

# Calibration of Computer Models with Multivariate Output

Rui Paulo<sup>a,\*</sup>, Gonzalo García-Donato<sup>b</sup>, Jesús Palomo<sup>c</sup>

<sup>a</sup>*Department of Mathematics and CEMAPRE, ISEG, Technical University of Lisbon, Portugal*

<sup>b</sup>*Department of Economics and Finance, Universidad de Castilla-La Mancha, Spain*

<sup>c</sup>*Department of Business Administration (Finance), Rey Juan Carlos University, Spain*

---

## Abstract

The problem of calibrating computer models that produce multivariate output is considered, with a particular emphasis on the situation where the model is computationally demanding. The proposed methodology builds on Gaussian process-based response-surface approximations to each of the components of the output of the computer model to produce an emulator of the multivariate output. This emulator is then combined in a statistical model involving field observations, which is then used to produce calibration strategies for the parameters of the computer model. The results of applying this methodology to a simulated example and to a real application are presented.

*Keywords:* Computer model, validation, Gaussian process, linear model of coregionalization, Bayesian analysis.

---

## 1. Introduction

Computer models are implementations of sophisticated mathematical models that aim at reproducing a particular real process. Before we can run these computer models, we have to specify a vector of inputs which typically includes calibration parameters. Scientists are thus often interested in combining data obtained in computer model runs and in physical experiments to determine estimates of these calibration parameters which make the output of the model “match” in some sense the real process. This is what is meant by calibration of a computer model.

When a physical experiment is conducted to obtain field data, it is sometimes the case that several (presumably correlated) measurements are taken. In the realm of the computer model, it is also not uncommon that the output contains in fact a wealth of information, some of which can even be considered irrelevant for the actual calibration problem at hand. Clearly, the evaluation of a model depends on how one intends to use it, and as part of the calibration process one must specify which particular aspect(s) of the output are of main interest. In practice, the spirit underlying the specification of these aspects is often a mix of genuine practical interest and the need to keep

---

\*Corresponding author. Tel.: +351 21 3925833

*Email addresses:* [rui@iseg.utl.pt](mailto:rui@iseg.utl.pt) (Rui Paulo), [gonzalo.garciadonato@uclm.es](mailto:gonzalo.garciadonato@uclm.es) (Gonzalo García-Donato), [jesus.palomo@urjc.es](mailto:jesus.palomo@urjc.es) (Jesús Palomo)

the analysis feasible. This paper proposes methodology to address the following situation: one is simultaneously interested in several univariate aspects of the output, but the joint calibration of the computer model against these aspects is perceived as too difficult to tackle. In this situation, a simplistic strategy would consider separate calibration exercises, one per univariate component of the output. Besides being formally incorrect, the main drawback associated with this approach is that it potentially produces one calibration strategy for each of the components considered. This may result in an illusory perception of the quality of the computer model as a representation of reality, as there may be calibration strategies which accommodate for the individual discrepancies very well, but not one that does so globally. Also, there may be too much variability — or several sensible calibration strategies — associated with calibrating each univariate output separately, which may render the analyses inconclusive. By combining all sources of information, one may be able to substantially reduce the uncertainty and pin down a particular calibration strategy. Section 2 illustrates these ideas with a motivating example.

The purpose of the present article is to propose methodology to deal with the calibration of a computer model that produces multivariate output by building on methodology already developed for the univariate case. The main motivation is to produce methods that are computationally not very demanding and hence have the potential to scale-up to more complicated problems. Although the emulator we propose is valid in complete generality, the calibration methodology that we develop in this paper is restricted to the case where only one configuration of the vector of controllable inputs is exercised to obtain the field data. Possible extensions of this setup are discussed in Section 8.

### *1.1. Background*

The approach to computer model calibration that we espouse is Bayesian and has roots in Craig et al. (1996), Kennedy and O’Hagan (2001), Kennedy et al. (2002), Higdon et al. (2004), and most directly in Bayarri et al. (2007b). It hinges on modeling the relationship between reality and computer model output in a Bayesian fashion introducing the notion of bias, effectively combining model and field data to produce estimates of the calibration parameters. A key ingredient of the methodology is the use of Gaussian processes as priors for unknown functions. This technique dates back at least to O’Hagan (1978) although its use in the realm of the computer model world is more recent: Sacks et al. (1989) and Currin et al. (1991). The end result is a Gaussian process response-surface approximation to the output of the computer model, and its associated measure of uncertainty, which is sometimes called the ‘emulator’. This is a key component of the analysis, as frequently the models are computationally very demanding and this precludes their direct evaluation within Markov chain Monte Carlo algorithms. The construction of this emulator and its subsequent integration in the calibration process clearly depends on the characteristics of the output. To give a few examples, Bayarri et al. (2007b) and Higdon et al. (2004) deal with scalar output, Bayarri et al. (2009) with functional output that is very smooth, Bayarri et al. (2007a) with very irregular functional output. Higdon et al. (2008) utilizes a related approach to deal with high-dimensional

output like images. The issue of emulating a computer model that exhibits nonstationary behavior is dealt with by means of partitioning in Gramacy and Lee (2008a) and Gramacy and Lee (2008b). Reichert et al. (2011) have recently proposed an emulator for dynamic models with application in hydrology.

Methodological developments to directly model multivariate aspects of a computer model’s output have been addressed in several papers, including Habib et al. (2007), Rougier (2008), McFarland et al. (2008), Higdon et al. (2008), Higdon et al. (2010), Conti and O’Hagan (2010), Fricker et al. (2010), Bhat et al. (2010) and Wilkinson (2010). In Section 4 we describe some of these approaches in the common framework of the linear model of coregionalization (Gelfand et al., 2004).

## 1.2. Overview

The remainder of this paper is organized as follows: in Section 2 we use a simple example to illustrate the drawbacks of the simplistic strategy of independent calibration analyses, one per dimension of the multivariate output. Section 3 precisely describes the problem we want to address and sets up the necessary notation. In Section 4 we describe previous contributions in the literature to produce a multivariate emulator, highlighting the fact that they can all be seen as special cases of the linear model of coregionalization. We also describe the approach that we propose in this paper. The problem of embedding this emulator in a calibration setting is addressed in Section 5. Section 6 describes how we propose to go about implementing our calibration methodology in practice, and Section 7 presents the results of applying it to a real application.

## 2. A motivating example

In order to highlight the drawbacks that may result from performing separate analyses to handle responses that are multivariate in nature, we introduce the following example. Additionally, we suggest how these problems can be addressed under the combined approach we propose. We skip technical details and formal definitions, since these will be accurately presented in the remainder of the paper in more general scenarios.

Suppose that the real process, depending on a scalar  $u$ , is the bivariate function  $\mathbf{y}^R(u) = (y_1^R(u), y_2^R(u))'$ ,  $u \in (0, 0.5)$ , where  $y_1^R(u) = \exp(-1.4u - 0.05/u) \cos(7\pi u/2)$  and

$$y_2^R(u) = 0.1 \exp(-1.4u - 0.05/u) \left[ -1.4 + \frac{0.05}{u^2} \cos(7\pi u/2) - \frac{7\pi}{2} \sin(7\pi u/2) \right].$$

The function  $y_1^R(u)$  has been considered by Santner et al. (2003), by Joseph (2006) and, more recently, by Liu et al. (2009) for illustrative purposes. Although essentially irrelevant for our intended use of  $y_2^R(u)$ , this function was constructed as a small distortion of the first derivative of  $y_1^R(u)$  times the factor 0.1. This factor was imposed to allow both functions to have a similar scale.

