International Journal for Uncertainty Quantification, 5 (1): 73–98 (2015)

BAYESIAN INFERENCE FOR INVERSE PROBLEMS OCCURRING IN UNCERTAINTY ANALYSIS

Shuai Fu,[1,3,*] Gilles Celeux,[2] Nicolas Bousquet,[3] & Mathieu Couplet[3]

[1] *Université Paris-Sud 11, France*

[2] *Inria Saclay-Île-de-France*

[3] *EDF, R&D, France*

Original Manuscript Submitted: 03/17/2014; Final Draft Received: 10/07/2014

The inverse problem considered here is the estimation of the distribution of a nonobserved random variable X, linked through a time-consuming physical model H to some noisy observed data Y. Bayesian inference is considered to account for prior expert knowledge on X in a small sample size setting. A Metropolis-Hastings-within-Gibbs algorithm is used to compute the posterior distribution of the parameters of the distribution of X through a data augmentation process. Since running H is quite expensive, this inference is achieved by a kriging emulator interpolating H from a numerical design of experiments (DOE). This approach involves several errors of different natures and, in this article, 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 DOE and the prior distribution. After describing the calculation of 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.*

1. INTRODUCTION

The probabilistic treatment of uncertainties is gaining fast growing interest in numerous industrial fields. Designing and predicting the behavior of complex mechanisms in various environments benefits from increasing computational means, and numerical simulation has become a well established domain of engineering. In such settings, a physical phenomenon of interest is implemented through a so-called computer code, often seen as a black-box, that involves a set of input variables to be calibrated. Such variables can be considered uncertain, because they reflect intrinsic randomness or because any information to calibrate them is noisy or indirect. Therefore, 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 major difficulty is linked to the limited sampling information directly available on uncertain input variables. A simple example, that motivated the present study, is the prediction of a river water level using hydraulical codes: one of their most influent inputs is the river bed friction, which is uncertain by nature and for which no observation is directly available. This parameter summarizes a set of local geomorphological effects, the consequence of which is discernible only by observing the fluctuations of the output variable, namely the water level.

For such variables, it is highly beneficial (a) to integrate expert judgment into the quantification, such as likely bounds on physical intervals or more elaborate probabilistic information, provided this judgment is reliable, and (b) to integrate indirect information connected to the uncertain variable of interest through a physical model. The recovery of indirect information generally involves the probabilistic inversion of a computer simulator H implementing this model. The situation of indirect information can be summarized by the following equation:

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

*Correspond to Shuai Fu, E-mail: fshvip@gmail.com

2152–5080/15/$35.00

where $Y_i \in \mathbb{R}$ is an observable output, $X_i \in \mathbb{R}^q$ is a nonobserved input, $d_i \in \mathbb{R}^{q_2}$ is an observed input related to the experimental conditions, and $U_i \in \mathbb{R}^p$ is a measurement error. The purpose is to estimate the distribution $\mathcal{F}$ of the random vectors X_i from the observations $(y_i, i = 1, \ldots, n)$, knowing that the function H cannot be inverted formally or numerically in due time.

Many approaches are possible to approximate the solution of this inverse problem (provided it exists), as linearizing the physical model H around a fixed point x_0 (see Celeux et al. [1]), or using a non linear approximation of the function H obtained through kriging and making use of a stochastic procedure with this nonlinear approximation of H (see Barbillon et al. [2] and Li and Sudjianto [3]). The innovation proposed here is to consider a Bayesian approach of the problem. Not only it allows to address both issues (a) and (b) simultaneously, but it can be helpful, especially, to avoid identifiability problems not investigated by the former approaches. In this framework, $\mathcal{F}$ is given a prior distribution, and the inversion problem becomes conditioning the prior knowledge to the available indirect observations, which results in the posterior distribution of $\mathcal{F}$.

Nonetheless, the technical choices implied by solving this inverse problem in a Bayesian statistical setting are likely to produce several kinds of error, the combination of which potentially threatens the relevance of the results. They are listed beneath:

- *Estimation error:* Usually the sample size n is small with respect to the dimension of the problem, and the variance of the estimators of the parameters defining $\mathcal{F}$ can be expected to be large, even if some additive knowledge may be incorporated through a choice of an informative prior distribution;
- *Emulator error:* Since H is too complex, it is needed to replace it with an emulator $\widehat{H}$, pursuing the ideas of Barbillon et al. [2], and the discrepancy between H and $\widehat{H}$ could induce an important error;
- *Algorithmic error:* To proceed to statistical inference, it is needed to use complex stochastic algorithms. In the Bayesian setting, those algorithms are Monte Carlo Markov Chain (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 essential to get reliable estimates.
- *Prior uncertainty:* The prior knowledge on the parameters defining $\mathcal{F}$ is expected to produce regularized estimates with smaller variances than maximum likelihood estimates. But eliciting reliable information from experts could be difficult and wrong prior information can severely influence the inference.

Beyond the estimation problem, this article 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, the data agreement 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 article is organized as follows. In Section 2, the MCMC algorithm for a Bayesian estimation of an emulator of the 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 the relevance of 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

The statistical problem is to estimate the probability distribution $\mathcal{F}$ of $\mathbf{X} = \{X_1, \ldots, X_n\}$ from the observations $\mathbf{y} = \{y_i, i = 1, \ldots, n\}$, where y_i is the realisation of the random variable Y_i. To get a manageable but rather generic problem, it is assumed that the unknown input X_i follows a multivariate Gaussian distribution $\mathcal{N}_q(m, C)$ with unknown mean m and variance matrix C. Other assumptions are the mutual independence of the X_i and the error U_i for $i = 1, \ldots, n$, and the independence of the Y_i. Besides, the observation error U_i is supposed to follow a Gaussian distribution $\mathcal{N}_p(0, R)$ with known diagonal variance matrix R. Finally, by considering only an independent measurement error, it is underlyingly assumed that the model discrepancy, namely the difference between the actual

physical process from which the observations Y_i are taken and the physical model H, is negligible. Details about the concrete handling of model discrepancy can be found, for instance, in Kennedy and O'Hagan [4] and Brynjarsdóttyr and O'Hagan [5].

In the Bayesian framework, the first task is to choose a prior distribution $\pi(\theta)$ for the parameter vector $\theta = (m, C)$ to be estimated in the model (1). A conjugate prior distribution appears as a rather natural and usual choice:

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

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

where $\mathcal{IW}_q(\Lambda, \nu)$ denotes an inverse-Wishart distribution, with $\nu > q-1$ the degrees of freedom and $\Lambda \in \mathcal{M}^{q\times q}$ the positive definite inverse scale matrix; the hyperparameters $\rho = (\mu, a, \Lambda, \nu)$ are assumed to be specified by the practitioner, typically from expert knowledge. See O'Hagan et al. [6] for a review of the dedicated methods.

The conjugation properties being, however, restricted to cases where the X_i are not missing, the multivariate posterior distribution $\pi(\theta|\mathbf{y})$ is not explicit. Therefore it must be described by simulation means. This can be approximated using Markov chains produced by a Gibbs sampler including a Metropolis-Hastings (MH) step (see, for instance, Tierney [7]). Actually, the calculation of the full conditional posterior distributions of m, the variance matrix C, and $\mathbf{X}$ leads to the following Gibbs sampler (below the $(r+1)$-th iteration):

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

---

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

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

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

   which is not belonging to a closed form family of distributions. Thus a 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 time-consuming, the Gibbs sampler must be adapted. In those situations, we propose to replace H with a *maximin*-LHD (latin hypercube design) kriging emulator $\widehat{H}$, following Barbillon [8]. This emulator is briefly described below.

- *Kriging* is a geostatistical method (Matheron [9]) that has been adapted by Sacks et al. [10, 11] 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 [12], Santner et al. [13], and Fang et al. [14], among others. According to this approach the function H is regarded as the realization of a Gaussian process (GP) $\mathcal{H} \sim \mathrm{GP}(\mu, c)$, characterized by its mean and variance functions: $\mu(z) = \mathbb{E}[\mathcal{H}(z)]$ and $c(z, z') = \mathrm{Cov}[\mathcal{H}(z), \mathcal{H}(z')] = \sigma^2 K_\epsilon(\|z - z'\|)$ for any $z = (x, d)$, K_ϵ being a symmetric positive kernel such that $K_\epsilon(0) = 1$ (see Mitchell et al. [15]). In a Bayesian perspective, GP modeling can be interpreted as providing H with a prior (Rasmussen and Williams [16]). The process $\mathcal{H}$ can be proved to be normally distributed knowing some evaluations $\mathbf{H}_{D_N} = \{H(z_{(1)}), \ldots, H(z_{(N)})\}$ on a *design of experiments* (DOE) $D_N = \{z_{(1)}, \ldots, z_{(N)}\}$ of N points $z_{(j)} = (x_{(j)}, d_{(j)})$.

The best MSPE (*mean squared prediction error*) predictor of H, denoted by $\widehat{H}$, is the conditional mean:

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

Then $\widehat{H}(z)$ is minimizing the conditional expectation of the loss function $(\mathcal{H}(z) - \widehat{H}(z))^2$, the so-called MSE (mean squared error) (see Johnson et al. for details [17]),

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

- The set $D_N = \{z_{(1)}, \ldots, z_{(N)}\}$ is chosen on $\Omega \in \mathbb{R}^{q+q_2}$ according to a *maximin*-LHD (see Joseph and Hung [18], McKay et al. [19], and Petelet et al. [20]): 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 maximizes

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

among the LHD of size N.

