



HAL
open science

Bayesian inference for inverse problems occurring in uncertainty analysis

Shuai Fu, Gilles Celeux, Nicolas Bousquet, Mathieu Couplet

► **To cite this version:**

Shuai Fu, Gilles Celeux, Nicolas Bousquet, Mathieu Couplet. Bayesian inference for inverse problems occurring in uncertainty analysis. [Research Report] RR-7995, INRIA. 2012. hal-00708814

HAL Id: hal-00708814

<https://inria.hal.science/hal-00708814v1>

Submitted on 15 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Bayesian inference for inverse problems occurring in uncertainty analysis

Shuai Fu, Gilles Celeux, Nicolas Bousquet, Mathieu Couplet

**RESEARCH
REPORT**

N° 7995

Juin 2012

Project-Teams SELECT Inria and
EDF R&D

ISRN INRIA/RR--7995--FR+ENG

ISSN 0249-6399



Bayesian inference for inverse problems occurring in uncertainty analysis

Shuai Fu^{*†}, Gilles Celeux^{*}, Nicolas Bousquet[†], Mathieu
Couplet[†]

Project-Teams SELECT Inria and EDF R&D

Research Report n° 7995 — Juin 2012 — 27 pages

Abstract: The inverse problem considered here is to estimate the distribution of a non-observed random variable X from some noisy observed data Y linked to X through a time-consuming physical model H . Bayesian inference is considered to take into account prior expert knowledge on X in a small sample size setting. A Metropolis-Hastings within Gibbs algorithm is proposed to compute the posterior distribution of the parameters of X through a data augmentation process. Since calls to H are quite expensive, this inference is achieved by replacing H with a kriging emulator interpolating H from a numerical design of experiments. This approach involves several errors of different nature and, in this paper, we pay effort to measure and reduce the possible impact of those errors. In particular, we propose to use the so-called DAC criterion to assess in the same exercise the relevance of the numerical design and the prior distributions. After describing how computing this criterion for the emulator at hand, its behavior is illustrated on numerical experiments.

Key-words: Inverse problems, Bayesian analysis, Kriging, Design of Experiments, Assessment Error.

E-mail addresses: shuai.fu@edf.fr, gilles.celeux@math.u-psud.fr, nicolas.bousquet@edf.fr, mathieu.couplet@edf.fr

* University of Paris-Sud 11, Mathematics Dept., Bat. 425, 91405 Orsay (France)

† EDF R&D, Industrial Risk Management Dept., 6 quai Watier, 78401 Chatou (France)

**RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE**

Parc Orsay Université
4 rue Jacques Monod
91893 Orsay Cedex

**Approche bayésienne pour les problèmes inverses en analyse
d'incertitude
rapport de recherche
Inria**

Résumé : Le problème inverse considéré est d'estimer la distribution d'une variable aléatoire non observée X à partir d'observations bruitées Y , à l'aide d'un modèle physique d'obtention coûteuse H . Le cadre bayésien nous permet de prendre en compte les connaissances préalables d'experts surtout avec peu de données disponibles. Un échantillonneur de Gibbs combiné avec l'algorithme de Metropolis-Hastings est utilisé pour approcher la distribution a posteriori de X . La fonction coûteuse H est remplacée par un émulateur de krigeage (méta-modèle) \hat{H} basé sur un plan d'expérience (*design*). Cette approche implique plusieurs erreurs de nature différente et, dans ce rapport, nous nous attachons à estimer et réduire l'impact de ces erreurs. En particulier, nous proposons d'utiliser le critère DAC pour évaluer la qualité du design ainsi que le choix de la loi a priori. Après avoir décrit le calcul de ce critère, son comportement est illustré par les expériences numériques.

Mots-clés : Problèmes inverses, Analyse bayésienne, Krigeage, Plan d'expérience, Évaluation d'erreur

1 Introduction

Probabilistic uncertainty treatment is gaining fast growing interest in the industrial field. Besides the uncertainty propagation challenges when dealing with complex and high CPU-time demanding physical models, one of the key issues regards the quantification of the sources of uncertainties. A key difficulty is linked to the highly-limited sampling information directly available on uncertain input variables. It can be highly beneficial (a) to integrate expert judgment, such as likely bounds on physical intervals or more elaborate probabilistic information, or (b) to integrate indirect information, such as data on other, more easily observable, parameters that are linkable to the uncertain variable of interest by a physical model. Methods for (b) are making use of probabilistic inverse methods since the recovering of indirect information involves generally the inversion of a physical model or a computer simulator H . It leads to the following uncertainty model

$$Y_i = H(X_i, d_i) + U_i, i \in \{1, \dots, n\}, \quad (1)$$

where $X_i \in \mathbb{R}^q$ is a non-observed input, $d_i \in \mathbb{R}^{q_2}$ an observed input related to the experimental conditions and $U_i \in \mathbb{R}^p$ a measurement error. The error U_i and X_i are assumed to be independent for $i = 1, \dots, n$. Moreover the $(Y_i, i = 1, \dots, n)$ are independent. The purpose is to estimate the distribution of the random vectors X_i s from the observations $(\mathbf{y}_i, i = 1, \dots, n)$, knowing that the function H (the physical model...) cannot be inverted. In what follows, the random vector X_i will be assumed to have a Gaussian distribution $\mathcal{N}_q(m, C)$, with mean m and variance matrix C to be estimated, and the error vector U_i will be assumed to have a Gaussian distribution $\mathcal{N}_p(\mathbf{0}, R)$, with known diagonal variance matrix R .

Many approaches are possible to approximate this inverse problem as linearizing the physical model H around a fixed point x_0 (see Celeux et al. 2010, [11]), or using a non linear approximation of the function H obtained through kriging and making use of a stochastic procedure with this non linear approximation of H (see Barbillon et al. 2011, [12]). In this paper we opt for a Bayesian approach allowing to take into account prior knowledge that can be helpful, in particular, to avoid identifiability problems.

The estimation problem related to this inverse problem involves many possible errors:

- *Estimation error*: Usually the sample size n is small with respect to the dimension of the problem and the variance of the estimates could be expected to be large;
- *Emulator error*: Since H is too complex, there is the need to replace it with an emulator \hat{H} and the discrepancy between H and \hat{H} could induce an important error;
- *Algorithmic error*: To proceed to statistical inference, there is the need to use complex stochastic algorithms. In the Bayesian setting, those algorithms are Monte Carlo Markov Chains (MCMC) algorithms which produce Markov chains converging to the desired posterior distributions. But, controlling the convergence of the MCMC algorithms towards their limit distributions is important to get reliable estimates.
- *Prior error*: The prior knowledge on the parameters m and C is expected to produce regularized estimates of smaller variances than maximum likelihood estimates. But, if the prior distributions are irrelevant, it could jeopardize the statistical analysis.

Beyond the estimation problem, this paper is mainly concerned with the assessment of the quality of the proposed estimates. It implies to measure and control the above mentioned error sources. In this context, we focus on the *prior error* which received little attention and propose to measure it with a criterion (DAC) well-adapted for emulators defined on a compact set. Obviously those

different error sources are linked and their relations for uncertainty analysis with small samples are discussed. The paper is organised as follows. In Section 2, the MCMC algorithm for a Bayesian estimation of an emulator of model (1) is presented and the possible error sources are precisely described. Then, the DAC criterion to measure the prior error is presented in Section 3 as the resulting strategy for assessing both the emulator and the prior distribution. Numerical experiments, where different criteria assessing the different error sources are illustrated and compared, are presented in Section 4 and a Discussion section ends the paper.

2 Bayesian inference with a Gaussian emulator

In the Bayesian framework, the first task is to choose a prior distribution $\pi(\theta)$ for the parameter $\theta = (m, C)$ to be estimated in the model (1). A conjugate prior distribution has been selected

$$m | C \sim \mathcal{N}_q(\mu, C/a); \quad (2)$$

$$C \sim \mathcal{IW}_q(\Lambda, \nu), \quad (3)$$

the hyperparameters $\rho = (\mu, a, \Lambda, \nu)$ being specified by the user.

The posterior distribution $\pi(\theta | \mathbf{y})$ is approximated with a Gibbs sampler including a Metropolis-Hastings step (see for instance Tierney, 1995, [23]). Actually, the calculation of the full conditional posterior distributions of m, C and $\mathbf{X} = \{X_1, \dots, X_n\}$ lead to the following Gibbs sampler (below the $(r + 1)$ -th iteration):

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

1. $C^{[r+1]} | \dots \sim \mathcal{IW}\left(\Lambda + \sum_{i=1}^n (m^{[r]} - X_i^{[r]})(m^{[r]} - X_i^{[r]})' + a(m^{[r]} - \mu)(m^{[r]} - \mu)', \nu + n + 1\right)$

2. $m^{[r+1]} | \dots \sim \mathcal{N}\left(\frac{a}{n+a}\mu + \frac{n}{n+a}\overline{\mathbf{X}_n^{[r]}} , \frac{C^{[r+1]}}{n+a}\right)$ where $\overline{\mathbf{X}_n^{[r]}}$ denotes the empirical mean of the n vectors $X_i^{[r]}, i = 1, \dots, n$

3. $\mathbf{X}^{[r+1]} | \dots \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^n \left[\left(X_i^{[r+1]} - m^{[r+1]}\right)' (C^{[r+1]})^{-1} \left(X_i^{[r+1]} - m^{[r+1]}\right) + \left(Y_i - H(X_i^{[r+1]}, d_i)\right)' R^{-1} \left(Y_i - H(X_i^{[r+1]}, d_i)\right) \right]\right\}$

which is not belonging to a closed form family of distributions. That is why a Metropolis-Hastings (MH) step is used to simulate $\mathbf{X}^{[r+1]}$ from its full conditional distribution.

Now, considering situations where extensive sampling of $H(X, d)$ is too time-consuming, we propose to replace H with a *maximin* LHD (Latin Hypercube Design) kriging emulator \hat{H} , following Barbillon (2010) [6]. This emulator is briefly described below.

