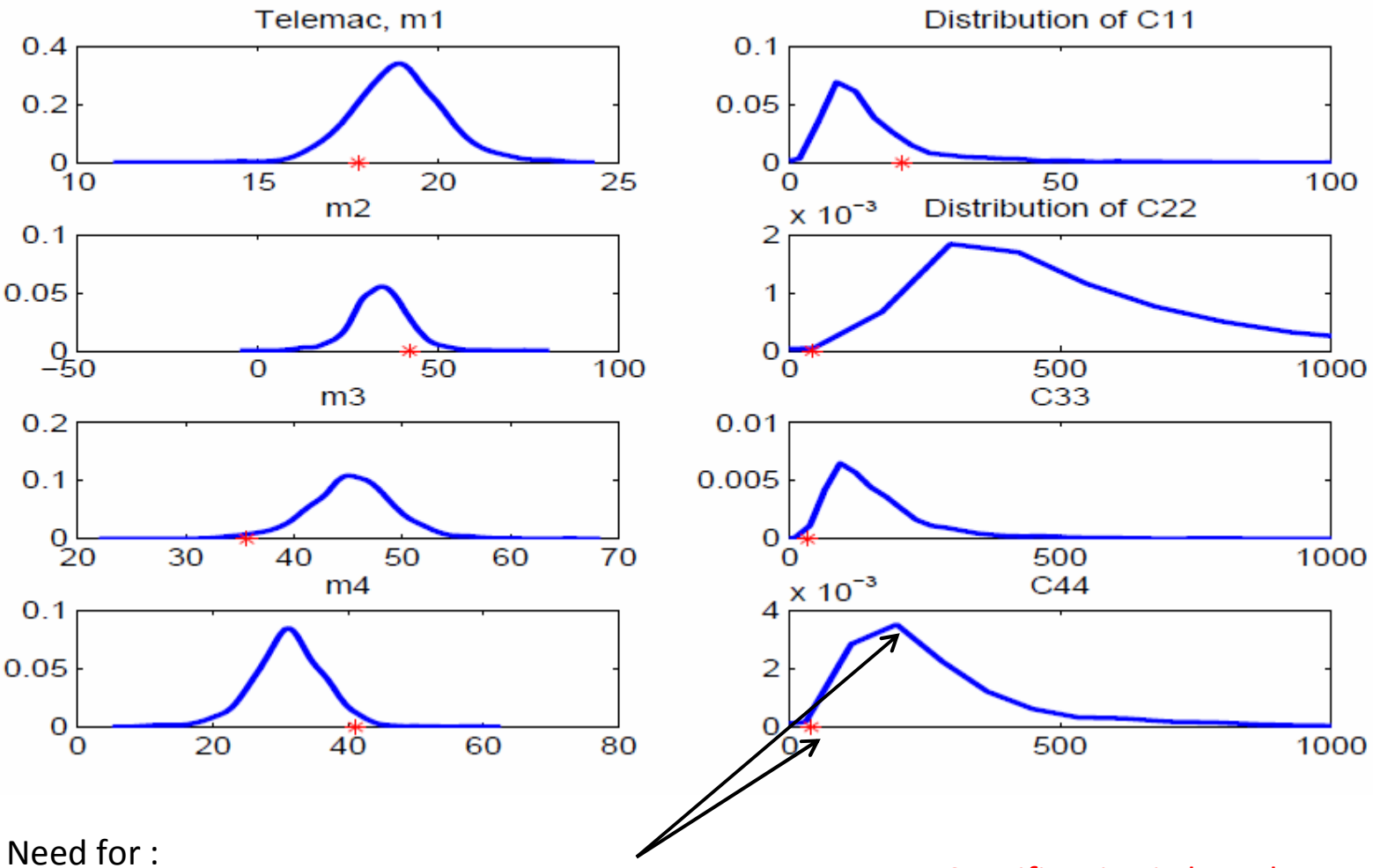
3 sections characterized
by 3 X_k = Strickler roughness coefficients