Finally, considering the new *emulator error* resulting from this version of kriging, the conditional distribution of $\mathbf{X}$ is described as follows. Note that simulating this conditional distribution $\mathbf{X}$ requires again a Metropolis-Hastings step inside the Gibbs sampler, the details of which being provided in Fu et al. [21].

$$\mathbf{X}^{[r+1]}|\cdots \propto |\mathbf{R} + \text{MSE}^{[r+1]}|^{-1/2} \cdot \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n}(X_i^{[r+1]} - m^{[r+1]})'\left[C^{[r+1]}\right]^{-1}(X_i^{[r+1]} - m^{[r+1]}) \right.$$

$$\left. -\frac{1}{2}\left(\left(y_1 - \widehat{H}_1^{[r+1]}\right)', \ldots, \left(y_n - \widehat{H}_n^{[r+1]}\right)'\right)\left(\mathbf{R} + \text{MSE}^{[r+1]}\right)^{-1}\begin{pmatrix} y_1 - \widehat{H}_1^{[r+1]} \\ \vdots \\ y_n - \widehat{H}_n^{[r+1]} \end{pmatrix} \right\},$$

where $\widehat{H}_i^{[r+1]} = \widehat{H}(X_i^{[r+1]}, d_i)$, by denoting the ith diagonal component of R by R_{ii},

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

and $\text{MSE}^{[r+1]} = \text{MSE}\left(\mathbf{X}^{[r+1]}, \mathbf{d} = (d_1, \ldots, d_n)\right)$ is a block diagonal matrix defined by

$$\text{MSE}(\mathbf{X}^{[r+1]}, \mathbf{d}) = \begin{pmatrix} \text{MSE}_1(\mathbf{X}^{[r+1]}, \mathbf{d}) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \text{MSE}_p(\mathbf{X}^{[r+1]}, \mathbf{d}) \end{pmatrix}, \begin{matrix} \} \ n \text{ lines} \\ \\ \} \ n \text{ lines} \end{matrix}$$

with each variance matrix $\text{MSE}_j(\mathbf{X}^{[r+1]}, \mathbf{d}) \in \mathcal{M}^{n \times n}$ defined by

$$\text{MSE}_j(\mathbf{X}^{[r+1]}, \mathbf{d}) = \mathbb{E}\left(\left(\mathcal{H}_j(\mathbf{X}^{[r+1]}, \mathbf{d}) - \widehat{H}_j(\mathbf{X}^{[r+1]}, \mathbf{d})\right)^2 \,|\, \mathbf{H}_{D_N}\right).$$

Here $\mathcal{H}_j$ denotes the jth dimension of the Gaussian process $\mathcal{H}$. In this conditional posterior distribution, the error term is composed of $\mathbf{R}$ and MSE, where the former represents the uncertainty from the physical model and the latter represents the uncertainty from the Gaussian emulator.

2.1 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 effort having been made on this question, there is not an absolute way to answer it. We chose to use the much employed Brooks-Gelman (BG) statistics (Brooks and Gelman [22]) computed from five replications of the Monte Carlo Markov chain (see Appendix A). The MCMC algorithm is stopped if the BG statistics remains smaller than 1.05. This threshold is more stringent than the standard threshold 1.2 suggested in [22].

2.2 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*). This could occur because kriging makes use of a Gaussian process approximation that is known to be smooth, while some types of physical model H are not and will require using larger number of points N to approximate them. Thus, a too small number of points N chosen for the design D_N can distinctly increase the emulator error. The two following criteria, among the most used criteria to measure the quality of a design, are investigated here.

(i) The coefficient of predictability Q_2 (see Vanderpoorten and Palm [23]) is

$$Q_2 = 1 - \frac{\text{PRESS}(D^*)}{\left\|H(D^*) - \overline{H}(D^*)\right\|^2}, \tag{4}$$

where

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

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

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

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

$$Q_{2\,\text{CV}} = 1 - \frac{\text{PRESS}_{\text{CV}}}{\sum_{i=1}^{N} \left\|H(z_{(i)}) - \overline{H}_{D_N}\right\|^2}, \tag{5}$$

with

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

and

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

where

- $e_{(i)}$ is the prediction error at $z_{(i)}$ of a fitted model without the point $z_{(i)}$;
- $\widehat{H}_{-i}(z_{(i)})$ 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 and the better the quality of the design D_N.

(ii) An alternative criterion is the Mahalanobis distance (MD) (see Bastos and O'Hagan [24]), computed on a validation sample D^* with N^* points as follows:

$$\text{MD} = \Big(H(D^*) - \widehat{H}(D^*)\Big)' \Big(\text{MSE}(D^*)\Big)^{-1} \Big(H(D^*) - \widehat{H}(D^*)\Big), \tag{6}$$

where $\text{MSE}(D^*)$ is the conditional variance matrix of the design D^* knowing $H_{D^*} = \{H(v_{(1)}), \ldots, H(v_{(N^*)})\}$. An interest of this criterion is to account for 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:

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

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

Now, the smaller the sample size n, the greater the *estimation error*. The two above-mentioned criteria are not aiming to measure this estimation error. But since H is complex, it is quite difficult to assess this error in an inverse modeling 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 may 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 [25]) for this task, as detailed in the next section.

3. ASSESSING A PRIOR DISTRIBUTION AND A DESIGN

3.1 The DAC Criterion

The data agreement criterion (DAC) (Bousquet [25]) has been conceived as a measure of the discrepancy between a prior distribution of model parameters and the data. Let $\mathbf{y}$ be a sample with pdf $f(\mathbf{y}|\theta)$. Let $\pi^J(\theta)$ be a benchmark noninformative prior (see for instance, Yang and Berger [26]) and $\pi(\theta)$ the prior distribution derived from the prior information on θ. DAC is defined as

$$\text{DAC}(\pi|\mathbf{y}) = \frac{\text{KL}(\pi^J(\theta|\mathbf{y})||\pi(\theta))}{\text{KL}(\pi^J(\theta|\mathbf{y})||\pi^J(\theta))}, \tag{7}$$

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

$$\text{KL}(p||q) = \int_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)}\, dx, \tag{8}$$

$\mathcal{X}$ being the set of all accessible values for x. The rationale underlying the definition of DAC is as follows: the posterior distribution $\pi^J(\theta|\mathbf{y})$ derived from the noninformative prior provides essentially the same information on theta as the data $\mathbf{y}$. This posterior can thus be interpreted as a benchmark prior perfectly in accordance with data information. The

divergence $\mathrm{KL}(\pi^J(\theta|\mathbf{y})||\pi(\theta))$ (or relative negative entropy) between the true prior π and $\pi^J(\cdot|\mathbf{y})$ provides a measure of the discrepancy between the two sources of information on θ.

If $\mathrm{DAC}(\pi|\mathbf{y}) \leq 1$, the informative prior π is closer to $\pi^J(\cdot|\mathbf{y})$ than the noninformative prior π^J, and the data $\mathbf{y}$ and the prior $\pi(\theta)$ are declared to be in agreement. Otherwise if $\mathrm{DAC}(\pi|\mathbf{y}) > 1$, the data $\mathbf{y}$ and the prior $\pi(\theta)$ are declared to be discrepant. DAC has been proved to be efficient when the noninformative prior $\pi^J(\theta)$ is proper, otherwise it must be adapted using techniques used for estimating Bayes factors (see Bousquet [25]).

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 to define a proper noninformative prior $\pi^J(m, C)$, which is 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 B and C.

It is important to notice that the 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,

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

A DAC value greater than 1 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 $(S/T) \leq 1 \Longleftrightarrow S - T \leq 0$, if $S \geq 0, T > 0$, a numerically more convenient version of DAC, denoted $\widetilde{\mathrm{DAC}}$, is

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

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

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

and

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

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

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

and checking $\widetilde{\mathrm{DAC}}(\pi|\mathbf{y}, \mathbf{H}_{D_N}) \leq 0$ means that the prior distribution $\pi(\theta)$ and the couple $(\mathbf{y}, \mathbf{H}_{D_N})$ are compatible. This criterion can be computed using the outputs of a Gibbs sampler run with a noninformative prior $\pi^J(\cdot)$:

$$\widetilde{\mathrm{DAC}}(\pi|\mathbf{y}, \mathbf{H}_{D_N}) \backsimeq \frac{1}{R}\sum_{r=1}^{R} \log \pi^J(\theta^r) - \frac{1}{R}\sum_{r=1}^{R} \log \pi(\theta^r), \tag{9}$$

where $\theta^r \sim \pi^J(\cdot|\mathbf{y}, \mathbf{H}_{D_N})$, $r \in \{1, ..., R\}$ is a simulated sequence obtained by Gibbs sampling. Denoting $\overline{X}_n = (1/n)\sum_{i=1}^n X_i$, the full conditional distribution of m verifies

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

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

$$\pi^J(C \,|\, m, \mathbf{Y}, \mathbf{X}, \rho, H_D) \propto \mathbf{I}_{\Omega_C} |C|^{-[(n+q+2)/2]} \exp\left[-\frac{1}{2}\mathrm{Tr}\left(n\,(m - \overline{X}_n)(m - \overline{X}_n)' \cdot C^{-1}\right)\right]. \tag{10}$$

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

$$\mathbf{I}_{\Omega_C} \cdot \mathcal{IW}\left(n\,(m - \overline{X}_n)(m - \overline{X}_n)', n + 1\right). \tag{11}$$

Using the full conditional posterior distributions of $\mathbf{X}$, the Gibbs sampler approximating the posterior distribution of (m, C) with a noninformative prior truncated to the domain $\Omega_m \times \Omega_C$ could be straightforwardly described (see Fu et al. [21]).