- *Kriging* is a geostatistical method (Matheron 1971, [7]) that has been adapted by Sacks and al. (1989b, [20]) to approximate a physical model H on a bounded hypercube Ω . This method has known a growing interest in meta-modeling since the works of Koehler and Owen (1996, [16]), Santner and al. (2003, [21]) and Fang and al. (2006, [14]), among others. According to this approach the function H is regarded as the realization of a Gaussian Process (GP) $\mathcal{H} \sim \text{GP}(\mu, c)$, characterised by its mean and variance functions: $\mu(z) = \mathbb{E}[\mathcal{H}(z)]$ and $c(z, z') = \text{Cov}[\mathcal{H}(z), \mathcal{H}(z')] = \sigma^2 K_\epsilon(\|z - z'\|)$

for any $z = (x, d)$, K_ϵ being a symmetric positive kernel such that $K_\epsilon(0) = 1$. In a Bayesian perspective, GP modelling can be interpreted as providing H with a prior (Rasmussen & Williams, 2006, [18]). The process \mathcal{H} can be proved to be normally distributed knowing some evaluations $\mathbf{H}_{D_N} = \{H(z_{(1)}), \dots, H(z_{(N)})\}$ on a *design of experiments* $D_N = \{z_{(1)}, \dots, z_{(N)}\}$ of N points $z_{(j)} = (x_{(j)}, d_{(j)})$.

The best MSPE (*Mean Squared Prediction Error*) predictor of H , denoted by \hat{H} , is the conditional mean:

$$\hat{H}(z) = \mathbb{E}(\mathcal{H}(z) | \mathbf{H}_{D_N}), \forall z \in \Omega.$$

Then $\hat{H}(z)$ is minimizing the conditional expectation of the loss function $(\mathcal{H}(z) - \hat{H}(z))^2$, so-called MSE (Mean Squared Error) (see Johnson et al. 1990 for details, [5]),

$$\text{MSE}(z) = \mathbb{E}\left((\mathcal{H}(z) - \hat{H}(z))^2 | \mathbf{H}_{D_N}\right), \forall z \in \Omega.$$

- The set $D_N = \{z_{(1)}, \dots, z_{(N)}\}$ is chosen on $\Omega \in \mathbb{R}^{q+q_2}$ according to a *maximin* LHD (see McKay, Beckman, and Conover 1979, [13]): each dimension of the multidimensional domain Ω is divided into N intervals of equal length and the set D_N of N points are selected such that when projected on any dimension, each interval contains one and only one of the N projected points. Moreover, D_N is chosen to be *maximin*, i.e. it maximises

$$\delta_D = \min_{i \neq j} \|z_{(i)} - z_{(j)}\|$$

amongst the LHD of size N .

For the kriging version, considering the new *emulator error*, the conditional distribution of \mathbf{X} is as follows

$$\begin{aligned} & \pi(\mathbf{X} | \mathbf{Y}, m, C, \rho, \mathbf{H}_{D_N}) \propto \pi(\mathbf{X} | m, C, \rho, \mathbf{H}_{D_N}) \cdot \pi(\mathbf{Y} | \mathbf{X}, m, C, \rho, \mathbf{H}_{D_N}) \\ = & |\mathbf{R} + \text{MSE}(\mathbf{Z})|^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^n [(X_i - m)' C^{-1} (X_i - m)] \right. \\ & \left. -\frac{1}{2} \left((Y_1 - \hat{H}(Z_1))', \dots, (Y_n - \hat{H}(Z_n))' \right) (\mathbf{R} + \text{MSE}(\mathbf{Z}))^{-1} \begin{pmatrix} (Y_1 - \hat{H}(Z_1)) \\ \vdots \\ (Y_n - \hat{H}(Z_n)) \end{pmatrix} \right\}, \end{aligned} \quad (4)$$

where

$$\mathbf{R} = \begin{pmatrix} R_{11} & & & & & \\ & \ddots & & & & \\ & & R_{11} & & & \\ & & & \ddots & & \\ & & & & R_{pp} & \\ & & & & & \ddots \\ & & & & & & R_{pp} \end{pmatrix}, \quad \left. \begin{array}{l} \phantom{R_{11}} \\ \\ \phantom{R_{11}} \\ \\ \phantom{R_{pp}} \\ \\ \phantom{R_{pp}} \end{array} \right\} \begin{array}{l} n \text{ lines} \\ n \text{ lines} \end{array}$$

with R_{ii} the i -th diagonal component of the diagonal variance matrix R , and $\text{MSE}(\mathbf{Z})$ is the block diagonal matrix

$$\text{MSE}(\mathbf{Z}) = \left(\begin{array}{ccc} \text{MSE}_1(\mathbf{Z}) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \text{MSE}_p(\mathbf{Z}) \end{array} \right) \left. \begin{array}{l} \vphantom{\text{MSE}(\mathbf{Z})} \\ \vphantom{\text{MSE}(\mathbf{Z})} \\ \vphantom{\text{MSE}(\mathbf{Z})} \end{array} \right\} \begin{array}{l} n \text{ lines} \\ \\ n \text{ lines} \end{array}$$

composed with the variance matrices $\text{MSE}_j(\mathbf{Z}) \in \mathcal{M}^{n \times n}$ described as

$$\text{MSE}_j(\mathbf{Z}) = \mathbb{E} \left((\mathcal{H}_j(\mathbf{Z}) - \widehat{H}_j(\mathbf{Z}))^2 \mid \mathbf{H}_{D_N} \right),$$

for $j = 1, \dots, p$, where \mathbf{Z} denotes the n sample points $\{Z_1, \dots, Z_n\}$ with $Z_i = (X_i, d_i)$, \mathcal{H}_j denotes the j th dimension of the Gaussian process \mathcal{H} and \mathbf{H}_{D_N} is the evaluations of the function H on the design D_N . Simulating this conditional distribution of \mathbf{X} requires the Metropolis-Hastings (MH) step described in Appendix A.

Controlling the algorithmic error An important problem when running MCMC algorithms is monitoring the convergence of the simulated Markov chain in order to minimize the above mentioned *algorithmic error*. Actually, MCMC algorithms can converge slowly and stopping a simulated chain too early could lead to a poor approximation of the target distribution. Monitoring the convergence of a MCMC algorithm is also a difficult problem. Despite many efforts have been paid on this question, there is not an absolute way to answer it. We chose to use the much employed Brooks-Gelman (BG) statistics (Brooks et Gelman, 1998, [1]) computed from five replications of the Monte Carlo Markov chain (see Appendix B). The MCMC algorithm is stopped if the BG statistics is smaller than 1.05. We select this severe threshold of 1.05, instead of the more standard 1.2 value suggested in [1], to make sure that a reasonable approximation of the target distribution has been reached.

Measuring the emulator error However, a good monitoring of the MCMC algorithm could be jeopardized if the emulator \widehat{H} is too far from the model H (the *emulator error*). Typically, the emulator error can be large if the number of points N of the design D_N is too small. Two much employed criteria to measure the quality of a design are experimented here.

- i) The coefficient of predictability Q_2 (see Vanderpoorten and Palm, 2001, [25]) is

$$Q_2 = 1 - \frac{\text{PRESS}(D^*)}{\|H(D^*) - \overline{H}(D^*)\|^2}, \quad (5)$$

with

$$\text{PRESS}(D^*) = \|H(D^*) - \widehat{H}(D^*)\|^2$$

the Euclidean distance between the true function value H and the approximated value \widehat{H} on a validation sample $D^* = \{v_{(1)}, \dots, v_{(N^*)}\}$, $\overline{H}(D^*)$ denoting the mean function value

on D^* :

$$\bar{H}(D^*) = \frac{1}{N^*} \sum_{i=1}^{N^*} H(v_{(i)}).$$

A cheaper version of Q_2 can be obtained by cross-validation, as follows (*leave one out* procedure):

$$Q_{2\text{CV}} = 1 - \frac{\text{PRESS}_{\text{CV}}}{\sum_{i=1}^N \|H(z_{(i)}) - \bar{H}_{D_N}\|^2}. \quad (6)$$

where

$$\bar{H}_{D_N} = \frac{1}{N} \sum_{i=1}^N H(z_{(i)}),$$

and

$$\text{PRESS}_{\text{CV}} = \sum_{i=1}^N e_{(i)}^2 = \sum_{i=1}^N \|H(z_{(i)}) - \hat{H}_{-i}(z_{(i)})\|^2$$

with

- $e_{(i)}$ is the prediction error at $z_{(i)}$ of a fitted model without the point $z_{(i)}$;
- $\hat{H}_{-i}(z_{(i)})$ is the approximation of H at $z_{(i)}$ derived from all the points of the design except $z_{(i)}$.

Both versions of Q_2 are related to the ratio of variance explained by an emulator. The closer Q_2 to 1, the smaller this ratio is and the better the quality of the design D_N is.

- ii) An alternative criterion is the Mahalanobis distance (MD) (see Bastos and O'Hagan 2009, [4]), computed on a validation sample D^* with N^* points as follows:

$$\text{MD} = \left(H(D^*) - \hat{H}(D^*) \right)' \left(\text{MSE}(D^*) \right)^{-1} \left(H(D^*) - \hat{H}(D^*) \right), \quad (7)$$

where $\text{MSE}(D^*)$ (Mean Squared Error) is the conditional variance matrix of the design D^* knowing $H_{D^*} = \{H(v_{(1)}), \dots, H(v_{(N^*)})\}$. An interest of this criterion is to take into account the correlations between the points through the $\text{MSE}(D^*)$ term. Obviously, the MD value is sensitive to the choice of D^* . D^* could be generated as a *maximin* LHD. A cheaper cross-validated version of MD is as follows:

$$\text{MD}_{\text{CV}} = \frac{1}{N} \sum_{i=1}^N \left(H(z_{(i)}) - \hat{H}_{-i}(z_{(i)}) \right)' \left(\text{MSE}_{-i}(z_{(i)}) \right)^{-1} \left(H(z_{(i)}) - \hat{H}_{-i}(z_{(i)}) \right),$$

where $\hat{H}_{-i}(z_{(i)})$ denotes the predictor of H at point $z_{(i)}$ by using the design $D_{-i} = \{z_{(1)}, \dots, z_{(i-1)}, z_{(i+1)}, \dots, z_{(N)}\}$ and $\text{MSE}_{-i}(z_{(i)})$ denotes the related squared error.