As in Santner et al. (2003), we assume that the computer model that aims at reproducing the first component of the real response is given by  $y_1^M(u) = \exp(-1.4u) \cos(7\pi u/2)$ , while we consider

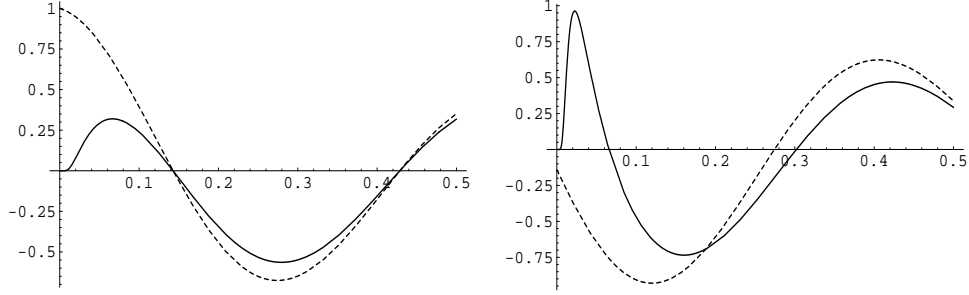


Figure 1: Left:  $y_1^R(u)$  (solid line) and  $y_1^M(u)$  (dashed line). Right:  $y_2^R(u)$  (solid line) and  $y_2^M(u)$  (dashed line).

its first derivative, that is,

$$y_2^M(u) = 0.1 \exp(-1.4u) \left[ -1.4 \cos(7\pi u/2) - \frac{7\pi}{2} \sin(7\pi u/2) \right],$$

for the second component of the real process. We show in Figure 1, by components, both the real process and the computer model characterized by the functions described above.

We will treat the computer model  $\mathbf{y}^M(u) = (y_1^M(u), y_2^M(u))'$  as unknown, and hence use throughout the calculations a response-surface approximation to its output. The information necessary to construct this approximation was obtained by evaluating the computer model at the set of equally-spaced design points  $\{0.05, 0.15, 0.25, 0.35, 0.45\}$ . By treating  $u$  as a calibration input, we want to simulate field experiments that will both allow us to estimate the “true value” of  $u$  — which we henceforth denote by  $u^*$  —, and to estimate all other unknowns in the statistical model. Let  $u^* = 0.2$  be the “true value” of  $u$ ; in this scenario,  $\mathbf{y}^R(u^*) = (-0.346, -0.649)'$ . We now consider  $K = 7$  simulated replicates of the field experiment,  $\mathbf{y}_k^F(u^*)$ ,  $k = 1, \dots, K$ , where  $\mathbf{y}_k^F(u^*) \sim N_2(\mathbf{y}^R(u^*), \Sigma^F)$ , with

$$\Sigma^F = \begin{pmatrix} 1/400 & 0.5/400 \\ 0.5/400 & 1/400 \end{pmatrix}.$$

Notice that, in this situation,  $y_1^M(u) = y_1^R(u^*)$  has two possible solutions,  $u \in \{0.185, 0.372\}$ ; similarly, the equation  $y_2^M(u) = y_2^R(u^*)$  has two roots,  $u \in \{0.050, 0.194\}$ . Consequently, the posterior densities of  $u^*$  are potentially bimodal.

Firstly, we *separately* apply the univariate methodology in Bayarri et al. (2007b) to each of the components. We represent the posterior densities obtained for  $u^*$  in Figure 2. Due to the multimodality of the densities, the mean and median are not very meaningful as summaries of the distributions. The two modes of  $u^*$ , using the first component, are  $\{0.186, 0.369\}$  (the second being the global maximum). The modes of the posterior density of  $u^*$  using the second component are  $\{0.051, 0.192\}$  (again the second being the global maximum). In a situation like this, one is then faced with the problem of how to combine the two conflicting analyses.

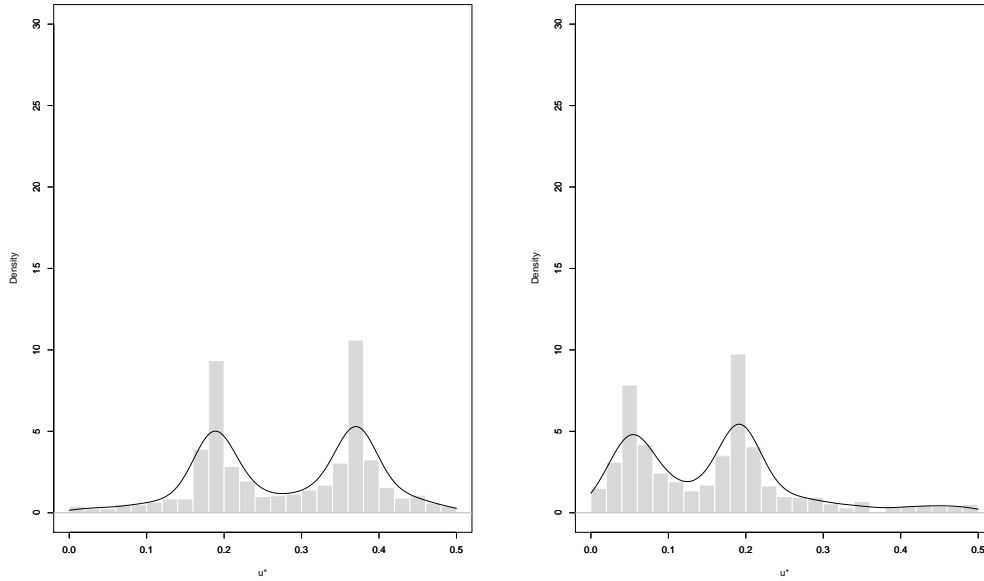


Figure 2: Posterior density of  $u^*$  obtained from the independent analysis of each component of the output (left panel: first component of the output; right panel: second component)

Next, we explicitly recognize the multivariate nature of the responses and apply the methodology introduced in the present article to these data. We now obtain only one posterior distribution for  $u^*$ , represented in Figure 3, which results from combining all the available information. The idea of producing a single calibration policy for the *same* quantity was the main motivation for this research. The posterior distribution of  $u^*$  now has three modes, located at  $\{0.006, 0.191, 0.354\}$ , of which two of them, 0.006 and 0.354, barely concentrate any probability mass. One way of interpreting this result is that the spurious modes are penalized by the multivariate approach since they provide (comparatively) unreasonable predictions for one of the components. On the other hand, the values for the calibration parameter that result in predictions that are close to the real value for both of the components are strengthened in the joint analysis. Notice also how the uncertainty about the calibration parameter decreases substantially in the combined analysis. By combining model and field data in a statistical model, we can produce estimates of reality that take into account both sources of information — the so-called bias-corrected prediction of Bayarri et al. (2007b). For the separate and joint analyses, these are summarized in Table 1. Again, what stands out is the decrease in uncertainty present in the conclusions when we move from the separate to the joint analysis.

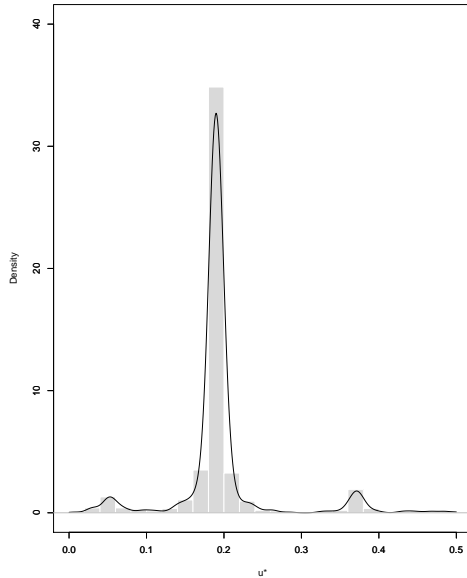


Figure 3: Posterior density of  $u^*$  obtained from the combined analysis.

Approach	$\hat{y}_1^R(u^*)$		$\hat{y}_2^R(u^*)$	
	Mean	95% CI	Mean	95% CI
Separate	-0.372	(-0.621, -0.285)	-0.673	(-1.030, -0.417)
Combined	-0.353	(-0.373, -0.334)	-0.661	(-0.690, -0.633)

Table 1: Bias-corrected prediction (see Bayarri et al. (2007b)) of reality. Posterior mean and 95% credible interval for the bias-corrected prediction of  $\mathbf{y}^R(u^*) = (-0.346, -0.649)'$ .