Remark: The simulation of C is difficult since $n\,(m - \overline{X}_n)(m - \overline{X}_n)'$ is not a definite but a semi-definite positive matrix and numerical problems can occur. However, up to an additive constant, the calculation (10) is proper. For this reason, we recommend to use a Metropolis-Hastings algorithm for simulating C:

Metropolis-Hastings (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 $\epsilon\mathbf{I}_q$ to the semi-positive definite matrix $(m - \overline{X}_n)(m - \overline{X}_n)'$, with ϵ a small positive value and $\mathbf{I}_q$ an identity matrix of dimension $q \times q$

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

   - Let

     $$\alpha(C^{[h-1]}, \xi) = \frac{g(\xi) f^*(C^{[h-1]})}{g(C^{[h-1]}) f^*(\xi)} \wedge 1, \tag{12}$$

     with g proportional to the target distribution which means the truncated Inverse-Wishart distribution (11)

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

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

     $$C^{[h]} = \begin{cases} \xi & \text{with probability } \alpha(C^{[h-1]}, \xi), \\ C^{[h-1]} & \text{otherwise.} \end{cases} \tag{13}$$

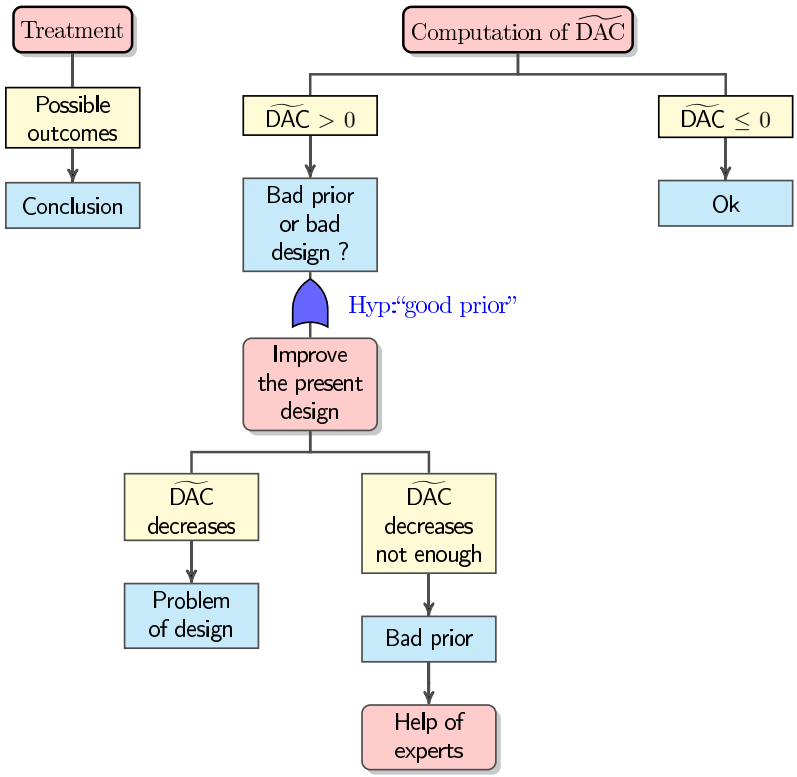
In this way, the produced Markov chain $(C^{[h]})$ converges to the target distribution (11).

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 relevance with respect to the observed data y. But this double assessment has to be done properly using the following procedure:

1. If $\widetilde{\text{DAC}} \leq 0$ then the prior and the design are declared to be acceptable.
2. If $\widetilde{\text{DAC}} > 0$, the following step is required:
   under a "good prior" assumption, efforts are made to improve the design by increasing N or modifying Ω. If $\widetilde{\text{DAC}}$ is not decreasing under zero, it means that the prior information is questionable and there is the need to go back to the experts or, maybe more honestly, to employ vague priors as the Jeffreys priors.

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 two statistical models. The first example is a simplified version of a hydraulic model, and the second example is a a real physical hydraulic code which is widely used at EDF (Electricité de France).

4.1 Simplified Hydraulic Model

We consider the following simplified two-dimensional hydraulic model used, for instance, in [27]:

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

where,

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 30 \\ 50 \end{pmatrix}, \begin{pmatrix} 5^2 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

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

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

Since we are mainly concerned in analyzing the behavior of $\widetilde{\text{DAC}}$, six different prior distributions on the model parameters are considered, which present several degrees of accordance with the simulated data X. They are summarized in Table 1. Note that the prior distributions or the parameters m and C are $m|C \sim \mathcal{N}(\mu, C/a)$ and $C \sim \mathcal{IW}(\Lambda, \nu)$ with $\Lambda = t \cdot \widetilde{C}_{\text{Exp}}$.

4.1.1 Checking for Good Posterior Approximation

A first task is to validate the approximation of the true posterior distribution provided by the Gibbs algorithm calling the kriging emulator. The cases when the prior is noninformative (needed later to compute DAC) or in accordance with the simulated data (FHV) should therefore lead to get posterior estimation results about (m, C, X, Y) close to the values used in the simulation task. Using the Brooks-Gelman heuristics detailed in Appendix A to test the convergence of MCMC chains, 1000 posterior samples (sorted using a classical ACF autocorrelation test) were produced, each one conditioned to a design of 20 points. The posterior means, standard deviations, and 95% credible intervals were calculated for the parameters and displayed in Table 2. The presence of the simulation values

TABLE 1: Description of the six prior distributions by their accordance with the simulated data: 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

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

TABLE 2: Posterior means, standard deviations, and 95% credible intervals for (m, C) given, respectively, the FHV informative prior and noninformative prior

Parameters	m_1	m_2	C_{11}	C_{22}
Informative prior (FHV)				
Estimate	30.9014	49.4802	26.4058	1.0778
Std. dev.	0.9895	0.1941	6.9559	0.2925
2.5%	29.0053	49.1112	15.5504	0.6528
97.5%	32.9227	49.8723	43.2018	1.7495
Non-informative prior				
Estimate	30.8865	49.5102	25.6555	1.1032
Std. dev.	0.9941	0.1965	7.5674	0.3210
2.5%	28.9870	49.1423	14.9727	0.6323
97.5%	32.9860	49.8951	44.3246	1.9190

within the posterior coverage illustrates the relevance of the approximation of H and the management of MCMC chains.

Furthermore, because an uncertainty study requires to sample relevant values for X, the posterior predictive missing data X were sampled from a multivariate normal distribution calibrated with these posterior values. Finally the posterior predictive observations Y were computed using these reconstructed X, the observed values d, and the true function H. Figures 1–4 superimpose the isolines of the posterior predictive density of X and Y and the true observations (marked as red stars). Again, a good coverage can be noticed. Note that using the informative prior very slightly improves the coverage with respect to the noninformative one.

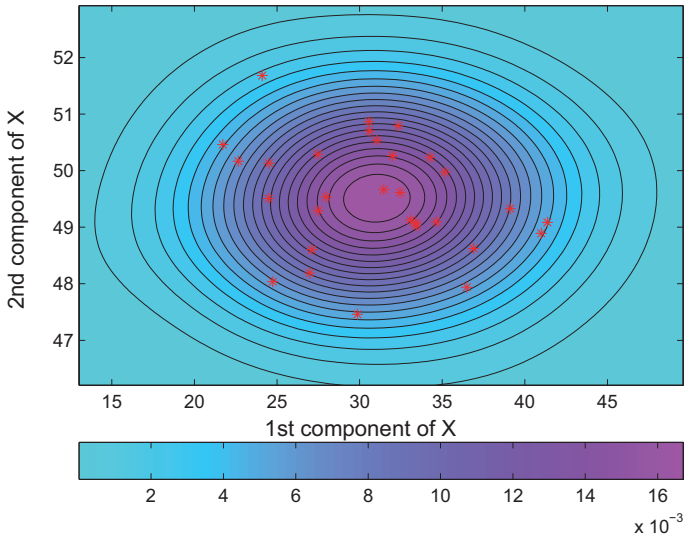


FIG. 1: Isolines of the posterior predictive density and true values (red stars) of the missing data X, given the FHV informative prior.

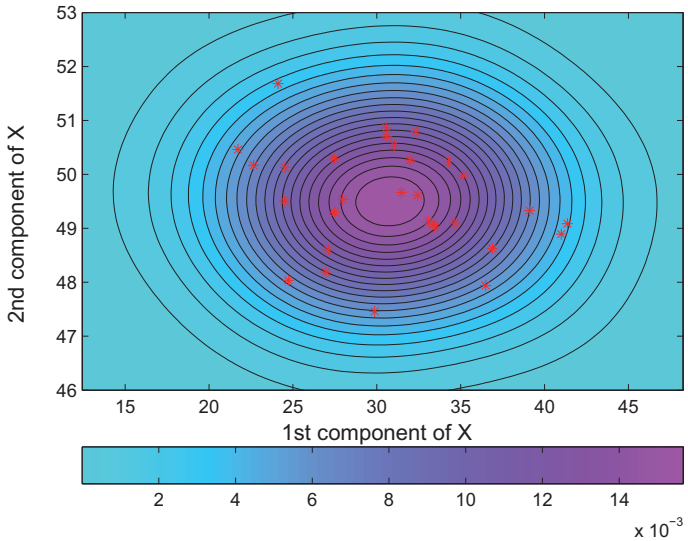


FIG. 2: Isolines of the posterior predictive density and true values (red stars) of the missing data X, given the noninformative prior.