Now, the smaller the sample size n , the greater the *estimation error* is. The two above mentioned criteria are not aiming to measure the *estimation error*. But since H is complex, it is quite difficult to assess this error in an inverse modelling context. Bayesian inference could be expected to be helpful to reduce the estimation error when n is small and when reliable prior information is available. However, if the prior information is not relevant, the *prior error* will be large and Bayesian inference could be harmful. For this very reason, it is important to be able to measure the relevance of the prior information. In the present context, it is possible to use a promising criterion, the so-called DAC criterion (Bousquet 2008, [2]) for this task, as detailed in the next section.

3 Assessing a prior distribution and a design

3.1 The DAC criterion

The DAC criterion (Bousquet 2008, [2]) has been conceived as a measure of the discrepancy between a prior distribution of a model parameters and the data. Let \mathbf{y} be a sample with pdf $f(\mathbf{y}|\theta)$. Let $\pi^J(\theta)$ be a benchmark non-informative prior (see for instance Yang and Berger 1998, [22]) and $\pi(\theta)$ the prior distribution derived from the prior information on θ . DAC is

$$\text{DAC}(\pi|\mathbf{y}) = \frac{\text{KL}(\pi^J(\theta|\mathbf{y})||\pi(\theta))}{\text{KL}(\pi^J(\theta|\mathbf{y})||\pi^J(\theta))}, \quad (8)$$

where $\text{KL}(p||q)$ is denoting the Kullback-Leibler distance between the probability distributions p and q , which is defined as

$$\text{KL}(p||q) = \int_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)} dx, \quad (9)$$

\mathcal{X} being the set of all accessible values for x . The rationale underlying DAC criterion is as follows: the posterior distribution $\pi^J(\theta|\mathbf{y})$ derived from the non-informative prior can be regarded as an ideal prior distribution on θ in perfect agreement with the data \mathbf{y} . Thus, $\text{KL}(\pi^J(\theta|\mathbf{y})||\pi(\theta))$ is measuring the distance between the prior π to be assessed and the ideal prior $\pi^J(\cdot|\mathbf{y})$.

If $\text{DAC}(\pi|\mathbf{y}) \leq 1$, the informative prior π is closer to the ideal prior than the non-informative prior π^J , and the data \mathbf{y} and the prior $\pi(\theta)$ are declared to be in agreement. Otherwise if $\text{DAC}(\pi|\mathbf{y}) > 1$, the data \mathbf{y} and the prior $\pi(\theta)$ are declared to be discrepant. DAC has been proved to be efficient when the non-informative prior $\pi^J(\theta)$ is proper (see Bousquet 2008, [2]).

3.2 The impact of the emulator

In the present context, a kriging emulator defined on a compact set Ω is used to compute an approximation of the posterior distribution of the parameter $\theta = (m, C)$. Since the emulator is defined on a compact set, the parameters m and C are also restricted to be in compact sets Ω_m and Ω_C . It allows us to define a proper non-informative prior $\pi^J(m, C)$ (chosen as the Jeffreys prior for the multivariate Gaussian model), then a tractable DAC. The technical precisions about Ω_m , Ω_C and the calculation of DAC are provided in Appendices C and D.

It is important to remark that the DAC criterion is depending on the design D_N . Denoting $\pi^J(\theta|\mathbf{y}, D_N)$ the posterior distribution of θ given the data \mathbf{y} and the current design $D_N = \{z_{(1)}, \dots, z_{(N)}\}$,

$$\text{DAC}(\pi|\mathbf{y}, \mathbf{H}_{D_N}) = \frac{\text{KL}(\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})||\pi(\theta))}{\text{KL}(\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})||\pi^J(\theta))}.$$

A DAC value greater than one is just indicating that there is something misleading between the data, the prior and the design. Thus, if the data and the prior are known (or assumed) to be relevant, DAC could be regarded as a criterion to assess the design as Q_2 or MD.

3.3 Computing DAC

Since $\frac{S}{T} \leq 1 \iff S - T \leq 0$, if $S \geq 0, T > 0$, a numerically more convenient version of DAC, denoted by $\widetilde{\text{DAC}}$ is

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

The critical value for $\widetilde{\text{DAC}}$ is 0. Since the support of $\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})$ is Ω , we have

$$\begin{aligned} \text{KL}(\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})||\pi(\theta)) &= \int_{\Omega} \pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N}) \log \frac{\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})}{\pi(\theta)} d\theta \\ &= \mathbb{E}_{\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})} [\log \pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})] - \mathbb{E}_{\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})} [\log \pi(\theta)], \end{aligned}$$

and

$$\begin{aligned} \text{KL}(\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})||\pi^J(\theta)) &= \int_{\Omega} \pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N}) \log \frac{\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})}{\pi^J(\theta)} d\theta \\ &= \mathbb{E}_{\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})} [\log \pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})] - \mathbb{E}_{\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})} [\log \pi^J(\theta)]. \end{aligned}$$

Therefore, the transformed $\widetilde{\text{DAC}}$ can be written as:

$$\begin{aligned} \widetilde{\text{DAC}}(\pi|\mathbf{y}, \mathbf{H}_{D_N}) &= \text{KL}(\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})||\pi(\theta)) - \text{KL}(\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})||\pi^J(\theta)) \\ &= \mathbb{E}_{\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})} [\log \pi^J(\theta)] - \mathbb{E}_{\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})} [\log \pi(\theta)], \end{aligned}$$

and the $\widetilde{\text{DAC}}$ criterion can be computed using the outputs of a Gibbs sampler run with a non-informative prior $\pi^J(\cdot)$ (In practice, we chose a Jeffreys non-informative prior.)

$$\widetilde{\text{DAC}}(\pi|\mathbf{y}, \mathbf{H}_{D_N}) \simeq \frac{1}{R} \sum_{r=1}^R \log \pi^J(\theta^r) - \frac{1}{R} \sum_{r=1}^R \log \pi(\theta^r), \quad (10)$$

where $\theta^r \sim \pi^J(\cdot|\mathbf{y}, \mathbf{H}_{D_N})$, $r \in \{1, \dots, R\}$ is a simulated sequence obtained by Gibbs sampling. For the purpose of simplicity, in the following we use the notation

$$\widetilde{\text{DAC}}_N := \widetilde{\text{DAC}}(\pi|\mathbf{y}, \mathbf{H}_{D_N}).$$

$\widetilde{\text{DAC}}_N \leq 0$ means the prior distribution $\pi(\theta)$ and the couple $(\mathbf{y}, \mathbf{H}_{D_N})$ are declared compatible. Now, computing DAC criterion requires to run an additional Gibbs sampler with the non-informative prior distribution. Denoting $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, the full conditional distribution of m verifies

$$\pi^J(m|C, \mathbf{Y}, \mathbf{X}, \rho, H_D) \propto \mathbf{I}_{\Omega_m} \exp \left[-\frac{1}{2} (m - \bar{X}_n)' \left(\frac{C}{n} \right)^{-1} (m - \bar{X}_n) \right].$$

Thus, it is a normal distribution truncated on Ω_m : $\mathbf{I}_{\Omega_m} \cdot \mathcal{N}(\bar{X}_n, \frac{C}{n})$. The full conditional distribution of the variance matrix C verifies

$$\pi^J(C|m, \mathbf{Y}, \mathbf{X}, \rho, H_D) \propto \mathbf{I}_{\Omega_C} |C|^{-\frac{n+q+2}{2}} \exp \left[-\frac{1}{2} \text{Tr} \left(n(m - \bar{X}_n)(m - \bar{X}_n)' \cdot C^{-1} \right) \right]. \quad (11)$$

Thus it is an Inverse-Wishart distribution truncated on Ω_C :

$$\mathbf{I}_{\Omega_C} \cdot \mathcal{IW}(n(m - \bar{X}_n)(m - \bar{X}_n)', n + 1). \quad (12)$$

Moreover, the full conditional distribution of the missing data \mathbf{X} given the current parameters θ , the observation \mathbf{y}, \mathbf{d} and the evaluations \mathbf{H}_{D_N} is given by (4).

Using those full conditional posterior distributions, the Gibbs sampler approximating the posterior distribution of (m, C) with a non-informative prior truncated to the domain $\Omega_m \times \Omega_C$ could be straightforwardly described. Obviously, it incorporates the MH step presented in Appendix A to simulate the missing data \mathbf{X} .

Remark: The simulation of C is difficult since $n(m - \bar{X}_n)(m - \bar{X}_n)'$ is not a definite but a semi-definite positive matrix and numerical problems can occur. However, up to an additive constant, the calculation (11) is proper. For this reason, we recommend to use a Metropolis-Hastings algorithm for simulating C :

Metropolis-Hasting (MH) algorithm

1. Iteration 0: Choose an arbitrary value $C^{[0]} = C_0$
2. Iteration h : Update $C^{[h]}$ as follows:

- Generate ξ from the following proposal distribution f^* , which is adding a small correction ϵI_q to the semi-positive definite matrix $(m - \bar{X}_n)(m - \bar{X}_n)'$

$$f^*(\xi) = \mathbf{I}_{\Omega_C}(\xi) \cdot \mathcal{IW}(n(m - \bar{X}_n)(m - \bar{X}_n)' + \epsilon I_q, n + 1).$$

- Let

$$\alpha(C^{[h-1]}, \xi) = \frac{g(\xi)f^*(C^{[h-1]})}{g(C^{[h-1]})f^*(\xi)} \wedge 1, \quad (13)$$

with g proportional to the target distribution which means the truncated Inverse-Wishart distribution (12)

$$g(C) = \mathbf{I}_{\Omega_C}(C) \cdot |C|^{-\frac{n+q+2}{2}} \exp\left[-\frac{1}{2}\text{Tr}(n(m - \bar{X}_n)(m - \bar{X}_n)' \cdot C^{-1})\right].$$