### 3. Formulation of the problem

We conduct a physical experiment on a real system to obtain noisy measurements of a  $p$ -dimensional quantity of interest  $\mathbf{y}^R = (y_1^R, \dots, y_p^R)'$ . We denote the resulting replicate observations by  $\mathbf{y}_k^F = (y_{k1}^F, \dots, y_{kp}^F)'$ ,  $k = 1, \dots, K$ . In addition, we have access to a computer implementation of a scientific model that purports to reproduce  $\mathbf{y}^R$ . Given  $\mathbf{u}$ , a concrete value for a vector of unknown characteristics of the real system, the model produces an output that we denote by  $\mathbf{y}^M(\mathbf{u}) = (y_1^M(\mathbf{u}), \dots, y_p^M(\mathbf{u}))'$ . The goal is to perform inference on the value of  $\mathbf{u}$  which makes the output of the computer model as close as possible, in some sense, to the quantity  $\mathbf{y}^R$ . To achieve that goal, we model the relationship between computer model output and real response by

$$\mathbf{y}^R = \mathbf{y}^M(\mathbf{u}^*) + \mathbf{b}$$

where  $\mathbf{u}^*$  is the “true value” of  $\mathbf{u}$ . The object  $\mathbf{b}$  is defined by subtraction, and is called the bias or model discrepancy term, a central quantity in the calibration exercise. For a recent account on

this notion see Goldstein (2010). Field measurements of the real response process are subject to error, so we model the  $K$  replicates of the field observation independently as

$$\mathbf{y}_k^F \mid \mathbf{y}^M(\mathbf{u}^*), \mathbf{b} \sim N_p(\mathbf{y}^M(\mathbf{u}^*) + \mathbf{b}, \boldsymbol{\Sigma}^F), \quad k = 1, \dots, K, \quad (1)$$

where  $\boldsymbol{\Sigma}^F$  is a  $p \times p$  positive definite matrix.

To estimate the unknowns in (1) we adopt a Bayesian approach. That entails using simulation methods that require the evaluation of the likelihood that results from (1) a very large number of times. This is incompatible with the fact that the evaluation of  $\mathbf{y}^M(\mathbf{u})$ , for a single  $\mathbf{u}$ , takes, in many instances, several minutes. A solution to this problem, when  $p = 1$ , has been to treat the computer model as an unknown and assign it a Gaussian process prior (see, e.g., Sacks et al. (1989), Currin et al. (1991), Kennedy and O’Hagan (2001), Higdon et al. (2004), Bayarri et al. (2007b) among many others.) The natural extension of this approach to the  $p > 1$  case would be to consider a  $p$ -dimensional Gaussian process prior. The difficulties associated with this extension include the increase in computational complexity, the specification of cross-correlation functions that are sufficiently flexible and yet are computationally tractable, and the need to construct new software to deal, efficiently and accurately, with this class of stochastic processes. The linear model of coregionalization can be thought of as a constructive definition of a class of multivariate Gaussian processes. In the next section, we describe this model and its different uses that we encounter in the literature in the emulation of computer models. We end by describing our own proposal.

#### 4. Coregionalization emulators

Suppose that the computer model  $\mathbf{y}^M(\mathbf{z})$  is exercised at a set of design points  $D^M = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$ , and denote by  $\mathbf{y}_D^M$  the  $p \times N$  matrix with column  $i$  equal to the corresponding observed  $\mathbf{y}^M(\mathbf{z}_i)$ . The input vector  $\mathbf{z}$  is  $q$ -dimensional, and, in this section, may include a vector of controllable inputs in addition to the vector of calibration inputs  $\mathbf{u}$ . Once a prior model has been specified for  $\mathbf{y}^M(\cdot)$ , we consider the conditional distribution

$$f(\mathbf{y}^M(\mathbf{z}^*) \mid \mathbf{y}_D^M, \boldsymbol{\theta}^M), \quad (2)$$

where  $\mathbf{z}^*$  is a generic input value and  $\boldsymbol{\theta}^M$  are unknown parameters. Strategies to use (2) as an emulator of the computer model include placing a prior distribution on  $\boldsymbol{\theta}^M$  and performing a full Bayesian analysis or, alternatively, finding an estimate of  $\boldsymbol{\theta}^M$  and plugging-in this estimate in the conditional posterior (2). One key ingredient in constructing the emulator is therefore the choice of the prior model for  $\mathbf{y}^M(\cdot)$ . Most of the models that have been used in the literature to construct emulators of multivariate computer models can be seen as particular cases of the formulation considered in Gelfand et al. (2004) called linear model of coregionalization. In a simpler form, this model dates back to Matheron (1982).

#### 4.1. Linear model of coregionalization

The linear model of coregionalization (LMC) constructs a  $p$ -variate Gaussian process  $\mathbf{y}^M(\cdot)$  using as building blocks  $r$  independent univariate Gaussian processes  $w_i(\cdot)$ . To be more precise, it assumes that

$$\mathbf{y}^M(\mathbf{z}) = \boldsymbol{\mu}(\mathbf{z}) + \mathbf{A}\mathbf{w}(\mathbf{z}) + \boldsymbol{\epsilon}, \quad (3)$$

where  $\mathbf{A}$  is a  $p \times r$  full column rank matrix with  $r \leq p$ . The nugget effect  $\boldsymbol{\epsilon}$  is independent of  $\mathbf{w}(\cdot)$ , and  $\boldsymbol{\epsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{D})$ , with  $\mathbf{D}$  a diagonal matrix. The mean of the multivariate process is specified through a linear model (one for each component) on a known function of the input  $\mathbf{z}$ , that is,

$$\boldsymbol{\mu}(\mathbf{z}) = (\mathbf{I}_p \otimes \mathbf{h}(\mathbf{z})')\boldsymbol{\eta}, \quad \mathbf{h}(\mathbf{z}) \in \mathbb{R}^s, \quad (4)$$

where  $\boldsymbol{\eta}$  is a  $(p \times s)$ -dimensional vector of unknown parameters. The components  $w_i(\mathbf{z})$  in  $\mathbf{w}(\mathbf{z})$  are independent Gaussian processes with zero mean, precision  $\lambda_i^M$  and correlation structure  $c_i^M(\mathbf{z}, \mathbf{z}^*)$ . One commonly used correlation function in computer models is the separable power exponential correlation function

$$c_i^M(\mathbf{z}, \mathbf{z}^*) = \exp \left( - \sum_{j=1}^q \beta_{ij}^M |z_j - z_j^*|^{2-\alpha_{ij}^M} \right), \quad (5)$$

with  $\alpha_{ij}^M \in [0, 1]$  and  $\beta_{ij}^M > 0$ . We denote by  $\boldsymbol{\alpha}_i^M$  and  $\boldsymbol{\beta}_i^M$  the corresponding vectors.