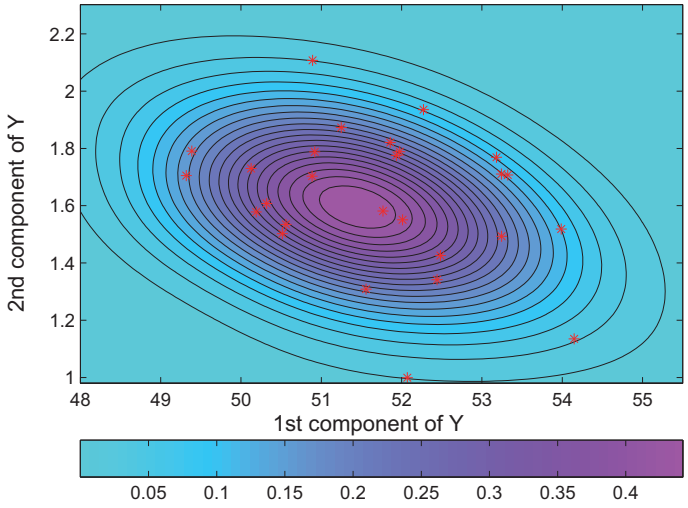


FIG. 3: Isolines of the posterior predictive density of Y and true observations $\mathbf{y}$ (red stars), given the FHV informative prior.

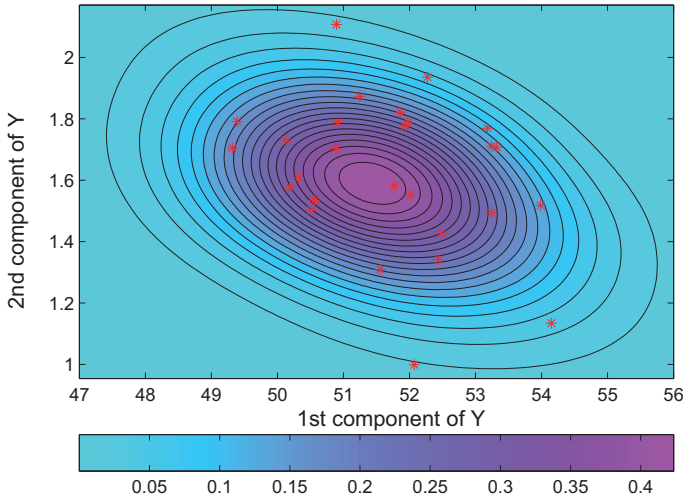


FIG. 4: Isolines of the posterior predictive density of Y and true observations $\mathbf{y}$ (red stars), given the noninformative prior.

4.1.2 Assessing the Relevance of the Design

The following experiments aim now at assessing 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 of 100 points within Ω_1. Figures 5 and 6 give the box plots of $1 - Q_2$ based on 20 repetitions computed on a validation sample and by cross-validation, respectively. The closer 1 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

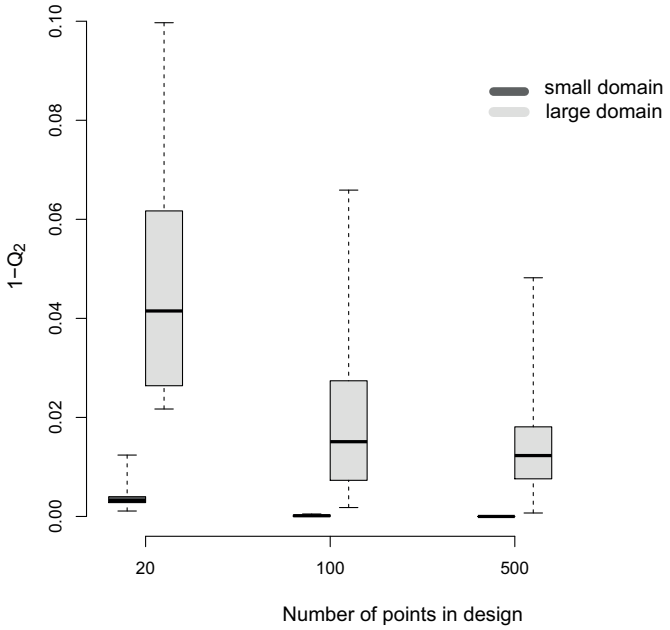


FIG. 5: $1 - Q_2$ boxplots based on 20 repetitions, calculated on a validation sample for six *maximin*-LHDs of 20, 100, and 500 points within small domain Ω_1 (—) and large domain Ω_2 (—), with the validation sample as a *maximin*-LHD of 100 points within Ω_1.

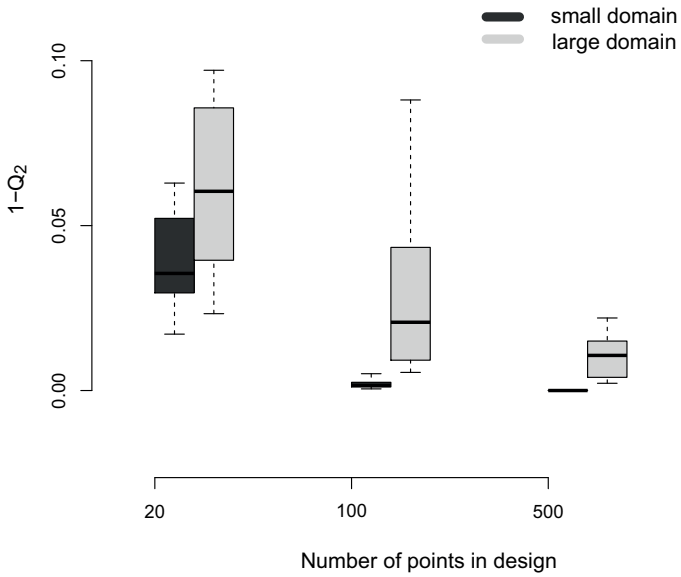


FIG. 6: $1 - Q_2$ boxplots based on 20 repetitions, calculated by cross-validation for six *maximin*-LHDs of 20, 100, and 500 points within small domain Ω_1 (—) and large domain Ω_2 (—).

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 7 and 8 display the boxplots of $\log(\text{MD})$ in the same conditions. As it could be expected, this criterion is decreasing when the number of design points increases. But, these figures show unexpected differences: the

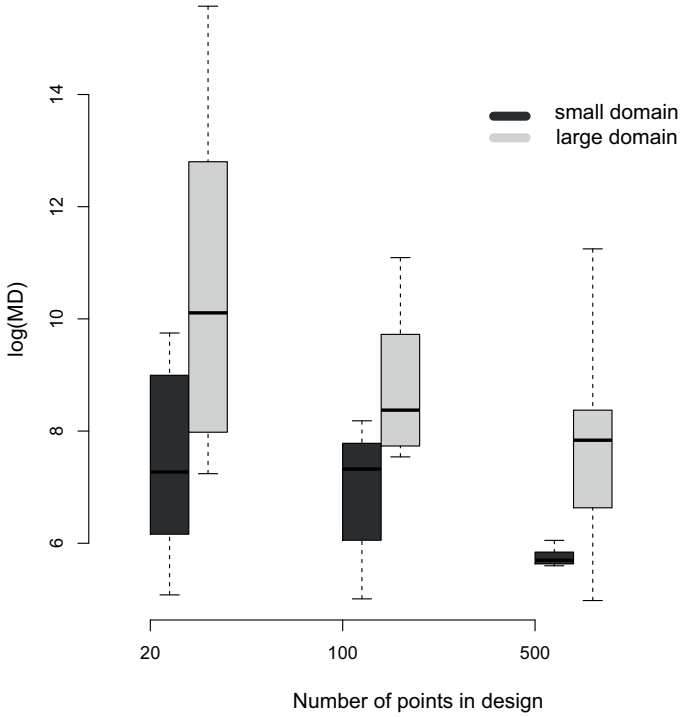


FIG. 7: MD boxplots based on 20 repetitions, calculated on a validation sample for six *maximin*-LHDs of 20, 100, and 500 points within small domain Ω_1 (—) and large domain Ω_2 (—), with the validation sample as a *maximin*-LHD of 100 points within Ω_1.

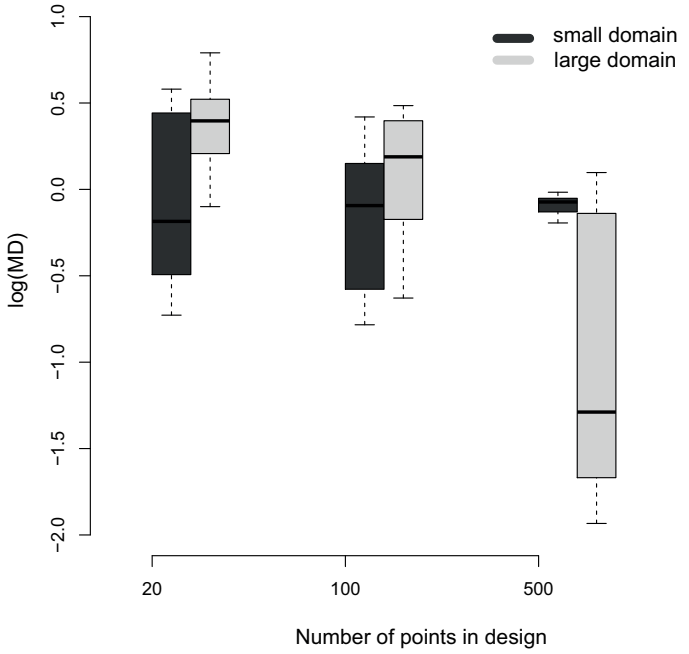


FIG. 8: MD boxplots based on 20 repetitions, calculated by cross-validation for six *maximin*-LHDs of 20, 100, and 500 points within small domain Ω_1 (—) and large domain Ω_2 (—).

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 amazingly scattered (see Fig. 8). Moreover, contrary to the 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 Fig. 7).

4.1.3 Assessing the Relevance of the Prior and the Design

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

Figure 9 depicts the behavior of DAC in the small domain, for 100 repeated estimations of the model with the six prior distributions and *maximin* LHDs with 20, 100, and 500 points. It appears that the "bad" priors are discarded in the first two cases while accepted with a design of 500 points. Other priors, even for a design of 20 points, seem 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 this is not problematic. 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 requires N calls to the function H and it could be faster 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 behavior of $\widetilde{\text{DAC}}$ for the larger domain Ω_2, which is not reported here, is quite similar to that for Ω_1. It shows that the choice of domain does not affect the agreement between the prior and the data.