- Choose $C^{[h]}$ as follows

$$C^{[h]} = \begin{cases} \xi & \text{with probability } \alpha(C^{[h-1]}, \xi), \\ C^{[h-1]} & \text{otherwise.} \end{cases} \quad (14)$$

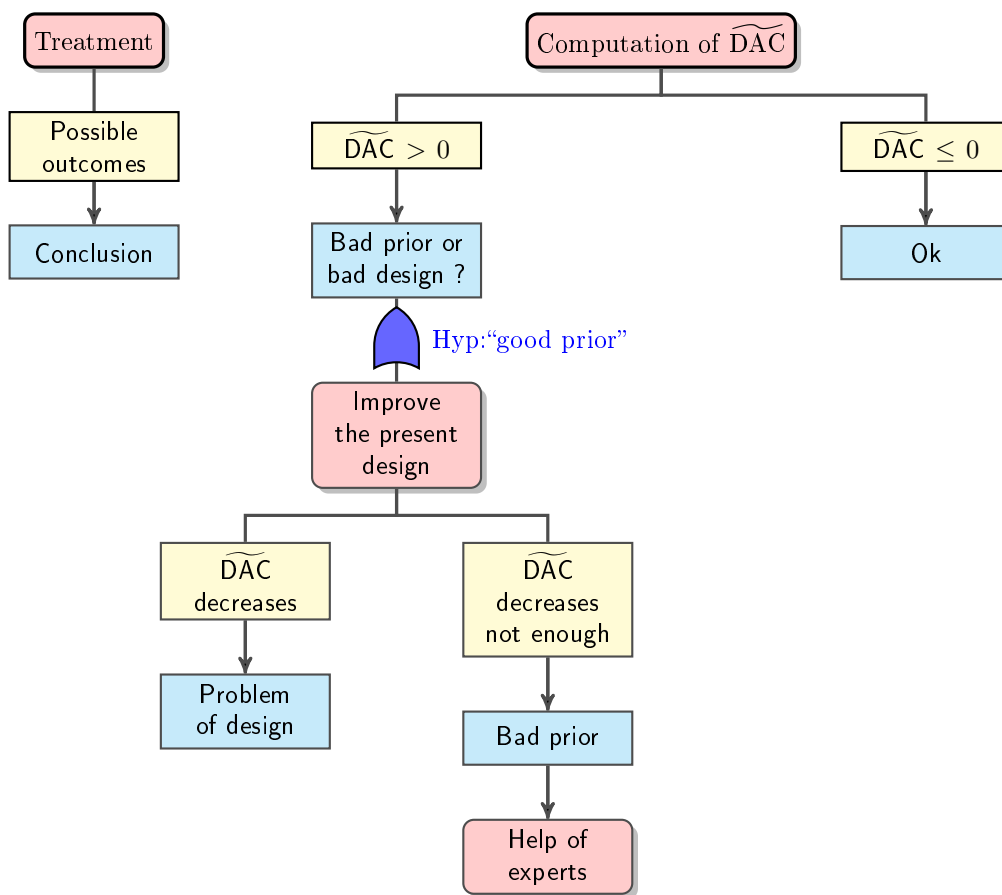
In this way, we obtain a Markov chain $(C^{[h]})$ which converges to the distribution (12).

3.4 Using the $\widetilde{\text{DAC}}$ criterion

By its very nature, the criterion $\widetilde{\text{DAC}}$ is measuring the agreement between the observed data and the prior distribution. As shown above, it could be computed without particular difficulties, despite it needs to run an additional Gibbs sampler, when the distribution H has been replaced by a kriging emulator \widehat{H} . Thus $\widetilde{\text{DAC}}$ is depending on the prior distribution and the design D_N . Hence $\widetilde{\text{DAC}}$ is a criterion allowing to assess both the prior and design relevances with respect to the observed data y . But this double assessment has to be done properly using the following procedure:

1. If $\widetilde{DAC} \leq 0$ then the prior and the design are declared to be acceptable.
2. If $\widetilde{DAC} > 0$, the following step is required:
 under a “good prior” assumption, efforts are paid to improve the design by increasing N or modifying Ω . If \widetilde{DAC} is not decreasing under zero, it means that the prior information is questionable and there is the need to go back to the experts.

This procedure is depicted by the following diagram:



4 Numerical experiments

In order to illustrate the behavior of the the above mentioned criteria, numerical experiments are performed from simulated data on a simplified version of a hydraulic model $Y = H(X, d) + U$ partly used in [17] where

$$H(X, d) = \left(X_2 + \left(\frac{\sqrt{5000}}{300\sqrt{55 - X_2}} \times \frac{d}{X_1} \right)^{0.6}, \frac{d^{0.4} X_1^{0.6} (55 - X_2)^{0.3}}{300^{0.4} \times 5000^{0.3}} \right),$$

with

$$X \sim \mathcal{N}\left(\begin{pmatrix} 30 \\ 50 \end{pmatrix}, \begin{pmatrix} 5^2 & 0 \\ 0 & 1 \end{pmatrix}\right),$$

$$d \sim \text{Gumbel}(1013, -458),$$

and the error term $U \sim \mathcal{N}(\mathbf{0}, 10^{-5} \cdot I_2)$.

Since we are mainly concerned in analyzing the behavior of DAC, six different prior distributions on the model parameters were considered. They are summarized in Table 4. Remind that the prior distributions on the parameters m and C are $m|C \sim \mathcal{N}(\mu, C/a)$ and $C \sim \mathcal{IW}(\Lambda, \nu)$ with $\Lambda = t \cdot \tilde{C}_{\text{Exp}}$.

Prior	PLV			PMV	PHV	FHV			BMV	BHV
μ	{30, 50}			{30, 50}	{30, 50}	{35, 49}			{10, 54}	{10, 54}
\mathbf{a}	1	10	10	1	1	1	5	10	1	1
\mathbf{t}	2	2	30	2	2	2			2	2
ν	5	5	33	5	5	5			5	5
\tilde{C}_{Exp}	$\begin{pmatrix} 1.5^2 & 0 \\ 0 & 1 \end{pmatrix}$			$\begin{pmatrix} 5^2 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 7.5^2 & 0 \\ 0 & 1.5^2 \end{pmatrix}$	$\begin{pmatrix} 7.5^2 & 0 \\ 0 & 1.5^2 \end{pmatrix}$			$\begin{pmatrix} 5^2 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 7.5^2 & 0 \\ 0 & 1.5^2 \end{pmatrix}$

Table 1: Description of the six prior distributions: **PLV** = perfect mean and low variance, **PMV** = perfect mean and medium variance, **PHV** = perfect mean and high variance, **FHV** = fair mean and high variance, **BMV** = bad mean and medium variance, **BHV** = bad mean and high variance.

4.1 Assessing the design

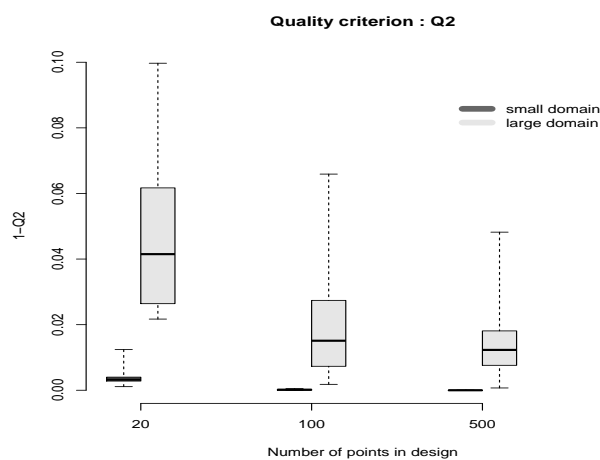
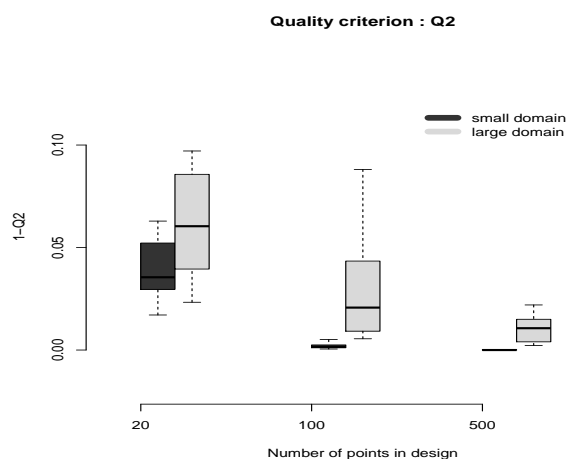
The first experiments are aiming to assess the ability of criteria Q_2 and MD to measure the quality of a design. In this purpose three different designs with 20 points, 100 points and 500 points have been considered on two different domains

$$\Omega_1 = [25.1001, 34.8999] \times [48.0400, 51.9600] \times [40, 1800]$$

$$\Omega_2 = [20, 40] \times [45, 55] \times [\min_i(d_i), \max_i(d_i)].$$

Ω_1 can be thought of as a realistic domain and Ω_2 is a larger domain. When using a validation sample D^* , we choose it as a *maximin* LHD with 100 points. Figures 1 and 2 give the box plots of $1 - Q_2$ based on 20 repetitions computed on a validation sample and by cross-validation respectively. The closer one and Q_2 are, the better the design is supposed to be. The observed differences on $1 - Q_2$ according to the designs are relevant but hardly perceptible as even a small design of 20 points on the large domain Ω_2 produces small $1 - Q_2$ values. The difficulty with criterion Q_2 is to choose a sensible threshold to declare that a design is acceptable.

Figures 3 and 4 display the box plots of $\log(\text{MD})$ in the same conditions. As it could be expected, this criterion is decreasing when the number of design points increases. The cross-validated MD does not seem very sensitive for the domain Ω_1 and the cross-validated MD values for the larger domain with a design of 500 points are amazing (see Figure 4). Moreover, contrary to Q_2 criterion, no reference value is available with MD and it seems difficult to use this more expensive criterion to assess a design (see Figure 3).


 Figure 1: $1 - Q_2$ box plots calculated on a validation sample for six designs

 Figure 2: $1 - Q_2$ box plots calculated by cross-validation for six designs

4.2 Assessing the prior and the design

The following numerical experiments aim to analyze the ability of $\widetilde{\text{DAC}}$ to assess either a design or a prior distribution.