Reduced to dimension 2 for MASCARET code or dimension 4 for TELEMAC code

Prior elicitation for the Garonne Strickler

Hyperparameters	Mascaret	Telemac-2D
μ	$\begin{pmatrix} 17 \\ 40 \end{pmatrix}$	$\begin{pmatrix} 17 \\ 45 \\ 38 \\ 40 \end{pmatrix}$
a	1	1
t	2	2
ν	5	7
C_{Exp}	$\begin{pmatrix} 4.1^2 & 0 \\ 0 & 7.1^2 \end{pmatrix}$	$\begin{pmatrix} 4.1^2 & 0 & 0 & 0 \\ 0 & 7.1^2 & 0 & 0 \\ 0 & 0 & 7.1^2 & 0 \\ 0 & 0 & 0 & 7.1^2 \end{pmatrix}$

Incomplete posterior approximation (TELEMAC example)



Need for :
- testing if the posterior approximation is fine
- adding points to the experimental design if needed

Specific criteria have been developed by S. Fu (2012)

A focus on « testing the design »

$$\widetilde{\mathrm{DAC}}(\pi|\mathbf{y},\mathbf{H}_{D_N}) = \mathrm{KL}\left(\pi^J(\theta|\mathbf{y},\mathbf{H}_{D_N})||\pi(\theta)\right) - \mathrm{KL}\left(\pi^J(\theta|\mathbf{y},\mathbf{H}_{D_N})||\pi^J(\theta)\right)$$

Non-informative (baseline) prior

Made proper on the compact set defined by :

- the knowledge of the kriging domain
- the « identifiability constraint »

Compact set for the covariance C:

$$\Omega_C = \left\{ C = VDV^T \in \mathcal{S}_q^+ \ st. \ |D| \geq |R|^{q/p}, 0 \leq D_{ii} \leq \sqrt{\sum_{j=1}^{q} \beta_i^2}, \ i = 1 \ldots, q \right\}$$

A focus on « improving the design »

Ideally, the next point $z_{(N+1)}$ to be added into D_N is to approach the target distribution :

$$z_{(N+1)} = \underset{z\in\Omega}{\operatorname{argmin}}\, \mathrm{KL}\Big(\underbrace{\pi(\theta|\mathbf{y},\mathbf{d},H)}_{\text{target distribution}} \,||\, \underbrace{\pi(\theta|\mathbf{y},\mathbf{d},\mathbf{H}_{D_N}\cup\{H(z)\})}_{\substack{\text{posterior distribution}\\ \text{by adding } z \text{ into } D_N}} \Big),$$

Obvious problems:

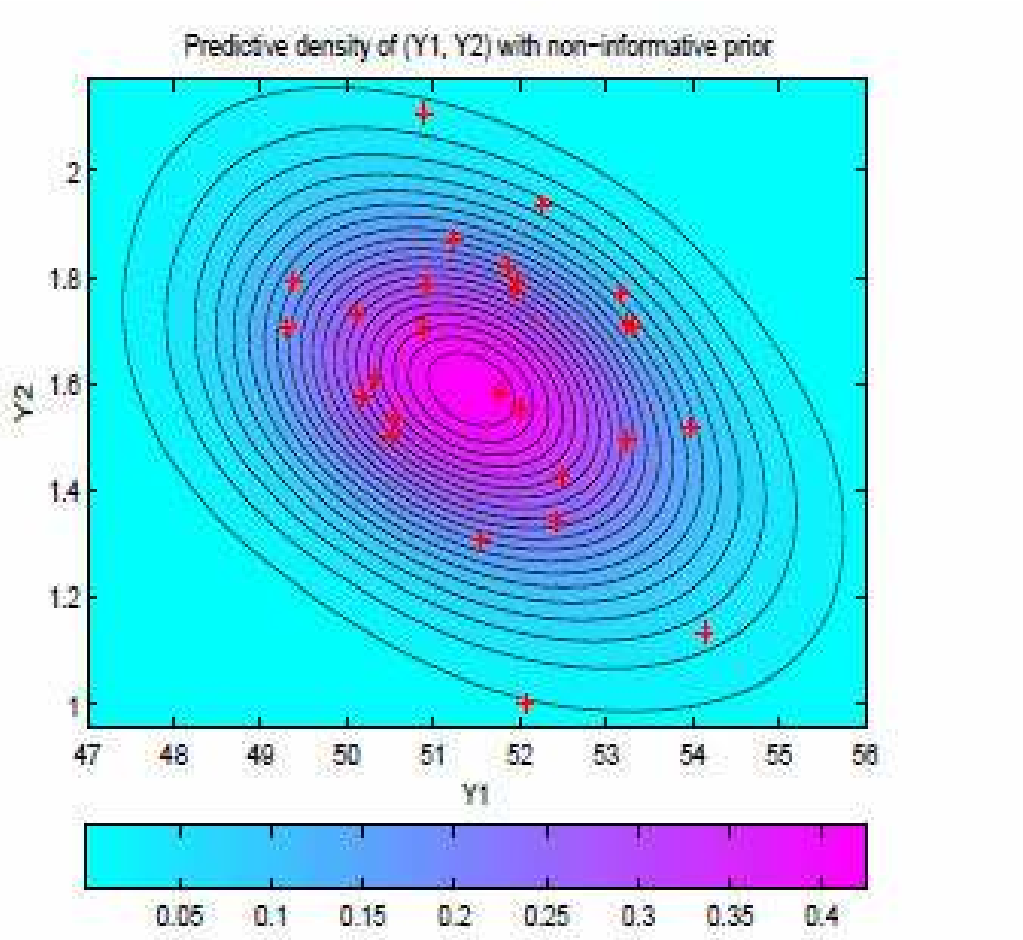
1. $\pi(\theta|\mathbf{y},\mathbf{d},H)$ is not accessible. 2. $H(z)$ is to be avoided.