#### 4.2. LMC in emulators of computer models

We now review some of the multivariate emulators proposed in the literature, with a particular emphasis on highlighting how they fit into the common framework of the LMC. The emulator proposed by Conti and O'Hagan (2010) is not introduced in the form (3) but by directly modeling the multivariate process. In particular, they assume for  $\mathbf{y}^M(\mathbf{z})$  a  $p$ -dimensional Gaussian process with mean  $\boldsymbol{\mu}(\mathbf{z})$  as in (4) with  $\mathbf{h}(\mathbf{z}) = (1, \mathbf{z}')'$ , and separable covariance

$$\text{Cov}(\mathbf{y}^M(\mathbf{z}), \mathbf{y}^M(\mathbf{z}^*)) = c^M(\mathbf{z}, \mathbf{z}^*) \boldsymbol{\Sigma}^M$$

with power exponential correlation function  $c^M(\cdot, \cdot)$  with  $\alpha_{ij}^M \equiv 0$ . It can be easily seen that this formulation is equivalent to the LMC model in (3) without the nugget effect, with  $c_i^M \equiv c^M$  and  $\mathbf{A}$  being any matrix such that  $\mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}^M$ . Note that the separability here is between the within-location and across-location correlation. This separability assumption is central to the way the authors implement the emulator. By assuming the objective (improper) prior  $\pi(\boldsymbol{\eta}, \boldsymbol{\Sigma}^M \mid \boldsymbol{\beta}^M) \propto |\boldsymbol{\Sigma}^M|^{-(p+1)/2}$  they show that  $(\boldsymbol{\eta}, \boldsymbol{\Sigma}^M)$  can be analytically integrated out resulting in an emulator (2) that, conditionally on  $\boldsymbol{\beta}^M$ , has a closed-form expression in terms of a matrix-variate Student distribution. Since  $\boldsymbol{\Sigma}^M$  is integrated out, the arbitrariness in defining the matrix  $\mathbf{A}$  is not a matter of concern. Unfortunately,  $\boldsymbol{\beta}^M$  cannot be integrated out analytically and has to be either integrated out numerically or replaced by an estimate. Any of these alternatives comes with a computational challenge (in terms of integration or optimization) for which there is not, to the best



of our knowledge, specific software readily available. In the specific context of spatial data, Bhat et al. (2010) implement a similar emulator, assuming the same type of separable cross-correlation structure. Another reference following the separability assumption is Rougier (2008).

Fricker et al. (2010) generalize Conti and O’Hagan (2010) by relaxing the separability assumption and making direct use of the LMC formulation. That is, they use the model in (3) but, unlike Conti and O’Hagan (2010), do not assume that the components of  $\mathbf{w}(\mathbf{z})$  are identically distributed independent Gaussian processes. In their implementation, firstly  $\boldsymbol{\eta}$  is analytically integrated out with respect to the improper prior  $\pi(\boldsymbol{\eta}) \propto 1$ . The remaining parameters (i.e.,  $\boldsymbol{\Sigma}^M$ ,  $\boldsymbol{\lambda}^M$  and  $\boldsymbol{\beta}^M$ ) are replaced by marginal maximum likelihood estimates. Although the authors do not report full details on how they perform such optimization task, our own experience with likelihood surfaces arising from Gaussian processes suggests that the calculation of such estimates is a delicate problem. Once  $\boldsymbol{\Sigma}^M$  is determined, one is still left with the task of determining  $\mathbf{A}$  (recall that any  $\mathbf{A}$  such that  $\mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}^M$  is valid). This ambiguity is resolved by defining  $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}$  where  $\boldsymbol{\Sigma}^M = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$  is the spectral decomposition of  $\boldsymbol{\Sigma}^M$ . It is then proved that, with this choice, the covariance matrix is invariant under arbitrary reordering of the outputs.

Neither Conti and O’Hagan (2010), Fricker et al. (2010) nor Rougier (2008) embed their proposed emulator in a statistical framework that allows for the calibration of the underlying computer model. This is an important difference with the work in Bhat et al. (2010) which addresses the calibration problem in the context of spatial data; and Habib et al. (2007) and Higdon et al. (2008) (HH in what follows) which explicitly consider calibration in multivariate computer models with high-dimensional output, for which a reduction in dimensionality is needed.

For reasons which have to do with the reduction in dimensionality, the model in HH is introduced in a principal components setting. Nevertheless, it can be seen that the HH emulator can be formulated as an LMC as in (3), with  $\mathbf{A}$  being an  $p \times r$  rectangular matrix with  $r \ll p$ . Notice that this is where the reduction in dimensionality takes place: the  $p$ -dimensional process  $\mathbf{y}^M(\mathbf{z})$  is mapped into the (lower)  $r$ -dimensional process  $\mathbf{w}(\mathbf{z})$ . As a consequence, a nugget effect is included in the model in an effort to compensate for the inexact mapping. As in the approaches above, for each of the components of  $\mathbf{w}$ , HH use separable power exponential correlation functions as in (5) and the mean is modeled as a constant (i.e.,  $h(\mathbf{z}) = 1$ ). HH implement this particular LMC in what may be perceived as a full Bayesian perspective, being the choice of priors a crucial aspect of their approach. To facilitate this aspect of the analysis, HH define priors which are strongly driven by preliminary data-based transformations of the model-run data. To be more specific, the mean  $\boldsymbol{\eta}$  is fixed at the sample mean  $\bar{\mathbf{y}}^M$  and, more importantly, the matrix  $\mathbf{A}$  is obtained from the singular value decomposition  $\mathbf{y}_D^M - \bar{\mathbf{y}}^M = \mathbf{P}_1\mathbf{D}_1\mathbf{Q}'$ , defined as being formed by the  $r$  columns of  $\mathbf{P}_1\mathbf{D}_1/\sqrt{N}$  which correspond to the  $r$  largest eigenvalues of  $\mathbf{D}_1$ . This particular choice of  $\mathbf{A}$ , based on the observed data, justifies taking priors for the variances  $1/\lambda_i^M$  centered at one, which in the words of the authors “helps stabilize the resulting posterior distribution for the correlation parameters” (correlation parameters are a one-to-one function of the  $\beta_{ij}^M$ ).

### 4.3. Our proposal

A central idea in our development is the combination of univariate emulators into a joint multivariate emulator in order to take advantage of the well-known univariate methodology, e.g. strategies and software to compute maximum likelihood estimates for univariate Gaussian processes. Examples of R packages which can be used to implement these univariate calculations include Roustant et al. (2011) and Dancik (2011). The main motivation is to construct a multivariate emulator which is easy to implement and, due to its simplicity, has the potential to scale-up to more complicated problems.

The model we propose is based on the LMC and, although its origins and motivation are different, it will in due course become apparent that it is strongly connected with the work of HH. For the type of problems we have in mind, high dimensionality is not an issue and hence our base model is (3) with  $r = p$ . The question of whether or not to include a nugget effect when constructing emulators for deterministic models is a topic of current research. Although traditionally small nuggets have been included in order to stabilize the numerical calculations, some authors advocate their use based on other considerations. For a recent contribution on this topic, see Gramacy and Lee (2011). Additionally, Pepelyshev (2010) shows that, contrary to what is often stated, the inclusion of the nugget may have an important impact on the estimated correlation structure. However, as will become apparent shortly, our approach is based on an argument that only applies if the nugget effect is not included in the mode, and hence, in what follows, we have  $\boldsymbol{\epsilon} = \mathbf{0}$ . Also, we assume a constant mean  $\boldsymbol{\mu}(\mathbf{z}) = \boldsymbol{\eta}$  (equivalently  $\mathbf{h}(\mathbf{z}) = 1$ ) and we use correlation functions of the type in (5) with no additional simplifications; in particular, we do not impose the restriction  $\alpha_{ij}^M \equiv 0$ . We consider this to be a distinguishing feature of our methodology when compared to the other LMC-based emulators described in this section.

In our implementation, the emulator is the predictive model (2) with the unknown parameters  $\mathbf{A}$ ,  $\boldsymbol{\eta}$ ,  $\boldsymbol{\alpha}^M$ ,  $\boldsymbol{\beta}^M$  and  $\boldsymbol{\lambda}^M$  replaced by estimates. What follows is a detailed description of how to proceed to obtain these estimates.