Figure 10, which displays the behavior of $\widetilde{\text{DAC}}$ for the PLV and FHV prior with different values for the hyperparameters a and t, shows that those hyperameters can have a sensitive impact on the result 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.

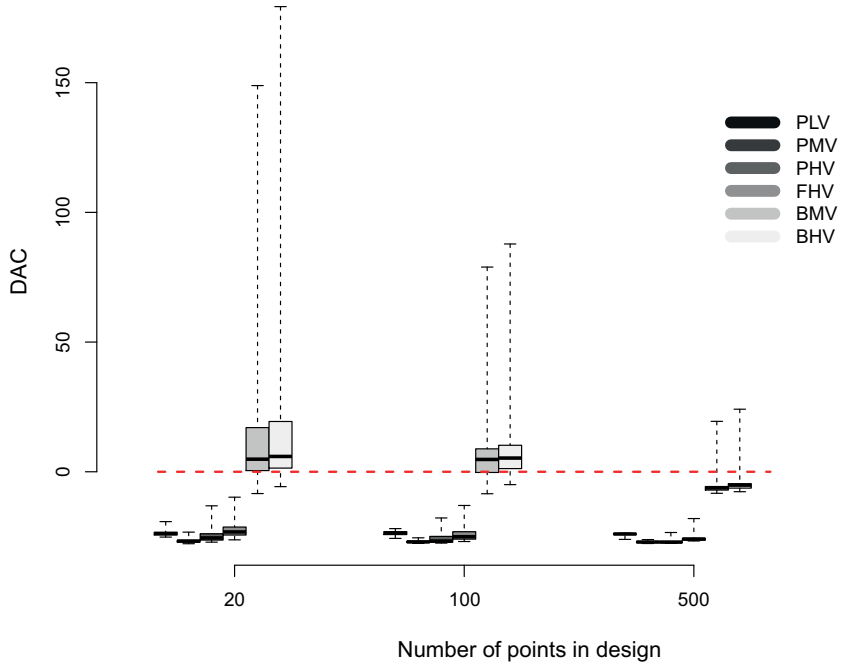


FIG. 9: $\widetilde{\text{DAC}}$ boxplots based on 100 repetitions, calculated within small domain Ω_1, for three *maximin*-LHDs of 20, 100, and 500 points and six priors PLV(—), PMV(—), PHV(—), FHV(—), BMV(—), and BHV(—).

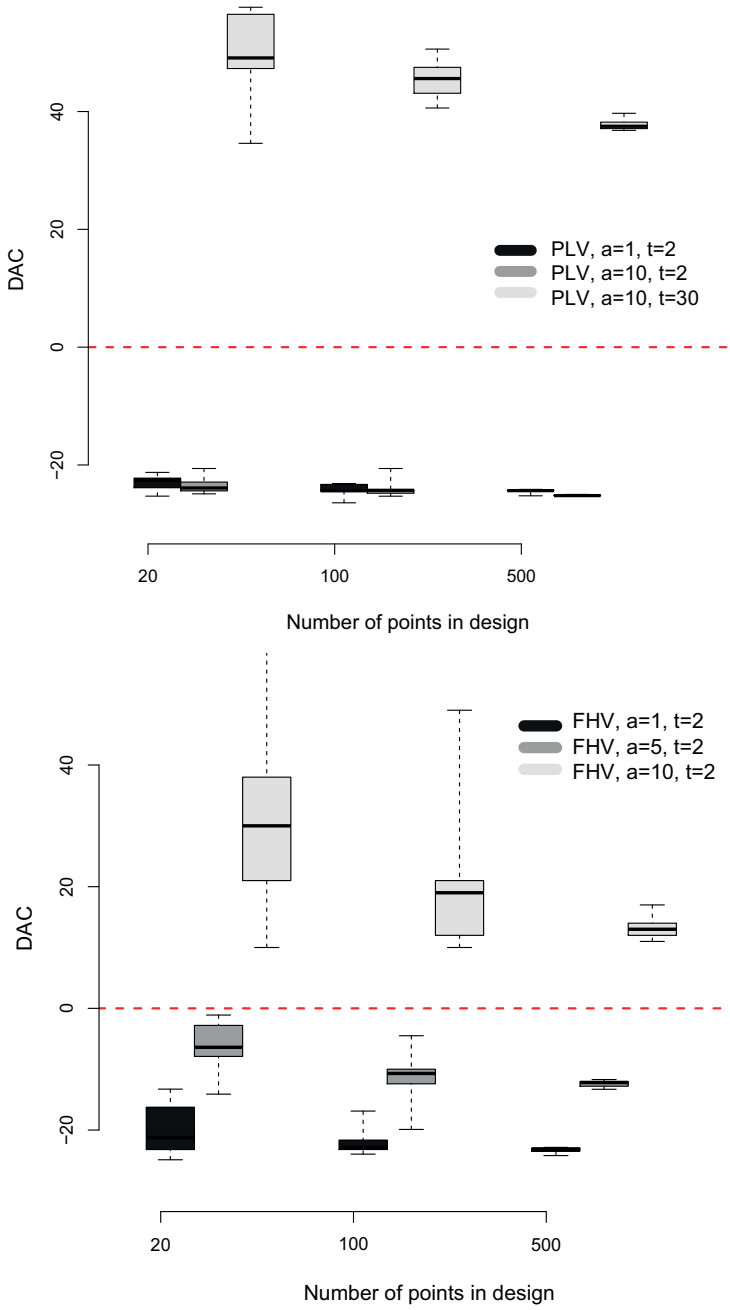


FIG. 10: $\widetilde{\text{DAC}}$ boxplots based on 100 repetitions, calculated within small domain Ω_1, for three *maximin*-LHDs of 20, 100, and 500 points and PLV and FHV priors with different values of the hyperparameters a and t.

Figures 11 and 12 display the marginal posterior distributions with a *maximum*-LHD of 100 points and 20 points. These 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, are quite similar. It is also important to notice than there is no sensitive differences between the posteriors derived from the 100 points and 20 points (which is not reported here) designs, as indicated by the $\widetilde{\text{DAC}}$ criterion.

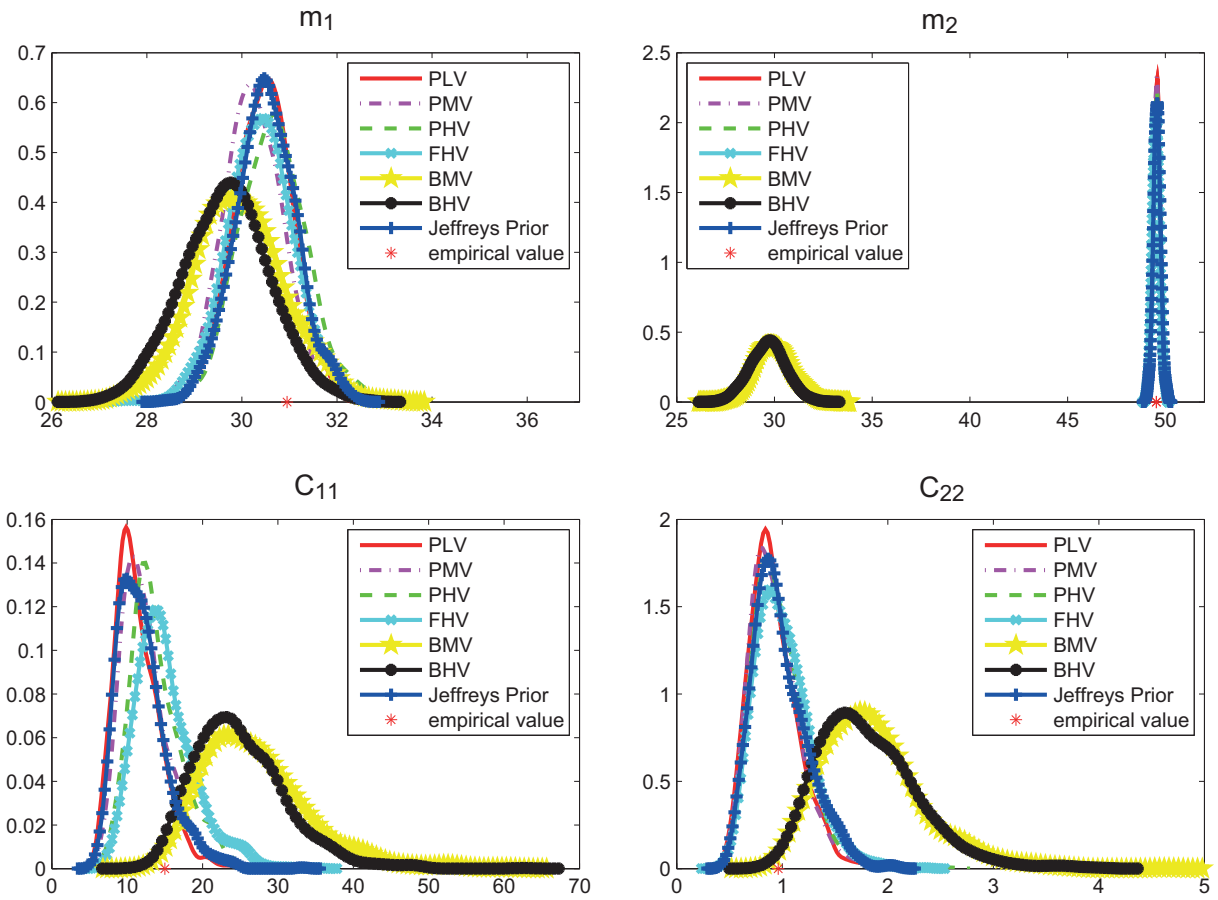


FIG. 11: Marginal posterior distributions of (m, C) with a *maximin*-LHD of 100 points in small domain Ω_1, based on six informative PLV, PMV, PHV, FHV, BMV, BHV priors and the Jefferys noninformative prior. The empirical estimation is marked by a star.