Figures 5 and 6 depict the behavior of $\widetilde{\text{DAC}}$ for the domains Ω_1 and Ω_2 , for 20 repetitions of the model with the six prior distributions and *maximin* LHDs with 20, 100 and 500 points. From those figures, it appears that the "bad" priors are discarded by $\widetilde{\text{DAC}}$ in all cases, but for the other priors even the design with 20 points seems acceptable. Obviously, for this poor design the Gibbs sampler converges dramatically slower (2 000 iterations for D_{500} and 100 000 iterations for D_{20}), but in many situations it is not an issue. Actually, the main computational burden is computing the highly CPU-time demanding physical model H . In the present context, running a Gibbs sampler with a design of N points require N calls to the function H and it could be faster

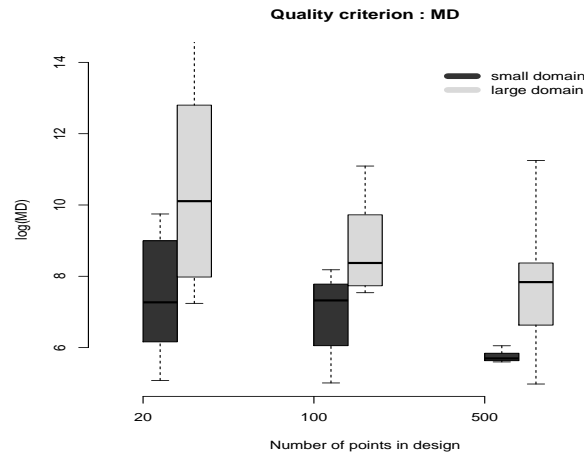


Figure 3: MD box plots calculated on a validation sample for six designs

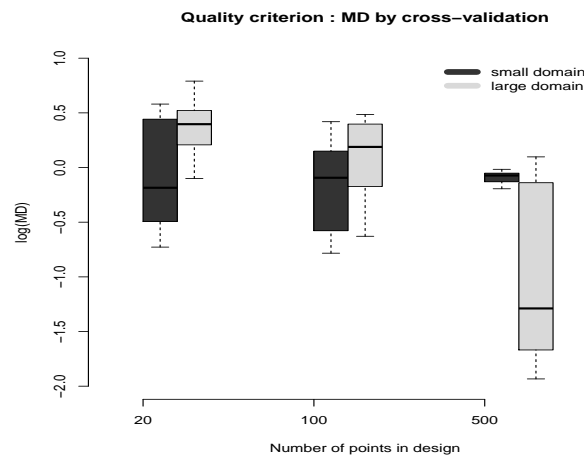


Figure 4: MD box plots calculated by cross-validation for six designs

to run a Gibbs sampler on a D_{20} for 100 000 iterations than a Gibbs sampler with a D_{500} for 2 000 iterations. . . Moreover, the similar behavior of $\widetilde{\text{DAC}}$ in Figures 5 and 6 for both domains shows that the domain choice does not affect the agreement between the prior and the data. For this reason, we only reported the next experiments for the small domain Ω_1 .

Figure 7 which displays the behavior of $\widetilde{\text{DAC}}$ for the PLV and FHV prior with different hyperparameters a and t values shows that those hyperparameters are sensitive and that too concentrated priors (related to large values of a and t) could lead to a doubtful Bayesian inference. For example, for the PLV prior increasing the value of a , which weights the prior mean μ , does not much change the value of $\widetilde{\text{DAC}}$ as μ is equal to the actual mean m ; while for the FHV prior, a larger a results in a larger $\widetilde{\text{DAC}}$ value as in this “fair” case, μ and the actual mean m are different.

Figures 8 displays the marginal posterior distributions with a LHD-*maximum* design of 100

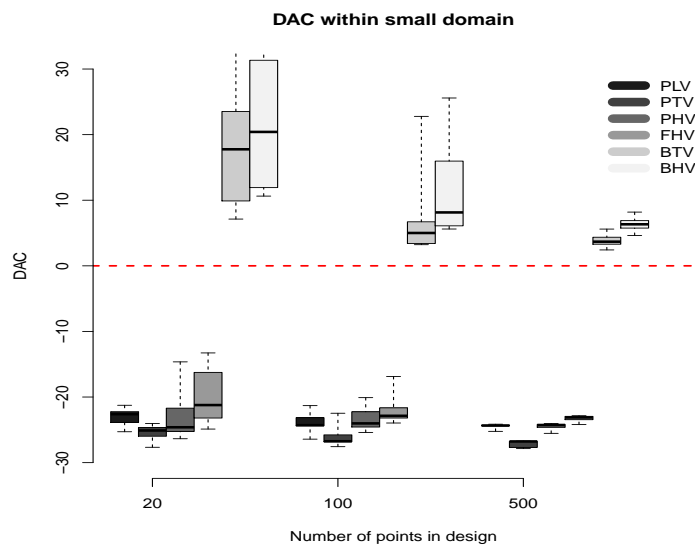


Figure 5: $\widetilde{\text{DAC}}$ in small domain with six priors and three designs

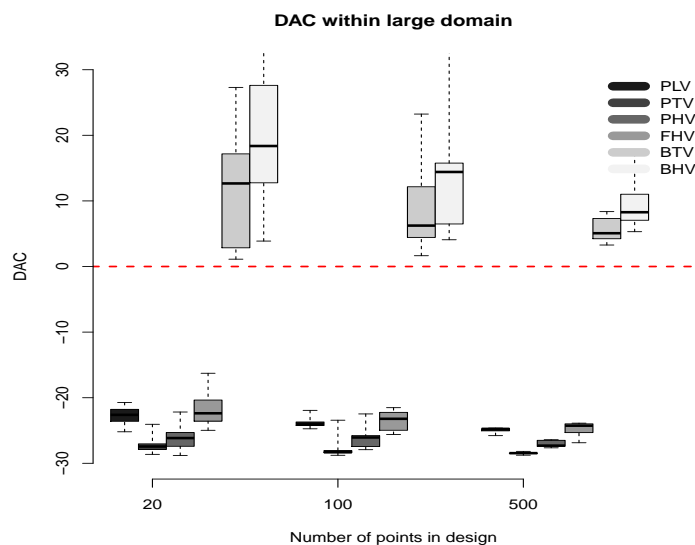


Figure 6: $\widetilde{\text{DAC}}$ in large domain with six priors and three designs

points and Figure 9 with a LHD-*maximum* design of 20 points. Those figures confirm the $\widetilde{\text{DAC}}$ diagnosis. There are great differences between the posteriors derived from "bad priors" and the other ones, including the posterior derived from the Jeffreys prior, which are quite similar. It is also important to notice that there are no sensitive differences between the posteriors derived from the 100 points and 20 points designs as indicated by the $\widetilde{\text{DAC}}$ criterion.

It seems that $\widetilde{\text{DAC}}$ is indicating that a reasonable prior can resist to a poor design. It is

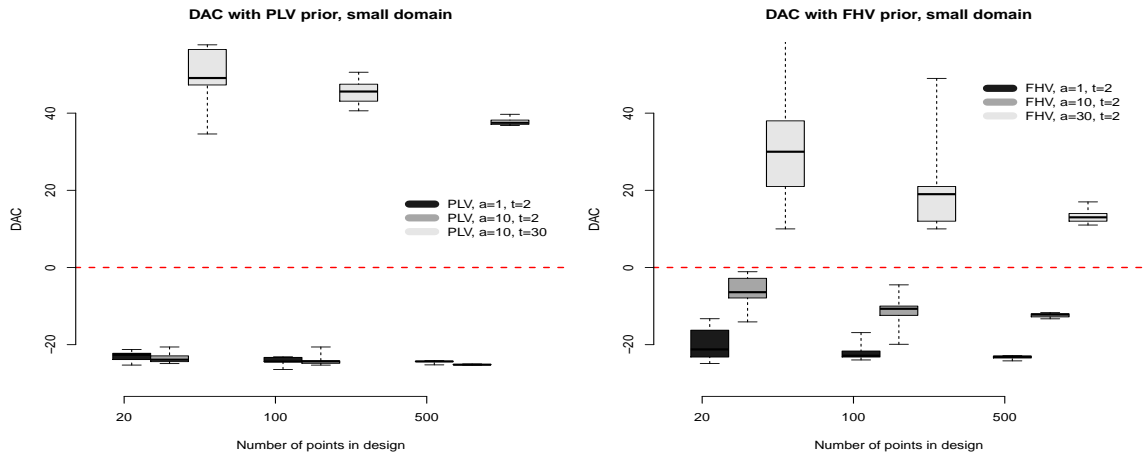


Figure 7: $\widehat{\text{DAC}}$ with **PLV** and **FHV** priors for the small domain Ω_1 and different values of the hyperparameters a and t