If the constant matrix  $\mathbf{A}$  is known, the model (3) can be equivalently written as  $\mathbf{y}^M(\mathbf{z}) = \mathbf{A}\mathbf{w}^a(\mathbf{z})$  where now the components in  $\mathbf{w}^a(\mathbf{z})$  are independent Gaussian processes with variance  $1/\lambda_i^M$ , correlation  $c_i^M(\cdot, \cdot)$  and with mean equal to the corresponding element in  $\boldsymbol{\eta}^a = \mathbf{A}^{-1}\boldsymbol{\eta}$ . Note that the no-nugget assumption is crucial here. We let  $\mathbf{w}_D^M = \mathbf{A}^{-1}\mathbf{y}_D^M$  denote the transformed model data. With this reparameterization, the maximum likelihood estimates  $(\hat{\lambda}_i^M, \hat{\eta}_i^a, \hat{\beta}_i^M, \hat{\alpha}_i^M)$  can easily be obtained using the available tools for univariate Gaussian processes (including specific software), since now the likelihood function factorizes into  $p$  separate factors. Obviously, the estimates for the original parameterization coincide with these except for the mean, that can be easily derived using  $\boldsymbol{\eta} = \mathbf{A}\boldsymbol{\eta}^a$ . At this point, what is left is the estimation of matrix  $\mathbf{A}$ . We propose to estimate  $\mathbf{A}$  as the orthogonal matrix  $\mathbf{P}$  appearing in the spectral decomposition (see e.g. Rencher, 1998) of the observed variance matrix of the columns in  $\mathbf{y}_D^M$ , that is, of the sample  $\mathbf{y}^M(\mathbf{z}_1), \dots, \mathbf{y}^M(\mathbf{z}_N)$ . The reasoning that justifies this choice is straightforward: with this matrix the pairwise observed

covariance matrix between  $\mathbf{w}^a(\mathbf{z}_i) = \mathbf{A}'\mathbf{y}^M(\mathbf{z}_i)$  and  $\mathbf{w}^a(\mathbf{z}_j) = \mathbf{A}'\mathbf{y}^M(\mathbf{z}_j)$  ( $i \neq j$ ) is diagonal. The fact that the resulting estimated matrix  $\mathbf{A}$ , which we denote by  $\widehat{\mathbf{A}}$ , is orthogonal has several positive implications. For instance, it can be easily shown that the total variance is conserved in the transformation, that is,  $\text{Tr} [\text{Var}(\mathbf{y}^M(\mathbf{z}))] = \text{Tr} [\text{Var}(\mathbf{w}^a(\mathbf{z}))]$ , where “Tr” stands for the trace of a matrix. This seems to be the simplest way of preventing potential identifiability problems which may appear since in the original parameterization (3) the variance of the process can be either introduced through the  $\lambda_i^M$  or through  $\mathbf{A}$ . For additional attractive consequences of the orthogonality of  $\widehat{\mathbf{A}}$ , see Fricker et al. (2010).

We are now in position to present our proposal for the emulator of a computer model producing multivariate output. With the estimates described above, it can be easily seen that the conditional predictive distribution (2) is

$$\mathbf{y}^M(\mathbf{z}) \mid \mathbf{y}_D^M, \widehat{\boldsymbol{\theta}}^M \sim N_p(\widehat{\mathbf{A}}\widehat{\mathbf{m}}(\mathbf{z}), \widehat{\mathbf{A}}\widehat{\mathbf{V}}(\mathbf{z})\widehat{\mathbf{A}}') \quad (6)$$

where  $\widehat{\mathbf{m}}(\mathbf{z}) = (\widehat{m}_i(\mathbf{z})', i = 1, \dots, p)'$ ,  $\widehat{\mathbf{V}}(\mathbf{z}) = \text{diag}(\widehat{V}_i(\mathbf{z}), i = 1, \dots, p)$ , and

$$\widehat{m}_i(\mathbf{z}) = \widehat{\eta}_i^a + \widehat{\gamma}_i(\mathbf{z})'(\widehat{\boldsymbol{\Gamma}}_i)^{-1}(\mathbf{w}_i^a - \widehat{\eta}_i^a \mathbf{1}), \quad \widehat{V}_i(\mathbf{z}) = \frac{1}{\lambda_i^M} - \widehat{\gamma}_i(\mathbf{z})'(\widehat{\boldsymbol{\Gamma}}_i)^{-1}\widehat{\gamma}_i(\mathbf{z}),$$

with  $\mathbf{1}$  being the vector of ones,  $\widehat{\boldsymbol{\Gamma}}_i$  being the estimated covariance matrix corresponding to the transformed model-run data  $\mathbf{w}_i^a = (w_i^a(\mathbf{z}_1), \dots, w_i^a(\mathbf{z}_N))'$ , and  $\widehat{\gamma}_i(\mathbf{z}) = (1/\lambda_i^M)(\widehat{c}_i^M(\mathbf{z}_1, \mathbf{z}), \dots, \widehat{c}_i^M(\mathbf{z}_k, \mathbf{z}))'$ , where  $\widehat{c}_i^M$  is the estimated correlation function. Recall that  $\widehat{\mathbf{A}}$  is obtained as follows: let  $\mathbf{PDP}'$  be the spectral decomposition of the variance matrix of  $\mathbf{y}_D^M$ ; then,  $\widehat{\mathbf{A}} = \mathbf{P}$ .

#### 4.4. Comparison with HH

Given the connection between our proposal and the work of HH we now stress the similarities and advantages of our approach.

The methods developed by HH were motivated by (and tailored to address) high-dimensional problems. For instance, HH’s prior specification for the parameters of the emulator puts considerable mass in areas of the parameter space that are associated with a small effect of the inputs in the output. That may be true for the large-dimensional problems that motivate their approach, but is arguably not true in general. Also, HH fix the roughness parameters of the power exponential correlation function at 2, whereas we estimate these from the data. Apart from these differences, it is clear that our proposed emulator and that in HH are based on connected arguments.

To understand the connection with the HH approach, first we have to consider their proposal without the reduction in dimensionality (i.e., when  $r = p$ ) and without the nugget effect. Recall that HH estimate  $\mathbf{A}$  as  $\mathbf{P}_1\mathbf{D}_1/\sqrt{N}$  where  $\mathbf{P}_1\mathbf{D}_1\mathbf{Q}'$  is the singular value decomposition of  $\mathbf{y}_D^M - \bar{\mathbf{y}}^M$ , which in this context coincides with the spectral decomposition (simply because  $r = p$ ). The key fact now is that the spectral decomposition of  $\mathbf{y}_D^M - \bar{\mathbf{y}}^M$  and that of the observed variance matrix of  $\mathbf{y}_D^M$  coincide, and hence our estimate of matrix  $\mathbf{A}$  coincides with  $\mathbf{P}_1$  in the HH setting (again

assuming no reduction in dimensionality). Loosely speaking, it could be said that we make use of the spectral decomposition to derive a matrix for which independency holds reasonably (as HH do) but without scaling by the observed variance (that is the difference between using  $\hat{\mathbf{A}} = \mathbf{P}$  in our approach or  $\hat{\mathbf{A}} = \mathbf{P}_1 \mathbf{D}_1 / \sqrt{N}$  in HH's approach). In our proposal, the marginal variances are estimated by the maximum likelihood estimates of the independent Gaussian processes.

Having said that, we do not expect the results to substantially differ in practice. For instance, for the example of Section 2, we have produced predictions of the first and second component of that computer model using our approach and using the natural adaptation of HH to this setting and the results essentially coincide.

However, our approach has the appealing of being computationally much simpler: our emulator is a closed-form MLE plug-in predictive density, not involving a full Bayesian analysis with MCMC techniques as in HH. Most importantly, as we have already stressed, our approach presents the important advantage of using software widely available for the univariate case and, additionally, does not require extensive transformations of the actual data.

#### 4.5. A synthetic example to compare emulators

As we have mentioned, we expect our emulator to produce essentially the same results as the emulator proposed in HH. Nevertheless, our emulator has the advantage of substantially reducing the computational burden since it comes given in closed form (a multivariate normal density) while the HH's emulator is defined as a posterior distribution with an involved expression (explorable only with simulation-based methods). This clearly supports our motivation of developing a methodology that has the potential to scale up to more complicated problems.