It seems that $\widetilde{\text{DAC}}$ is indicating that a reasonable prior can be resistant to a poor design. This is 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 when replacing H by the Sobol function:

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

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 Fig. 13, $\widetilde{\text{DAC}}_{18}$ remains positive for the four prior choices, which indicates the need to improve the design.

Remark. In these numerical experiments, different priors are compared using DAC in an illustrative purpose. It does mean that DAC might be used as a selection tool to select one prior over another. Indeed, prior distributions reflect prior knowledge and uncertainty about X before seeing the data. DAC is just a criterion measuring the discrepancy between the prior distribution and the data.

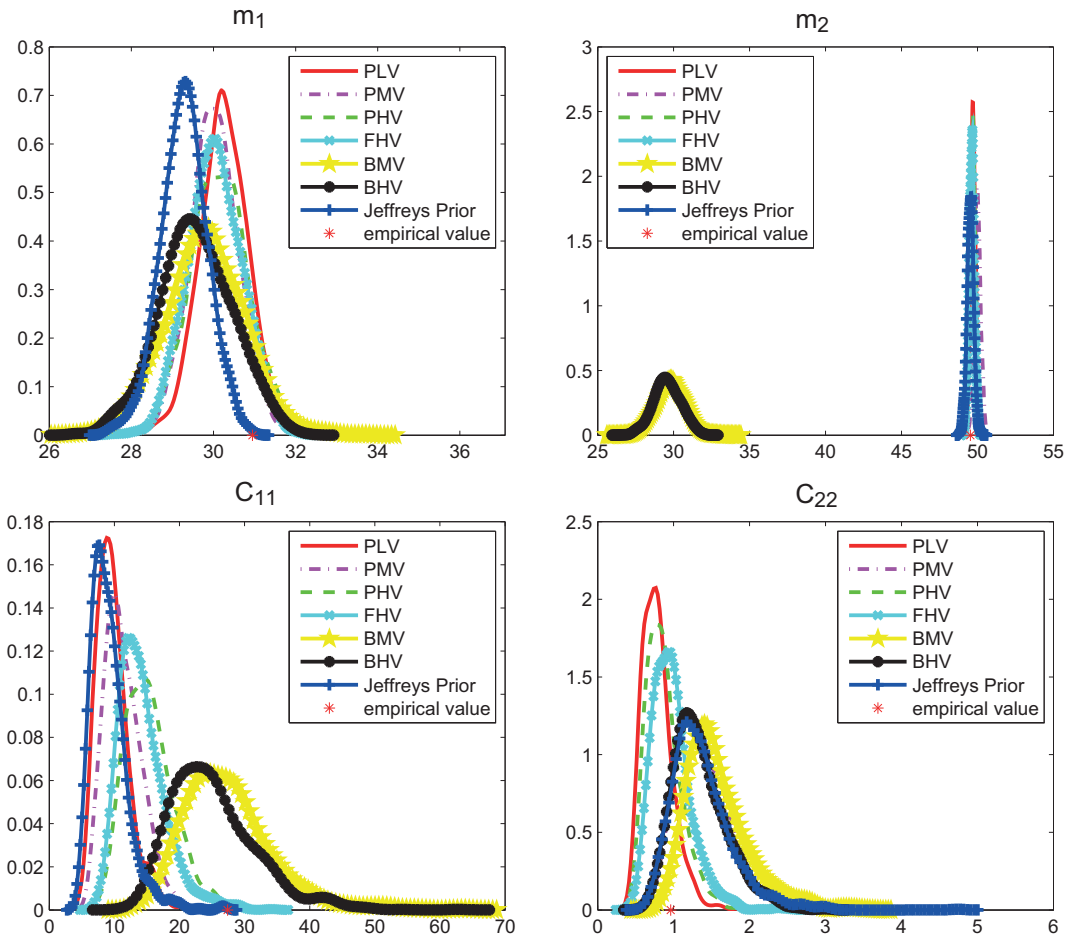


FIG. 12: Marginal posterior distributions of (m, C) with a *maximin*-LHD of 20 points in small domain Ω_1, based on six informative PLV, PMV, PHV, FHV, BMV, BHV priors and the Jefferys noninformative prior. The empirical estimation is marked by a star.

4.2 Real Case Study: the MASCARET Code

The second example considered here is a real hydraulic model, the MASCARET code, which implements an approximating solution to the St-Venant equation through finite difference methods. It is developed at EDF in collaboration with the Centre d'Études Techniques Maritimes et Fluviales (CETMEF), which puts together the computer code of free surface.

In this two-dimensional model, the main sources of uncertainty are the frictions on the riverbed and the floodplain, denoted by the vector K_s, the output of the hydraulic function is the water level, denoted by Y, and the observed input is the river flow, denoted by Q. In this case study, accounting for the observation error U, the dataset can thus be generated with the help of the MASCARET code H, as follows.

$$Y = H(K_s, Q) + U, \tag{14}$$

with $Q \sim \text{Gumbel}(1550, 780)$ and the missing data

$$K_s \sim \mathcal{N}(m, C) = \mathcal{N}\left(\begin{pmatrix} 17 \\ 40 \end{pmatrix}, \begin{pmatrix} 4.1^2 & 0 \\ 0 & 7.1^2 \end{pmatrix}\right). \tag{15}$$

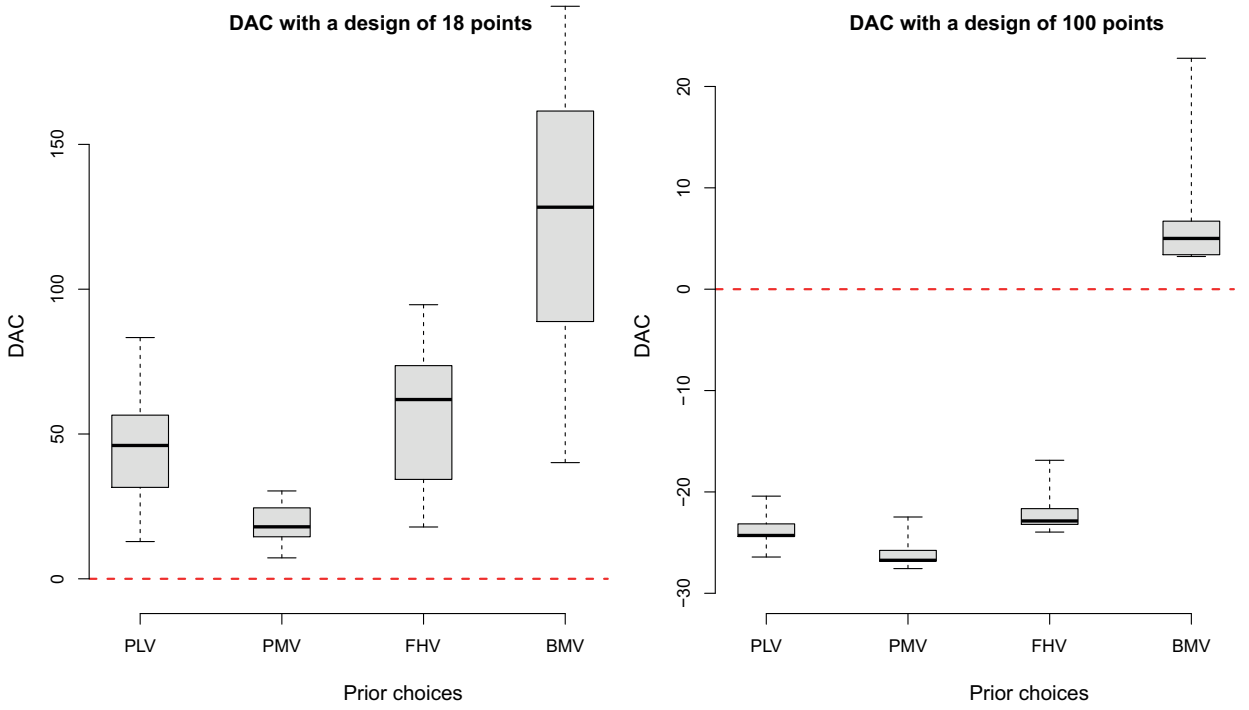


FIG. 13: $\widetilde{\text{DAC}}$ boxplots based on 20 repetitions with the Sobol function, calculated for two designs D_{18} and D_{100} and six priors PLV, PMV, FHV, and BMV.

To check the behavior of $\widetilde{\text{DAC}}$, two different prior distributions on the K_s hyperparameters are considered and summarized in Table 3.

Figure 14 displays the boxplot of $\widetilde{\text{DAC}}$ for 20 repetitions of the PMV and BLV priors, with sample size $n = 10, 50$ and a *maximin*-LHD of 20 and 200 points. It appears that the BLV prior is rejected by $\widetilde{\text{DAC}}$ in all the four cases as this criterion remains positive, and it seems almost acceptable for the last case, with 50 observed data and 200 points in the design, as $\widetilde{\text{DAC}}$ is quite near zero. Moreover, the PMV prior is obviously acceptable in each case study, thanks to the negative $\widetilde{\text{DAC}}$s. In each case this result testifies to a perfect agreement between the prior, the data, and the design.

Figure 15, displaying the corresponding marginal posterior distributions of θ, confirms the performance of $\widetilde{\text{DAC}}$. The PMV prior leads to reasonable posterior values for all the parameters while the bad prior leads to posterior values far from the good values, especially for the variance parameters.