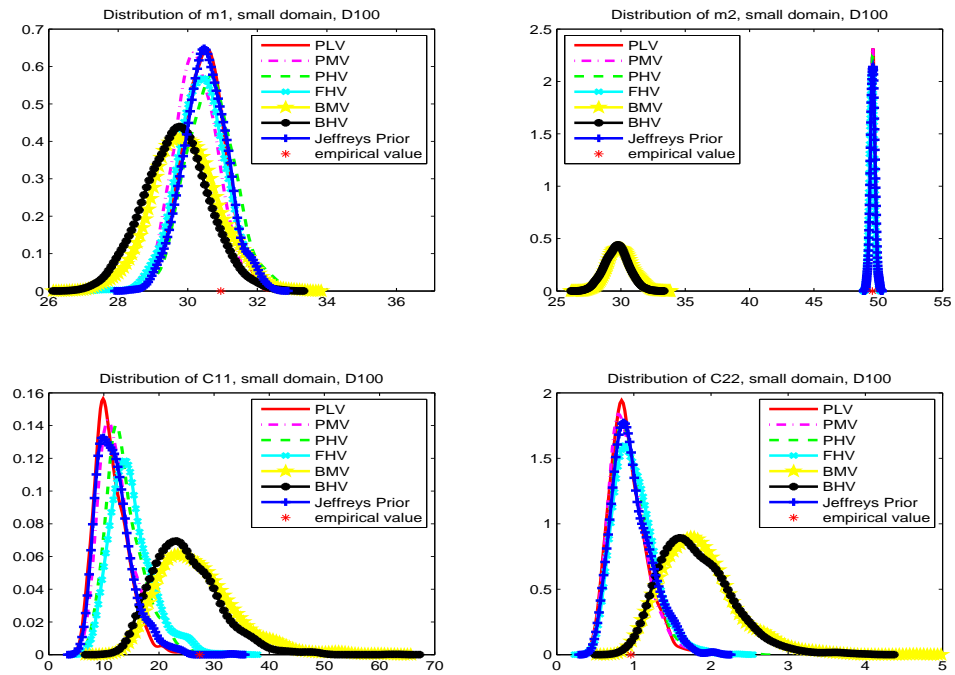
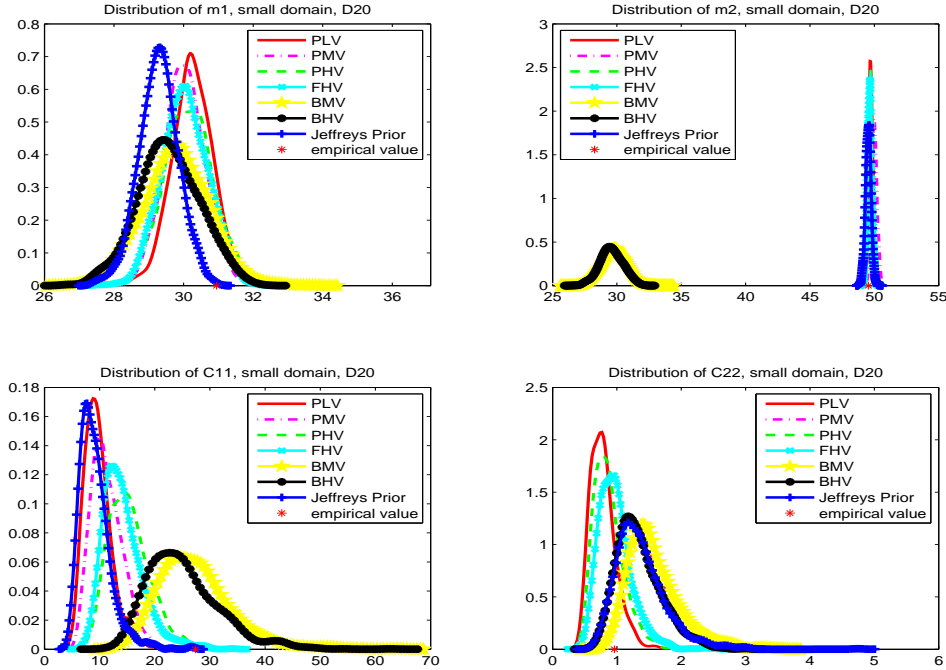


Figure 8: Posterior of θ with help of D_{100} in small domain Ω_1

not always true. For instance a poor design of 18 randomly generated points on the faces of a cube (Three points were generated on each face.) has been considered with the same model


 Figure 9: Posterior of θ with help of D_{20} in small domain Ω_1

$Y = H(X, d) + U$ but with the Sobol function H

$$H(X, d) = \prod_{k=1}^2 g_k(|\sin(X_k)|) g_3(|\sin(d)|), \text{ where } g_k(x) = \frac{|4x - 2| + a_k}{1 + a_k},$$

with $a_k = 1$. A Gibbs sampler of 800 000 runs has been run to estimate the posterior distribution $\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})$. As shown in the left graph of Figure 10, $\widetilde{\text{DAC}}_{18}$ remains positive for the four prior choices indicating the need to improve the design.

5 Discussion

We have shown that Bayesian analysis was possible and beneficial to solve inverse problems by estimating the parameters of highly complex uncertainty models. Bayesian analysis is possible thanks to MCMC algorithms such as Gibbs sampling and the approximation of the physical model by a kriging emulator using a *maximin* LHD. Bayesian analysis is beneficial since it allows to take into account properly prior knowledge and avoids linearization of the physical model H . Our analysis has shown that Bayesian inference could be beneficial because MCMC algorithms could be hoped to be rapid even with a *maximin* LHD with few points in comparison to the huge time needed to compute H . From this point of view, it could be helpful to translate the time to get a realization of H as a number of iterations of the MCMC algorithm in order to choose the number of points of the emulator's design. Let us suppose that the computation time of one

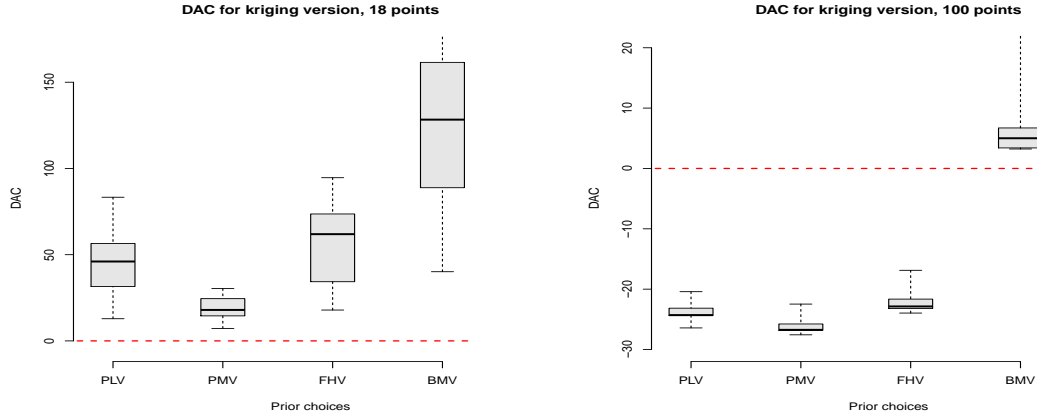


Figure 10: $\widetilde{\text{DAC}}$ for D_{18} and D_{100} with the Sobol function

call to H equals the computation time of $L(N)$ iterations of the MCMC algorithm¹. The integer $L(N)$ is expected to be quite large and is a decreasing function of the number N of points of the design D_N which is as well the number of "possible" calls to H . Our analysis proved that even when N is small, it is possible to increase the number of iterations of the MCMC algorithm to get a good approximation of the model parameter posterior distribution in an acceptable CPU time. For instance, with the hydraulic model, the CPU time (in seconds) has been 999 for $N = 500$, 1930 for $N = 100$ and 10 100 for $N = 20$ on a laptop PC, with two Intel P9700 cores of 2.80GHz

In this perspective, there is the need to control the four error sources listed in the introduction.

- By its very nature, Bayesian inference is helpful to control the *estimation error* when the number n of observations is small.
- The *algorithmic error* can be efficiently controlled with the BG statistics. To make sure that this error is not too big, we advocate a more severe threshold value 1.05 than the standard threshold 1.2.
- We propose to use the so-called $\widetilde{\text{DAC}}$ criterion which could be thought of as a relevant measure of the discrepancy between the observed sample and the prior distribution in order to control both the *emulator error* and the *prior error*. In our context, this criterion can be computed without major difficulties: the emulator is defined on a compact set and, consequently, proper non-informative priors are available. Our experiments show a promising behavior of this criterion. Obviously, computing $\widetilde{\text{DAC}}$ is not free since it involves to run an additional MCMC algorithm for non-informative priors. But we think that the result is worth the trouble. Moreover, as soon as the MCMC with a non-informative prior has been run, any informative prior can be assessed. On the other hand when $\widetilde{\text{DAC}}$ is greater than zero, it could be difficult to separate the *emulator* and the *prior* errors since both errors could be quite intricate. More experiments are needed to assess the relevance and sensibility of this criterion. But we think that it is a promising tool to drive Bayesian inference using an emulator for dealing with complex inverse problems in uncertainty analysis.

¹Recall that N is the total allowed number of calls to H and it is also the number of points of the emulator design.

Finally, the conclusion of this study can be stated as follows. When the prior knowledge on the model parameters is relevant, Gibbs sampling or other MCMC algorithms on an appropriate emulator could be expected to lead to a sensible estimation of these parameters with well calibrated prior distributions while dramatically saving the number of calls to the expensive function H . And, the criterion $\widetilde{\text{DAC}}$ could be expected to be helpful to honestly calibrate the prior distributions and choose a good design for the emulator.

A the Gibbs sampling with a Jeffreys non-informative prior

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

1. $C^{[r+1]} | m^{[r]}, \mathbf{X}^{[r]}, \mathcal{Y}, \rho, \mathbf{H}_{D_N} \sim \mathcal{IW}\left(\sum_{i=1}^n (m^{[r]} - X_i^{[r]})(m^{[r]} - X_i^{[r]})', n + 1\right) \cdot \mathbb{1}_{\Omega_C}$;
2. $m^{[r+1]} | C^{[r+1]}, \mathbf{X}^{[r]}, \mathcal{Y}, \rho, \mathbf{H}_{D_N} \sim \mathcal{N}\left(\overline{\mathbf{X}^{[r]}}^{\top}, \frac{C^{[r+1]}}{n}\right) \cdot \mathbb{1}_{\Omega_m}$,
with $\overline{\mathbf{X}^{[r]}} = \sum_{i=1}^n X_i^{[r]}$;
3. $\mathbf{X}^{[r+1]} | m^{[r+1]}, C^{[r+1]}, \mathcal{Y}, \rho, \mathbf{H}_{D_N} \Rightarrow$ Metropolis-Hastings algorithm
More precisely, this full conditional posterior distribution is proportional to

$$\begin{aligned} & |\mathbf{R} + \text{MSE}^{[r+1]}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2} \sum_{i=1}^n (X_i^{[r+1]} - m^{[r+1]})' [C^{[r+1]}]^{-1} (X_i^{[r+1]} - m^{[r+1]})\right. \\ & \left. - \frac{1}{2} \left((\mathcal{Y}_1 - \widehat{H}_{N,1}^{[r+1]})', \dots, (\mathcal{Y}_n - \widehat{H}_{N,n}^{[r+1]})' \right) (\mathbf{R} + \text{MSE}^{[r+1]})^{-1} \begin{pmatrix} \mathcal{Y}_1 - \widehat{H}_{N,1}^{[r+1]} \\ \vdots \\ \mathcal{Y}_n - \widehat{H}_{N,n}^{[r+1]} \end{pmatrix} \right\}, \end{aligned}$$

with $\widehat{H}_{N,i}^{[r+1]} = \widehat{H}_N(X_i^{[r+1]}, d)$ and $\text{MSE}^{[r+1]} = \text{MSE}(\mathbf{X}^{[r+1]}, d)$.