To illustrate these arguments, we have compared HH's emulator and ours in the context of a mathematical model recently proposed by Guerrero et al (2011) which describes the dynamics of subpopulations with respect to smoking habits in Spain. It is defined by a system of ordinary differential equations whose solution gives the proportion of non-smokers,  $y_1^M(z)$ , normal smokers,  $y_2^M(z)$ , and excessive smokers,  $y_3^M(z)$  in a specific year  $t$  given initial conditions at certain year  $t_0$ . For illustrative purposes, we use  $t_0 = 2003$  and produce predictions for year  $t = 2009$  and denote  $\mathbf{y}^M(z) = (y_1^M(z), y_2^M(z), y_3^M(z))$ . The response depends on a number of parameters that were accurately estimated based on population level data published by the Spanish Statistic Institute (e.g. like birth and death rate) and  $z$ , the transmission rate due to social pressure to adopt smoking habit. The calibration of  $z$  is more difficult as there is no direct data about  $z$ , and this was one of the main challenges this parameter was fitted by minimum least squared errors with some observed data.

We observed the model responses in the design set  $D^M = \{0.1, 0.3, 0.5, 0.7, 0.9\}$  and produce predictions for  $y^M(z^*)$  with the corresponding confidence bands at the unobserved points  $z^* = \{0.2, 0.8\}$ . The results are summarized in Table 2

	$z^* = 0.2$	$z^* = 0.8$
Our proposal		
$y_1^M(\mathbf{u}^*)$	0.5327(0.5327, 0.5327)	0.4780(0.4780,0.4780)
$y_2^M(\mathbf{u}^*)$	0.1563(0.1563, 0.1564)	0.1936(0.1936,0.1937)
$y_2^M(\mathbf{u}^*)$	0.1116(0.1115, 0.1116)	0.1221(0.1220,0.1221)
HH's emulator		
$y_1^M(\mathbf{u}^*)$	0.5327(0.5327, 0.5328)	0.4780(0.4780,0.4780)
$y_2^M(\mathbf{u}^*)$	0.1563(0.1563, 0.1564)	0.1936(0.1936,0.1937)
$y_2^M(\mathbf{u}^*)$	0.1116(0.1115, 0.1116)	0.1221(0.1220,0.1221)

Table 2: Predictions and 99% confidence bands based on our proposed emulator and HH's emulator.

## 5. Calibration of computer models with multivariate response

In this section, we take advantage of the multivariate emulator constructed in the previous section and construct a Bayesian model that will facilitate the calibration of computer models producing multivariate output.

### 5.1. The basic setting

In our approach to the calibration problem, we use (6) as the prior distribution for  $\mathbf{y}^M(\mathbf{u})$  in (1). This forms the basis of the proposed methodology which we detail below for a simple setting and which we later show to be easily extendable to slightly more complicated scenarios. In order to complete the specification of our statistical model, we need to postulate a prior on  $\mathbf{b}$ , on the covariance matrix  $\Sigma^F$ , and on the unknown input  $\mathbf{u}$ . It is natural and convenient to place a normal prior on  $\mathbf{b}$ . As the data provides information about  $\mathbf{b}$  only indirectly, we parameterize this prior parsimoniously:  $\mathbf{b} \mid \mathbf{T} \sim N_p(\mathbf{0}, \mathbf{T})$ , where  $\mathbf{T} = \text{diag}(\tau_i^2, i = 1, \dots, p)$ . We can think of the prior specification as in

$$\pi(\mathbf{y}^M(\mathbf{u}), \mathbf{b}, \mathbf{T}, \Sigma^F, \mathbf{u}) = \pi(\mathbf{y}^M(\mathbf{u}), \mathbf{b} \mid \mathbf{T}, \Sigma^F, \mathbf{u}) \pi(\mathbf{T}, \Sigma^F \mid \mathbf{u}) \pi(\mathbf{u}),$$

where the first factor on the right-hand side has already been specified; the last factor is typically specified using expert knowledge. We model  $(\mathbf{T}, \Sigma^F)$  as *a priori* independent of  $\mathbf{u}$ . Additionally, the prior  $\pi(\mathbf{T}, \Sigma^F) = \pi(\mathbf{T} \mid \Sigma^F) \pi(\Sigma^F)$  is constructed as follows: on  $\Sigma^F$ , we place the objective prior for a covariance matrix of a normal model that is recommended in Chang and Eaves (1990):

$$\pi(\Sigma^F) \propto |\Sigma^F|^{-\frac{p+1}{2}} |\mathbf{I} + \Sigma^F \circ (\Sigma^F)^{-1}|^{-1/2} \quad (7)$$

where the symbol “ $\circ$ ” stands for the Hadamard product of matrices. For the  $\tau_i^2$ , given the lack of direct information provided by the data about these parameters, we propose the use of a variation of a typical noninformative prior for the variance of a random effect:

$$\pi(\tau_i^2 \mid \sigma_i^2) \propto \frac{1}{\tau_i^2 + \sigma_i^2/K}, \quad (8)$$

where we are denoting by  $\sigma_i^2$  the elements of the main diagonal of  $\Sigma^F$ . Hence, this prior is appropriately scaled by  $\sigma_i^2/K$ , the variance of the average of the replicate observations. We will establish propriety of the ensuing posterior distribution, in a more general setting, in Theorem 5.1.

## 5.2. Hierarchical Scenarios

The setup of the previous section fits the example of Section 2. However, that setup can be naturally extended to accommodate more complex situations. Namely, it is straightforward to extend the methodology to the situation where the multivariate output is partitioned into groups that can be argued to be conditionally independent given the value of the common set of calibration inputs. That is, to the situation where we would perform independent analyses for each of the groups if it were not for the fact that these need to share the same value of the vector of calibration inputs.

A referee has pointed out that the recent paper by Higdon et al. (2010) describes an application concerning the calibration of cosmological parameters where this situation holds. The authors combine two types of observations (namely large structure data of the universe and measures of the cosmic microwave background) assuming conditional independence given a common set of calibrating parameters of a cosmological mathematical model. Unlike us, they do not include a bias term in the analysis.

Formally, we have  $\mathbf{y}^M(\mathbf{u}) = (\mathbf{y}_1^M(\mathbf{u}), \dots, \mathbf{y}_L^M(\mathbf{u}))'$  where each  $\mathbf{y}_\ell^M(\mathbf{u})$  is  $p$ -dimensional (the dimensions could actually differ, but that would only complicate the notation.) Similarly, the field data are denoted as

$$\mathbf{y}_k^F = (\mathbf{y}_{1k}^F(\mathbf{u}), \dots, \mathbf{y}_{Lk}^F(\mathbf{u}))', \quad k = 1, \dots, K$$

and are modeled independently as  $\mathbf{y}_{\ell k}^F(\mathbf{u}) \mid \mathbf{y}_\ell^M(\mathbf{u}), \mathbf{b}_\ell \sim N_{p_\ell}(\mathbf{y}_\ell^M(\mathbf{u}) + \mathbf{b}_\ell, \Sigma_\ell^F)$ ,  $k = 1, \dots, K$ ,  $\ell = 1, \dots, L$ . Extending the previous notation, the elements on the main diagonal of  $\Sigma_\ell^F$  are denoted  $\sigma_{i\ell}^2$ , for  $i = 1, \dots, p_\ell$ .

In terms of prior specification, everything follows through with the obvious modifications: essentially, adding a subscript  $\ell$  and imposing independence across the quantities associated with group  $\ell$ . Since the setting of the previous section is a particular case of the framework just described with  $L = 1$ , we state here the theorem that guarantees propriety of the resulting posterior:

**Theorem 5.1.** *The posterior distribution associated with our model is proper.*

Proof: See Appendix A.

Our application in Section 7 provides an illustration of the hierarchical scenario concerning the analysis of a vehicle suspension system over time.