TABLE 3: Description of the two prior distributions by their accordance with the generated data: PMV = perfect mean and medium variance, BLV = bad mean and low variance.

Prior	**PMV**	**BLV**
μ	$\{17, 40\}$	$\{5, 60\}$
a	1	1
t	2	2
ν	5	5
$\widetilde{C}_{\text{Exp}}$	$\begin{pmatrix} 4.1^2 & 0 \\ 0 & 7.1^2 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

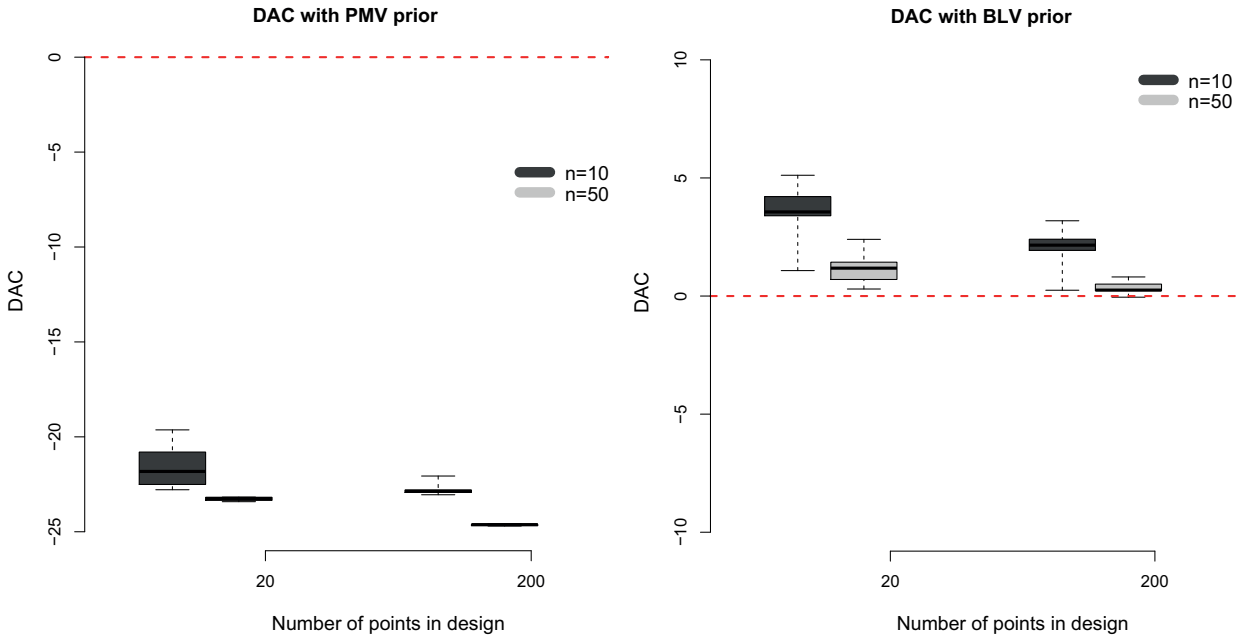


FIG. 14: $\widetilde{\text{DAC}}$ boxplots based on 20 repetitions in the MASCARET code, for two *maximin*-LHDs of 20 and 200 points, with 10 and 50 observations and PMV and BLV priors.

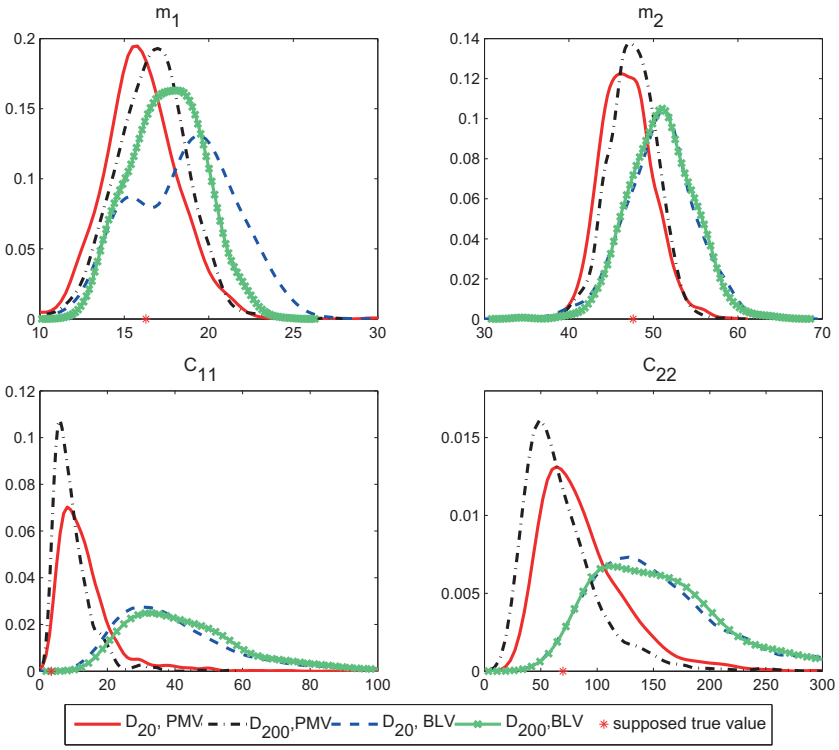


FIG. 15: Marginal posterior distributions of (m, C) in the MASCARET code, based on two informative PMV and BLV priors, two *maximin*-LHDs of 20 and 200 points and 50 observations. The supposed true value is marked by a star.

5. DISCUSSION

We have shown that Bayesian analysis is possible and beneficial to solve inverse problems by estimating the parameters of highly complex uncertainty models. Bayesian analysis is feasible 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 account properly for prior knowledge and to avoid a 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 is important 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 call to H equals the computation time of $L(N)$ iterations of the MCMC algorithm.[1] 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 real 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 2.80 GHz.

In this perspective, the four error sources listed in the Introduction can be controlled.

- 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 large, we advocate a more stringent 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 noninformative priors are available. However, this proper prior distribution is subject to the choice of this compact set. It is implicitly assumed here that this compact set has been chosen in a proper way. Note also that in our numerical experiments, the choice of this compact set does not appear to influence highly the posterior $\pi^J(\theta|\mathbf{y}, \mathbf{H}_{D_N})$ (see Fu et al. [21]).

  Our experiments show a promising behavior of this criterion. Obviously, computing $\widetilde{\text{DAC}}$ is not free since it involves running an additional MCMC algorithm for noninformative priors. But we think that the result is worth the trouble. Moreover, as soon as the MCMC with a noninformative 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 intricated. More experiments are needed to assess the relevance and sensibility of this criterion. Nonetheless, it is a promising tool to drive Bayesian inference using an emulator for dealing with complex inverse problems in uncertainty analysis.

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 elicited prior distributions while dramatically saving the number of calls to the expensive function H. And, DAC might be used as a diagnostic tool to detect if the prior distribution can be trusted. It might be also be helpful to choose a good design for the emulator.

REFERENCES

1. Celeux, G., Grimaud, A., Lefebvre, Y., and De Rocquigny, E., Identifying intrinsic variability in multivariate systems through linearised inverse methods, *Inverse Prob. Eng.*, 18:401–415, 2010.

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

2. Barbillon, P., Celeux, G., Grimaud, A., Lefebvre, Y. and Rocquigny (de), E., Non linear methods for inverse statistical problems, *Comput. Stat. Data Anal.*, 55:132–142, 2011.
3. Li, R. and Sudjianto, A., Analysis of computer experiments using penalized likelihood in Gaussian kriging models, *Technometrics*, 47:111–120, 2005.
4. Kennedy, M. C. and O'Hagan, A., Bayesian calibration of computer models, *J. R. Stat. Soc. B*, 63:425–464, 2001.
5. Brynjarsdóttir, J. and O'Hagan, A., Learning about physical parameters: The importance of model discrepancy, Technical Report, available at http://tonyohagan.co.uk/academic/pdf/simmach.pdf, 2013.
6. O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T., *Uncertain Judgements: Eliciting Experts' Probabilities*, Chichesker, UK: Wiley, 2006.
7. Tierney, L., Introduction to general state-space Markov chain theory, In *Markov Chain Monte Carlo in Practice*, Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., Eds., pp. 59–74, London: Chapman and Hall, 1995.
8. Barbillon, P., Méthodes d'interpolation à noyaux pour l'approximation de fonctions type boîte noire coûteuses, PhD thesis, Université Paris-Sud, 2010.
9. Matheron, G., The theory of regionalised variables and its applications, PhD Thesis, École Nationale Supérieure des Mines de Paris, 1971.
10. Sacks, J., Schiller, S. B., Mitchell, T. J., and Wynn, H. P., Design and analysis of computer experiments (with discussion), *Stat. Sinica*, 4:409–435, 1989.
11. Sacks, J., Schiller, S. B., and Welch, W. J., Designs for computer experiments, *Technometrics*, 31:41–47, 1989.
12. Koehler, J. R. and Owen, A. B., Computer experiments, In: *Handbook of Statistics*, Ghosh, S., Rao, C. R., Eds., Amsterdam: Elsevier, pp. 261–308, 1996.
13. Santner, T. J., Williams, B., and Notz, W., *The Design and Analysis of Computer Experiments*. Berlin: Springer-Verlag, 2003.
14. Fang, K.-T., Li, R., and Sudjianto, A., *Design and Modeling for Computer Experiments*, Boca Raton, FL: Chapman and Hall/CRC, 2006.
15. Mitchell, T., Morris, M., and Ylvisaker, D., Existence of smoothed stationary processes on an interval, *Stochastic Processes Appl.*, 35:109–119, 1990.
16. Rasmussen, C. E. and Williams, C., *Gaussian Processes for Machine Learning*, Cambridge, MA: The MIT Press, 2006.
17. Johnson, M. E., Moore, L. M., and Ylvisaker, D., Minimax and maximin distance designs, *J. Stat. Planning Inference*, 26:131–148, 1990.
18. Joseph, V. R., and Hung, Y., Orthogonal-maximin latin hypercube designs, *Stat. Sinica*, 18:171–186, 2008.
19. McKay, M. D., Beckman, R. J., and Conover, W. J., A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 21:239–245, 1979.
20. Petelet, M., Iooss, B., Asserin, O., and Marrel, A., Latin hypercube sampling with inequality constraints, *Adv. Stat. Anal.*, 94:325–339, 2010.
21. Fu, S., Celeux, G., Bousquet, N., and Couplet, M., Bayesian inference for inverse problems occurring in uncertainty analysis, Technical Report INRIA RR799, hal.inria.fr/hal-00708814/PDF/RR-7995.pdf, 2012.
22. Brooks, S. P. and Gelman, A., General methods for monitoring convergence of iterative simulations, *J. Comput. Graphical Stat.*, 7:434–455, 1998.
23. Vanderpoorten, A. and Palm, R., Compared regression methods for inferring ammonium nitrogen concentrations in rivers from aquatic bryophyte assemblages, *Hydrobiologia*, 452:181–190, 2001.
24. Bastos, L. S. and O'Hagan, A., Diagnostics for Gaussian process emulators, *Technometrics*, 51:425–438, 2008.
25. Bousquet, N., Diagnostics of prior-data agreement in applied Bayesian analysis, *J. Appl. Stat.*, 35:1011–1029, 2008.
26. Yang, R. and Berger, J. O., A Catalog of non-informative priors, *ISDS Discus. Paper* 97–42, 1998.
27. Munoz-Zuniga, M., Garnier, J., Remy, E. and de Rocquigny, E., Analysis of adaptive directional stratification for the controlled estimation of rare event probabilities, *Stat. Comput.*, 22:809–821, 2011.
28. Gelman, A. and Rubin, D. Inference from iterative simulation using multiple sequences, *Stat. Sci.*, 7:457–511, 1992.
29. Thisted, R. A., *Elements of Statistical Computing*, London: Chapman and Hallm, 1988.