The Metropolis-Hastings step inside the Gibbs sampler At step $r+1$ of Gibbs sampling, after simulating $m^{[r+1]}, C^{[r+1]}$, the missing data $\mathbf{X}^{[r+1]}$ have to be updated with a Metropolis-Hasting (MH) algorithm. The MH step is updating $\mathbf{X}^{[r]} = (X_1, \dots, X_n)'$ in the following way:

- For $i = 1, \dots, n$

1. Generate $\widetilde{X}_i \sim J(\cdot | X_i^r)$ where J is the proposal distribution.
2. Let

$$\alpha(X_i^r, \widetilde{X}_i) = \min\left(\frac{\pi_{\widehat{H}}(\widetilde{\mathbf{X}} | \mathcal{Y}, \theta^{[r+1]}, \rho, \mathbf{d}, H_D) J(X_i^r | \widetilde{X}_i)}{\pi_{\widehat{H}}(\mathbf{X}^{[r]} | \mathcal{Y}, \theta^{[r+1]}, \rho, \mathbf{d}, H_D) J(\widetilde{X}_i | X_i^r)}, 1\right),$$

where

$$\begin{aligned} \widetilde{\mathbf{X}} &= (X_1^{r+1}, \dots, X_{i-1}^{r+1}, \widetilde{X}_i, X_{i+1}^r, \dots, X_n^r)' \\ \mathbf{X}^{[r]} &= (X_1^r, \dots, X_{i-1}^r, X_i^r, X_{i+1}^r, \dots, X_n^r)' \end{aligned}$$

3. Take

$$X_i^{r+1} = \begin{cases} \widetilde{X}_i & \text{with probability } \alpha(X_i^r, \widetilde{X}_i), \\ X_i^{r+1} & \text{otherwise.} \end{cases}$$

Remarks:

- Many choices are possible for the proposal distribution J . It appears that choosing an independent MH sampler with J chosen to be the normal distribution $\mathcal{N}(m^{[r+1]}, C^{[r+1]})$ give satisfying results for the model (1).
- In practice, it can be beneficial to choose the order of the updates by a random permutation of $\{1, \dots, n\}$ to accelerate the convergence of the Markov chain to its limit distribution.

B Brooks-Gelman Statistics

In 1998, Brooks and Gelman proposed a method derived from the method proposed by Gelman and Rubin (1992a), for monitoring the convergence of iterative simulations ([1]). Supposing m parallel chains have been simulated, the statistic \widehat{R}_{BG} is constructed on the final M iterations after the “burn-in” period, as follows:

1. For each individual chain j , calculate the empirical $100(1 - \alpha)\%$ interval δ_j , which is the difference between the $100(1 - \frac{\alpha}{2})\%$ and $100\frac{\alpha}{2}\%$ percentile of the M simulated points. Thus, form the m within-sequence interval length estimates.
2. For the entire set of mM simulated draws from all chains, calculate the empirical $100(1 - \alpha)\%$ interval to construct a total-sequence interval length estimate.
3. Evaluate the statistic \widehat{R}_{BG} defined as

$$\widehat{R}_{BG} = \frac{\Delta}{\bar{\delta}}$$

- Δ the total-sequence interval length;
- $\bar{\delta} = \frac{1}{m} \sum_{j=1}^m \theta_j$, with θ_j the length of the within-sequence interval for the j -th chain.

The threshold value 1.2 is advocated by the authors ($\widehat{R}_{BG} < 1.2$) to declare that the simulation procedure has converged. In our experiments, we make use of a more conservative threshold and procedure to ensure that the MCMC algorithms have converged to their stationary distribution. A MCMC chain has been declared to have converged if the \widehat{R}_{BG} statistics is smaller than 1.05 for 3,000 iterations.

C Computing DAC for the kriging emulator

The compact set $\Omega_m = \Omega = \Omega_1 \times \dots \times \Omega_q$ where Ω_i denotes the domain for the i -th coordinate of X . To determine the compact set Ω_C related to the variance matrix C , it is convenient to consider its eigenvalue decomposition $C = VDVT$ where D is the diagonal matrix of eigenvalues of C with $|C| = |D|$ and V the orthogonal matrix of eigenvectors of C . For each dimension $i = 1, \dots, q$, $X_i^2 \leq \beta_i = \max(\max \Omega_i)^2, \min \Omega_i)^2$. On the other hand, recalling that R is the variance matrix of the measurement error in model (1), it is reasonable to assume that the measurement error is smaller than the variance and thus $|R|^{1/p} \leq |C|^{1/q} = |D|^{1/q}$. Finally, the domain of variance Ω_C can be defined as follows:

$$\Omega_C = \left\{ C = VDVT \in \mathcal{S}_q^+ \text{ st. } |D| \geq |R|^{q/p}, 0 \leq D_{ii} \leq \sqrt{\sum_{j=1}^q \beta_j^2}, i = 1 \dots, q \right\}, \quad (15)$$

where \mathcal{S}_q^+ is the set of symmetric positive definite matrices of rank q .

The benchmark prior $\pi^J(\theta)$ is chosen here as the Jeffreys prior for a multivariate Gaussian distribution restricted to Ω_m , ie.

$$\pi^J(\theta) = \frac{\mathbf{I}_{\Omega_m}(m)}{\text{Vol}(\Omega_m)} \cdot \frac{\Delta_C}{|C|^{\frac{q+2}{2}}} \mathbf{I}_{\Omega_C}(C) \quad (16)$$

with

$$\Delta_C = \left(\int_{\Omega_C} \frac{1}{|C|^{\frac{q+2}{2}}} dC \right)^{-1}.$$

Thus

$$\begin{aligned} \Delta_C^{-1} &= \int_{\Omega_C} \frac{1}{|C|^{\frac{q+2}{2}}} dC \\ &= \int_{\Omega_C} \frac{1}{|D|^{\frac{q+2}{2}}} d(VDV^T) \\ &= \int dV \left[\int_{\Omega_D} \frac{1}{|D|^{\frac{q+2}{2}}} dD \right], \end{aligned}$$

where

$$\Omega_D = \left\{ D \in \mathcal{DS}_q^+ \text{ st. } |D| \geq |R|^{q/p}, 0 \leq D_{ii} \leq \sqrt{\sum_{j=1}^q \beta_j^2}, i = 1 \dots, q \right\}. \quad (17)$$

Now, any orthogonal matrix V of dimension q is characterised by the composition of $q(q-1)/2$ rotations $(\psi_1, \dots, \psi_{q(q-1)/2})$ (cf. Thisted 1988, [24]),

$$\int dV = \underbrace{\int_0^\pi \dots \int_0^\pi}_{q(q-1)/2 \text{ times}} d\psi_1 \dots d\psi_{q(q-1)/2} = \pi^{q(q-1)/2}.$$

Thus

$$\Delta_C^{-1} = \pi^{q(q-1)/2} \left[\int_{\Omega_D} \frac{1}{|D|^{\frac{q+2}{2}}} dD \right].$$

Finally, it remains to calculate the integral $\int_{\Omega_D} \frac{1}{|D|^{\frac{q+2}{2}}} dD$. Denoting it $I(q, a, \beta_1, \dots, \beta_q)$, with $a = |R|^{q/p}$ it is derived by induction on q (the detailed calculation is given in Appendix D).

$$I(q, a, \beta_1, \dots, \beta_q) = \left(\frac{q-1}{q} \right)^{q-1} I \left(q-1, \left(\frac{a}{\beta_q} \right)^{\frac{q}{q-1}}, \beta_1^{\frac{q}{q-1}}, \dots, \beta_{q-1}^{\frac{q}{q-1}} \right), \quad (18)$$

and

$$I(2, a, \beta_1, \beta_2) = \frac{1}{a} \log \frac{\beta_1 \beta_2}{a} + \frac{1}{\beta_1 \beta_2} - \frac{1}{a}.$$

D Computing the normalising constant of the diagonal variance matrix domain

We are aiming to calculate

$$I = \int_{\Omega_C} \frac{1}{|C|^{\frac{q+2}{2}}} dC, \quad (19)$$

when the variance matrix C is diagonal and the domain Ω_C is defined as follows:

$$\Omega_C = \left\{ C \in \mathcal{S}_q^+ \text{ st. } |C| \geq |R|^{q/p}, |C_{ij}| \leq \sqrt{\beta_i \beta_j}, i, j = 1 \dots, q \right\}. \quad (20)$$

Since C is diagonal, the above definition is equivalent to

$$\begin{cases} 0 \leq C_i \leq \beta_i \\ \prod_{i=1}^q C_i \geq a, \end{cases} \quad (21)$$

where $\{C_i, 1 \leq i \leq q\}$ are the diagonal elements of C . Conditions (21) are equivalent to the conditions

$$\begin{cases} \frac{a}{\beta_2 \dots \beta_q} \leq C_1 \leq \beta_1 \\ \frac{a}{C_1 \beta_3 \dots \beta_q} \leq C_2 \leq \beta_2, \\ \vdots \\ \frac{a}{C_1 C_2 \dots C_{q-1}} \leq C_q \leq \beta_q \end{cases} \quad (22)$$