## 6. Implementation strategy

Denote by  $\mathbf{D}$  the set of observed data, that is, the set of field and computer-model data that has been recorded. Also, denote by  $\mathbf{y}^M = (\mathbf{y}_\ell^M, \ell = 1, \dots, L)'$  the vector that consists of the output of

the computer model evaluated at the current value of  $\mathbf{u}$ . The field data can be summarized with the independent sufficient statistics  $\bar{\mathbf{y}}_\ell^F = \sum_{k=1}^K \mathbf{y}_{k\ell}^F / K$  and  $\mathbf{S}_\ell = \sum_{k=1}^K (\mathbf{y}_{k\ell}^F - \bar{\mathbf{y}}_\ell^F)(\mathbf{y}_{k\ell}^F - \bar{\mathbf{y}}_\ell^F)'$ , which have sampling densities given by  $\bar{\mathbf{y}}_\ell^F | \mathbf{y}_\ell^M, \mathbf{b}_\ell, \Sigma_\ell^F \sim N_p(\mathbf{y}_\ell^M + \mathbf{b}_\ell, \Sigma_\ell^F / K)$  and  $\mathbf{S}_\ell | \Sigma_\ell^F \sim W_p(\Sigma_\ell^F, K - 1)$ . We are denoting by  $W_p(\mathbf{B}, \nu)$  a  $p$ -dimensional Wishart distribution with mean  $\nu \mathbf{B}$ .

Let  $\boldsymbol{\tau}^2 = (\tau_1^2, \dots, \tau_L^2)'$ , where  $\tau_\ell^2 = (\tau_{1\ell}^2, \dots, \tau_{p\ell}^2)'$ , for  $\ell = 1, \dots, L$ ; similarly,  $\Sigma^F = (\Sigma_1^F, \dots, \Sigma_L^F)$ . Then, the posterior distribution of all unknowns can be decomposed in the form

$$\pi(\mathbf{y}^M, \mathbf{b}, \Sigma^F, \boldsymbol{\tau}^2, \mathbf{u} | \mathbf{D}) = \pi(\mathbf{y}^M | \mathbf{b}, \Sigma^F, \boldsymbol{\tau}^2, \mathbf{u}, \mathbf{D}) \pi(\mathbf{b} | \Sigma^F, \boldsymbol{\tau}^2, \mathbf{u}, \mathbf{D}) \pi(\Sigma^F, \boldsymbol{\tau}^2, \mathbf{u} | \mathbf{D}) \quad (9)$$

and is proportional to

$$\begin{aligned} & \prod_{\ell=1}^L N_p(\bar{\mathbf{y}}_\ell^F | \mathbf{y}_\ell^M + \mathbf{b}_\ell, \Sigma_\ell^F / K) W_p(\mathbf{S}_\ell | \Sigma_\ell^F, K - 1) \times \\ & \times \pi(\mathbf{T}_\ell | \Sigma_\ell^F) N_p(\mathbf{y}_\ell^M | \hat{\mathbf{A}}_\ell \hat{\mathbf{m}}_\ell(\mathbf{u}), \hat{\mathbf{A}}_\ell \hat{\mathbf{V}}_\ell(\mathbf{u}) \hat{\mathbf{A}}_\ell') N_p(\mathbf{b}_\ell | \mathbf{0}, \mathbf{T}_\ell) \pi(\Sigma_\ell^F) \pi(\mathbf{u}) . \end{aligned} \quad (10)$$

We are denoting by  $\pi(\mathbf{T}_\ell | \Sigma_\ell^F)$  the prior on  $\mathbf{T}_\ell$ , which is given by the product of (8) (with  $\tau_i^2$  replaced  $\tau_{i\ell}^2$ ) from  $i = 1$  to  $p$ . Also,  $\pi(\Sigma_\ell^F)$  is formally identical to (7). The first two distributions in (9) can be sampled directly:

$$\begin{aligned} \pi(\mathbf{y}^M | \mathbf{b}, \Sigma^F, \boldsymbol{\tau}^2, \mathbf{u}, \mathbf{D}) &= \prod_{\ell=1}^L N_p(\mathbf{y}_\ell^M | \mathbf{m}_\ell^1, \Sigma_\ell^1) , \\ \pi(\mathbf{b} | \Sigma^F, \boldsymbol{\tau}^2, \mathbf{u}, \mathbf{D}) &= \prod_{\ell=1}^L N_p(\mathbf{b}_\ell | \mathbf{m}_\ell^2, \Sigma_\ell^2) , \end{aligned}$$

where  $\Sigma_\ell^j$  and  $\mathbf{m}_\ell^j$ ,  $j = 1, 2$ , are specified in Appendix B.1. To sample from the last distribution in (9), one could in principle use a Gibbs sampling algorithm, iteratively sampling from the full-conditionals  $[\Sigma^F | \boldsymbol{\tau}^2, \mathbf{u}, \mathbf{D}]$ ,  $[\boldsymbol{\tau}^2 | \Sigma^F, \mathbf{u}, \mathbf{D}]$ , and  $[\mathbf{u} | \boldsymbol{\tau}^2, \Sigma^F, \mathbf{D}]$ . These would all be sampled from using a Metropolis-Hastings step.

What complicates matters, and makes the previous approach untenable in practice, is the fact that some components of our statistical model are potentially confounded. For instance, it is clearly possible to accommodate a large discrepancy between the computer model and reality by allowing a large enough field error variance. One possible solution is to estimate  $\Sigma^F$  using the replicate information only. This is dependent of having a reasonable number of replicates, which is the case in the application of Section 7. Replicates are very much recommended, but it's true that they will not be always available. To be more precise, the route we follow in order to sample from the last distribution in (9) is to take advantage of the decomposition

$$\pi(\Sigma^F, \boldsymbol{\tau}^2, \mathbf{u} | \mathbf{D}) = \pi(\Sigma^F | \mathbf{D}) \pi(\boldsymbol{\tau}^2, \mathbf{u} | \Sigma^F, \mathbf{D}) . \quad (11)$$

Sampling from the last distribution in (11) is dealt with by means of a Gibbs sampler: alternately sample from  $[\boldsymbol{\tau}^2 | \Sigma^F, \mathbf{u}, \mathbf{D}]$  and  $[\mathbf{u} | \boldsymbol{\tau}^2, \Sigma^F, \mathbf{D}]$ . Although these are not available in closed form, it

is relatively easy to find a suitable proposal for the implementation of a Metropolis-Hastings step, as we detail in Appendix B.2 and Appendix B.3. In the case of the full conditional of  $\mathbf{u}$ , the success of the proposal will depend on the prior on these quantities. In Appendix B.3 we report on the strategy that we implemented for the application of Section 7. The first distribution in (11) cannot be obtained in closed form; our approximation is to ignore the integral with respect to  $(\mathbf{u}, \boldsymbol{\tau}^2)$  and utilize only the replicate information to determine the posterior of the  $\boldsymbol{\Sigma}_\ell^F$ ; that is, we sample from

$$\tilde{\pi}(\boldsymbol{\Sigma}^F | \mathbf{D}) \propto \prod_{\ell=1}^L W_p(\mathbf{S}_\ell | \boldsymbol{\Sigma}_\ell^F, K - 1) \pi(\boldsymbol{\Sigma}_\ell^F). \quad (12)$$

We do so by means of a Metropolis-Hastings step, which is described in Appendix B.4. This strategy to address the issue of confounding between measurement error and computer model bias was already utilized in Bayarri et al. (2007a) and a full discussion of these ideas can be found in Liu et al. (2009).