A more reasonable criterion :

$$z_{(N+1)} = \underset{z\in\Omega}{\operatorname{argmax}}\, \mathbb{E}_{\pi(h_N)}\Big[\mathrm{KL}\Big(\pi(\theta|\mathbf{y},\mathbf{d},\mathbf{H}_{D_N}\cup\{h_N(z)\})\,||\,\pi(\theta|\mathbf{y},\mathbf{d},\mathbf{H}_{D_N})\Big)\Big],$$

with $h_N(z) := \mathcal{H}(z)\,|\,\mathbf{H}_{D_N} \sim \mathcal{N}(\mu(z),\sigma^2(z))$.

Checking for agreement between posterior predictive distribution and observed data



Take-home messages

A kriging emulation method is needed to carry out the inversion process

A sequential kriging can accelerate the posterior estimation

A methodology of prior elicitation can be easily adaptable from the motivating case-study to many other cases

- *take account of prior constraints inducing that the inversion problem is well-defined*
- *use the prior predictive distribution for assessing a relevant domain for designing numerical experiments*
- *provide a clear sense to hyperparameters (virtual size, expert mean opinion about roughness…)*

Multi-fidelity : inverse calibration / validation ?

« Identifiability » conditions : should be improved ?

Accounting for errors in experimental conditions

Some references

- Leonardo S. Bastos and Anthony O'Hagan (2009) Diagnostics for Gaussian Process Emulators. University of Sheffield
- Dubourg V., Deheeger F. (2010) Une alternative à la substitution pour les méta-modèles en analyse de fiabilité. *Journées Nationales de la Fiabilité, Toulouse.*
- Picheny V., Ginsbourger D., Roustant O., Haftka R.T., Kim N-H. Adaptive Designs of Experiments for Accurate Approximation of a Target Region.
- Scheidt C. (2006) Analyse statistique d'expériences simulées: Modélisation adaptative de réponses non-régulières par krigeage et plans d'expériences. *Thesis, Louis Pasteur University, Strasbourg.*
- Bettinger R. (2009) Inversion d'un système par krigeage. *Thesis, Nice-Sophia Antipolis University.*
- Barbillon P., Celeux G., Grimaud A., Lefebvre Y. and De Rocquigny E. (2009) Non linear methods for inverse statistical problems. *Rapport de research INRIA.*
- Williams B., Santner T. and Notz W. (2000) Sequential design of computer experiments to minimize integrated response functions. *Statistica Sinica.*

Rachdi, N. (2011). Apprentissage statistique et computer experiments : approche quantitative du risque et des incertitudes en modélisation. Thèse de l'Université Toulouse III

Fu, S. (2012) Bayesian probabilistic inversion in uncertainty analysis. Thèse de l'Université Paris Sud

Fu, S., G., Couplet, M. Bousquet, N. (2013) . **An adaptive kriging method for characterizing uncertainty in inverse problems** (submitted soon)

Fu, S., Celeux, G., Bousquet, N., Couplet, M. (2012). **Bayesian inference for inverse problems occuring in uncertainty analysis. INRIA RR-7995**