Considering I as a function of $(q, a, \beta_1, \dots, \beta_q)$, the integral (19) can be developed as follows

$$\begin{aligned} I(q, a, \beta_1, \dots, \beta_q) &= \int_{\frac{a}{\beta_2 \dots \beta_q}}^{\beta_1} \frac{1}{C_1^{\frac{q+2}{2}}} dC_1 \int_{\frac{a}{C_1 \beta_3 \dots \beta_q}}^{\beta_2} \frac{1}{C_2^{\frac{q+2}{2}}} dC_2 \dots \int_{\frac{a}{C_1 \dots C_{q-1}}}^{\beta_q} \frac{1}{C_q^{\frac{q+2}{2}}} dC_q \\ &= \frac{2}{qa^{\frac{q}{2}}} \int_{\frac{a}{\beta_2 \dots \beta_q}}^{\beta_1} \frac{1}{C_1} dC_1 \int_{\frac{a}{C_1 \beta_3 \dots \beta_q}}^{\beta_2} \frac{1}{C_2} dC_2 \dots \int_{\frac{a}{C_1 \dots C_{q-2} \beta_q}}^{\beta_{q-1}} \frac{1}{C_{q-1}} dC_{q-1} \\ &\quad - \frac{2}{q\beta_q^{\frac{q}{2}}} \int_{\frac{a}{\beta_2 \dots \beta_q}}^{\beta_1} \frac{1}{C_1^{\frac{q+2}{2}}} dC_1 \int_{\frac{a}{C_1 \beta_3 \dots \beta_q}}^{\beta_2} \frac{1}{C_2^{\frac{q+2}{2}}} dC_2 \dots \int_{\frac{a}{C_1 \dots C_{q-2} \beta_q}}^{\beta_{q-1}} \frac{1}{C_{q-1}^{\frac{q+2}{2}}} dC_{q-1} \\ &= \frac{2}{qa^{\frac{q}{2}}} I_{q-1} - \frac{2}{q\beta_q^{\frac{q}{2}}} \left(\frac{q-1}{q} \right)^{q-1} I \left(q-1, \left(\frac{a}{\beta_q} \right)^{\frac{q}{q-1}}, \beta_1^{\frac{q}{q-1}}, \dots, \beta_{q-1}^{\frac{q}{q-1}} \right), \end{aligned} \quad (23)$$

where

$$\begin{aligned} I_{q-1} &= \int_{\frac{a}{\beta_2 \dots \beta_q}}^{\beta_1} \frac{1}{C_1} dC_1 \int_{\frac{a}{C_1 \beta_3 \dots \beta_q}}^{\beta_2} \frac{1}{C_2} dC_2 \dots \int_{\frac{a}{C_1 \dots C_{q-2} \beta_q}}^{\beta_{q-1}} \frac{1}{C_{q-1}} dC_{q-1} \\ &= \frac{1}{(q-1)!} \left(\log \frac{\beta_1 \dots \beta_q}{a} \right)^{q-1}, \end{aligned} \quad (24)$$

is obtained by induction and

$$\begin{aligned} &\int_{\frac{a}{\beta_2 \dots \beta_q}}^{\beta_1} \frac{1}{C_1^{\frac{q+2}{2}}} dC_1 \int_{\frac{a}{C_1 \beta_3 \dots \beta_q}}^{\beta_2} \frac{1}{C_2^{\frac{q+2}{2}}} dC_2 \dots \int_{\frac{a}{C_1 \dots C_{q-2} \beta_q}}^{\beta_{q-1}} \frac{1}{C_{q-1}^{\frac{q+2}{2}}} dC_{q-1} \\ &= \left(\frac{q-1}{q} \right)^{q-1} I \left(q-1, \left(\frac{a}{\beta_q} \right)^{\frac{q}{q-1}}, \beta_1^{\frac{q}{q-1}}, \dots, \beta_{q-1}^{\frac{q}{q-1}} \right), \end{aligned}$$

by the variable change

$$y_i = C_i^{\frac{a}{q-1}}.$$

Thus step by step thanks to equation (23), the integral can be calculated when C is diagonal. For instance for $q = 2, 3, 4$ we get

$$\begin{aligned} I(2, a, \beta_1, \beta_2) &= \frac{1}{a} \log \frac{\beta_1 \beta_2}{a} + \frac{1}{\beta_1 \beta_2} - \frac{1}{a}, \\ I(3, a, \beta_1, \beta_2, \beta_3) &= \frac{1}{3a^{\frac{3}{2}}} \left(\log \frac{\beta_1 \beta_2 \beta_3}{a} \right)^2 - \frac{4}{9a^{\frac{3}{2}}} \log \frac{\beta_1 \beta_2 \beta_3}{a} - \frac{8}{27(\beta_1 \beta_2 \beta_3)^{\frac{3}{2}}} + \frac{8}{27a^{\frac{3}{2}}}, \\ I(4, a, \beta_1, \beta_2, \beta_3, \beta_4) &= \frac{1}{12a^2} \left(\log \frac{\beta_1 \beta_2 \beta_3 \beta_4}{a} \right)^3 - \frac{1}{8a^2} \left(\log \frac{\beta_1 \beta_2 \beta_3 \beta_4}{a} \right)^2 + \frac{1}{8a^2} \left(\log \frac{\beta_1 \beta_2 \beta_3 \beta_4}{a} \right) \\ &+ \frac{1}{16(\beta_1 \beta_2 \beta_3 \beta_4)^2} - \frac{1}{16a^2}. \end{aligned}$$

References

- [1] Brooks, S.P. and Gelman, A. (1998), General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, **7**, 434-455.
- [2] Bousquet, N. (2008), Diagnostics of prior-data agreement in applied Bayesian analysis, *J. Appl. Statist.*, **35**, 1011-1029.
- [3] Horn, R. A. and Johnson, C. R. (1991), *Topics in Matrix Analysis*, Cambridge University Press.
- [4] Bastos, L.S. and O'Hagan, A. (2009), Diagnostics for Gaussian Process Emulators. Technical report, University of Sheffield.
- [5] Johnson, M. E., Moore, L. M. et Ylvisaker, D. (1990), Minimax and maximin distance designs, *Journal of Statistical Planning and Inference*, **26**, 131-148.
- [6] Barbillon, P. (2010), Méthodes d'interpolation à noyaux pour l'approximation de fonctions type boîte noire coûteuses. Université Paris-Sud.
- [7] Matheron, G. (1971). *The theory of regionalised variables and its applications*. École Nationale Supérieure des Mines de Paris.
- [8] Mitchell, T., Morris, M. and Ylvisaker, D. (1990), Existence of smoothed stationary processes on an interval, *Stochastic Processes and Their Applications*, **35**, 109-119,
- [9] Joseph V.R., and Hung Y. (2008), Orthogonal-Maximin Latin Hypercube Designs, *Statistica Sinica*, **18**, 171-186.
- [10] Petelet M., Iooss. B., Asserin O., Marrel, A.(2010), Latin hypercube sampling with inequality constraints, *Advances in Statistical Analysis*, **94**, 325-339
- [11] Celeux, G., Grimaud, A., Lefebvre, Y., and De Rocquigny, E. (2010), Identifying intrinsic variability in multivariate systems through linearised inverse methods, *Inverse Problems in Engineering*, **18**, 401-415.
- [12] Barbillon, P., Celeux, G., Grimaud, A., Lefebvre, Y. and Rocquigny (De), E. (2011), Non linear methods for inverse statistical problems, *Computational Statistics & Data Analysis*, **55**, 132-142.
- [13] McKay, M.D., Beckman, R.J., and Conover, W.J. (1979), A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, **21**, 239-245.
- [14] Fang, K.-T., Li, R. and Sudjianto, A. (2006), *Design and Modeling for Computer Experiments*, Chapman & Hall/CRC.
- [15] Li, R. and Sudjianto, A. (2005), Analysis of computer experiments using penalized likelihood in Gaussian kriging models, *Technometrics*, **47**, 111-120.
- [16] Koehler, J.R. and Owen, A.B. (1996), *Computer experiments*, in: Ghosh, S., Rao, C.R. (Eds.), *Handbook of Statistics*. Elsevier, 261-308.
- [17] Munoz-Zuniga, M., Garnier, J., Remy, E. and de Rocquigny, E. (2011). Analysis of adaptive directional stratification for the controlled estimation of rare event probabilities. *Statistics & Computing*, **22**, 809-821.

-
- [18] Rasmussen, C. E. and Williams, C. (2006) *Gaussian Processes for Machine Learning*, the MIT Press, 2006.
- [19] Sacks, J., Schiller, S.B., Mitchell, T.J. and Wynn, H.P. (1989a), Design and analysis of computer experiments (with discussion), *Statistica Sinica*, **4**, 409-435.
- [20] Sacks, J., Schiller, S.B. and Welch, W.J. (1989b), Designs for computer experiments, *Technometrics*, **31**, 41-47.
- [21] Santner, T.J., Williams, B. and Notz, W. (2003), *The Design and Analysis of Computer Experiments*, Springer-Verlag.
- [22] Yang, R. and Berger, J.O. (1998), A Catalog of Non-informative Priors, *ISDS Discussion Paper* 97-42.
- [23] Tierney, L. (1995), Introduction to general state-space Markov chain theory, in *Markov Chain Monte Carlo in Practice*, 59-74. Chapman & Hall.
- [24] Thisted, R.A. (1988), *Elements of Statistical Computing*, Chapman & Hall.
- [25] Vanderpoorten, A. and Palm, R. (2001), Compared regression methods for inferring ammonium nitrogen concentrations in rivers from aquatic bryophyte assemblages. *Hydrobiologia*, **452**, 181-190.

Contents

1	Introduction	3
2	Bayesian inference with a Gaussian emulator	4
3	Assessing a prior distribution and a design	8
3.1	The DAC criterion	8
3.2	The impact of the emulator	8
3.3	Computing DAC	9
3.4	Using the $\widehat{\text{DAC}}$ criterion	10
4	Numerical experiments	11
4.1	Assessing the design	12
4.2	Assessing the prior and the design	13
5	Discussion	17
A	the Gibbs sampling with a Jeffreys non-informative prior	20
B	Brooks-Gelman Statistics	21
C	Computing DAC for the kriging emulator	21
D	Computing the normalising constant of the diagonal variance matrix domain	23



**RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE**

Parc Orsay Université
4 rue Jacques Monod
91893 Orsay Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399