30. Horn, R. A. and Johnson, C. R., *Topics in Matrix Analysis*, Cambridge, UK: Cambridge University Press, 1991.

APPENDIX A. BROOKS-GELMAN STATISTICS

In 1998, Brooks and Gelman proposed a method derived from the method proposed by Gelman and Rubin [28], for monitoring the convergence of iterative simulations [22]. 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 $100(1-(\alpha/2))\%$ and $100(\alpha/2)\%$ 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;
- $\bar{\delta} = 1/m \sum_{j=1}^{m} \theta_j$, with θ_j the length of the within-sequence interval for the jth 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 3000 consecutive iterations.

APPENDIX B. COMPUTING DAC FOR THE KRIGING EMULATOR

The compact set $\Omega_m = \Omega = \Omega_1 \times \ldots \times \Omega_q$ where Ω_i denotes the domain for the ith coordinate of X. To determine the compact set Ω_C related to the variance matrix C, it is convenient to consider its eigenvalue decomposition $C = VDV^T$ 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 $i = 1, \ldots, q$, $X_i^2 \leq \beta_i = \max((\max \Omega_i)^2, (\min \Omega_i)^2)$. On the other hand, recalling that R is the variance matrix of the measurement error in (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:

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

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, i.e.,

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

with

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

Thus

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

where

$$\Omega_D = \left\{ D \in \mathcal{DS}_q^+ \ st.\ |D| \geq |R|^{q/p},\, 0 \leq D_{ii} \leq \sqrt{\sum_{j=1}^{q} \beta_j^2},\ i = 1, \ldots, q \right\}. \tag{B.3}$$

Now, any orthogonal matrix V of dimension q is characterized by the composition of $q(q-1)/2$ rotations $(\psi_1, \ldots, \psi_{q(q-1)/2})$ (cf. Thiested [29]),

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

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 $\int_{\Omega_D} (1/|D|^{(q+2)/2}) dD$. Denoting it $I(q, a, \beta_1, \ldots, \beta_q)$, with $a = |R|^{q/p}$ it is derived by induction on q (the detailed calculation is given in Appendix C and related theories can be found in Horn and Johnson [30]).

$$I\left(q, a, \beta_1, \ldots, \beta_q\right) = \left(\frac{q-1}{q}\right)^{q-1} I\left(q-1, \left(\frac{a}{\beta_q}\right)^{\frac{q}{q-1}}, \beta_1^{\frac{q}{q-1}}, \ldots, \beta_{q-1}^{\frac{q}{q-1}}\right), \tag{B.4}$$

and

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

APPENDIX C. COMPUTING THE NORMALIZING CONSTANT OF THE DIAGONAL VARIANCE MATRIX DOMAIN

Consider

$$I = \int_{\Omega_C} \frac{1}{|C|^{\frac{q+2}{2}}} dC, \tag{C.1}$$

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

$$\Omega_C = \left\{ C \in \mathcal{S}_q^+ \ st.\ |C| \geq |R|^{q/p},\, |C_{ij}| \leq \sqrt{\beta_i \beta_j},\ i, j = 1, \ldots, q \right\}. \tag{C.2}$$

Since C is diagonal, the above definition is equivalent to

$$\begin{cases} 0 \leq C_i \leq \beta_i, \\ \prod_{i=1}^{q} C_i \geq a, \end{cases} \tag{C.3}$$

where $\{C_i, 1 \leq i \leq q\}$ are the diagonal elements of C. Conditions (C.3) are equivalent to the conditions

$$\begin{cases} \dfrac{a}{\beta_2 \cdots \beta_q} \leq C_1 \leq \beta_1, \\ \dfrac{a}{C_1 \beta_3 \cdots \beta_q} \leq C_2 \leq \beta_2, \\ \vdots \\ \dfrac{a}{C_1 C_2 \cdots C_{q-1}} \leq C_q \leq \beta_q. \end{cases} \tag{C.4}$$

Considering I as a function of $(q, a, \beta_1, \ldots, \beta_q)$, the integral (C.1) can be developed as follows:

$$\begin{aligned} I(q, a, \beta_1, \ldots, \beta_q) &= \int_{\frac{a}{\beta_2 \cdots \beta_q}}^{\beta_1} \frac{1}{C_1^{\frac{q+2}{2}}} dC_1 \int_{\frac{a}{C_1 \beta_3 \cdots \beta_q}}^{\beta_2} \frac{1}{C_2^{\frac{q+2}{2}}} dC_2 \cdots \int_{\frac{a}{C_1 \cdots C_{q-1}}}^{\beta_q} \frac{1}{C_q^{\frac{q+2}{2}}} dC_q \\ &= \frac{2}{q a^{\frac{q}{2}}} \int_{\frac{a}{\beta_2 \cdots \beta_q}}^{\beta_1} \frac{1}{C_1} dC_1 \int_{\frac{a}{C_1 \beta_3 \cdots \beta_q}}^{\beta_2} \frac{1}{C_2} dC_2 \cdots \int_{\frac{a}{C_1 \cdots C_{q-2} \beta_q}}^{\beta_{q-1}} \frac{1}{C_{q-1}} dC_{q-1} \\ &- \frac{2}{q \beta_q^{\frac{q}{2}}} \int_{\frac{a}{\beta_2 \cdots \beta_q}}^{\beta_1} \frac{1}{C_1^{\frac{q+2}{2}}} dC_1 \int_{\frac{a}{C_1 \beta_3 \cdots \beta_q}}^{\beta_2} \frac{1}{C_2^{\frac{q+2}{2}}} dC_2 \cdots \int_{\frac{a}{C_1 \cdots C_{q-2} \beta_q}}^{\beta_{q-1}} \frac{1}{C_{q-1}^{\frac{q+2}{2}}} dC_{q-1} \\ &= \frac{2}{q a^{\frac{q}{2}}} I_{q-1} - \frac{2}{q \beta_q^{\frac{q}{2}}} \left(\frac{q-1}{q} \right)^{q-1} I\left(q-1, \left(\frac{a}{\beta_q} \right)^{\frac{q}{q-1}}, \beta_1^{\frac{q}{q-1}}, \ldots, \beta_{q-1}^{\frac{q}{q-1}} \right), \end{aligned} \tag{C.5}$$

where

$$\begin{aligned} I_{q-1} &= \int_{\frac{a}{\beta_2 \cdots \beta_q}}^{\beta_1} \frac{1}{C_1} dC_1 \int_{\frac{a}{C_1 \beta_3 \cdots \beta_q}}^{\beta_2} \frac{1}{C_2} dC_2 \cdots \int_{\frac{a}{C_1 \cdots C_{q-2} \beta_q}}^{\beta_{q-1}} \frac{1}{C_{q-1}} dC_{q-1} \\ &= \frac{1}{(q-1)!} \left(\log \frac{\beta_1 \ldots \beta_q}{a} \right)^{q-1}, \end{aligned} \tag{C.6}$$

is obtained by induction and

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

by the variable change

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

Thus step by step thanks to Eq. (C.5), the integral can be calculated when C is diagonal. For instance, for $q = 2, 3, 4$ we get

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

$$I\left(3,a,\beta_1,\beta_2,\beta_3\right)=\frac{1}{3a^{\frac{3}{2}}}\left(\log\frac{\beta_1\beta_2\beta_3}{a}\right)^2-\frac{4}{9a^{\frac{3}{2}}}\log\frac{\beta_1\beta_2\beta_3}{a}-\frac{8}{27\left(\beta_1\beta_2\beta_3\right)^{\frac{3}{2}}}+\frac{8}{27a^{\frac{3}{2}}},$$

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