## 7. An illustrative application

The problem that originally motivated this research is about predicting loads resulting from stressful events on a vehicle suspension system over time. A prototype vehicle is driven along a road that has two potholes, and the load history is recorded by sensors located at two different sites on this vehicle, hereafter referred to as sites 1 and 2. There are  $K = 7$  replicate field data experiments; the computer model was evaluated at a 65-point design set. More details about the data and some added insight about the application itself can be found in Bayarri et al. (2007a). In this application, the vector of calibration parameters is 4-dimensional, where  $u_3$  and  $u_4$  are associated with controllable inputs whose value for the vehicle tested in the field is known only up to manufacturing variability. From a methodological perspective they pose no challenge as they can simply be perceived as inputs to the computer model which must be estimated, as any calibration parameter. In Figure 4 we have plotted the first replicate of the field experiment for Site 1 (two top panels) and Site 2 (two bottom panels) and the two main regions of interest, named regions 1 and 2. These two regions of the load history correspond to the time at which the suspension system hits the two different potholes. We do not specify a numerical scale on the  $y$ -axis for proprietary reasons, but the scale is the same across all plots. Of particular interest to the engineers is the (maximum) range of the load history in regions 1 and 2 for both sites, i.e. the distance between the dashed lines in Figure 4. This is important for design purposes because, for safety reasons, the suspension system must be able to accommodate the largest load amplitude. In this paper, we will restrict attention exclusively to that function of the output of the computer model. This problem, as stated, fits exactly the modeling situation that we have described in Section 5.2: denote by  $\mathbf{y}_\ell^M(\mathbf{u}) = (y_{i\ell}^M(\mathbf{u}), i = 1, 2)'$  the computer model-calculated maximum load range for sites  $i = 1, 2$  at region  $\ell = 1, 2$ . We argue that the ranges at different regions may be treated as independent



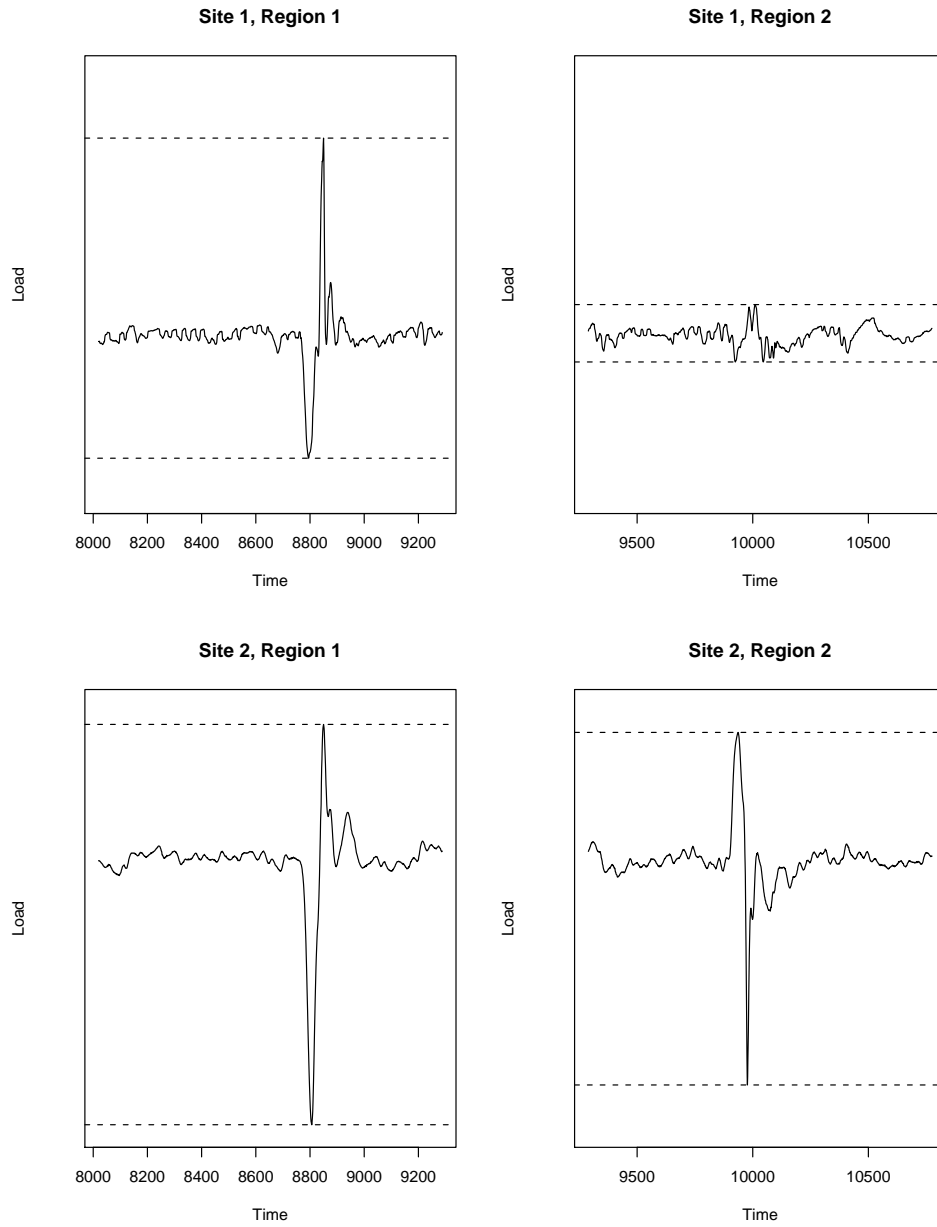


Figure 4: Sections of the field response (first replicate): sites 1 and 2 correspond to two different locations on the suspension system; regions 1 and 2 correspond to the suspension system hitting each of the two potholes.

(given  $\mathbf{u}$ ), but that within the same region these two quantities are obviously related, because they are the response to the same impact load but measured at two different sites on the vehicle.

### 7.1. Results

The prior on  $\mathbf{u}$  that we utilized was elicited with the help of experts and is such that the components are *a priori* independent with

$$\begin{aligned}\pi(u_1) &= \pi(u_2) = \text{Uni}(0.125, 0.875) \\ \pi(u_3) &= \text{N}(0.5, 2.4 \times 10^{-3}) \text{ restricted to } (0.3259, 0.6471) \\ \pi(u_4) &= \text{N}(0.5, 1.4 \times 10^{-2}) \text{ restricted to } (0.1471, 0.8529) .\end{aligned}$$

Estimates of the marginal posterior distribution of the uncertain parameters are depicted in Figure 5, where the dashed and dotted lines correspond to the posteriors obtained by analyzing separately the data corresponding to sites 1 and 2, respectively, using the methodology in Bayarri et al. (2007b). Namely, within each site we are treating the data obtained at each of the two regions as two independent analysis given the same calibration parameter. The solid lines correspond to the posteriors obtained using the combined analysis proposed in this paper; the shaded areas represent the prior distributions.

Clearly, things are not as transparent here as in the example of Section 2, as this is a real application. However, it is clear that the posterior distribution of the parameters  $u_3$  and  $u_4$  does not change much depending on the analysis. This is reassuring, as these are the parameters which represent manufacturing variability. Regarding the actual calibration parameters  $u_1$  and  $u_2$ , the results are more interesting. The posteriors of  $u_1$  do not seem to have many common features under the two analyses that use only site 1 or site 2 data. This poses the problem of which calibration strategy to follow. However, the posterior under the combined analysis seems to establish a compromise between the two differing posteriors, resulting in a single calibration strategy for the engineer. The situation with  $u_2$  is perhaps not as clear, but again we end up with a single posterior using our approach, which results from combining all the available information. Figure 6 depicts the resulting bias-corrected prediction of reality using our combined strategy (lower line of each panel) and the one that results from analyzing each site separately (upper line.) The dashed lines represent the means of the field data. What stands out is the considerable reduction in uncertainty under the analysis we propose, in line with what we saw in the context of the example of Section 2. We take it as evidence that our approach is appropriately combining information.

## 8. Summary and possible extensions

We have introduced statistical methodology to deal with calibration of computer models producing multivariate output. A key component in our proposal is the development of a non-separable multivariate emulator based on the linear model of coregionalization. Unlike other related proposals in the literature, our emulator can be easily implemented and, since it comes expressed in a closed-form, produces very fast responses.

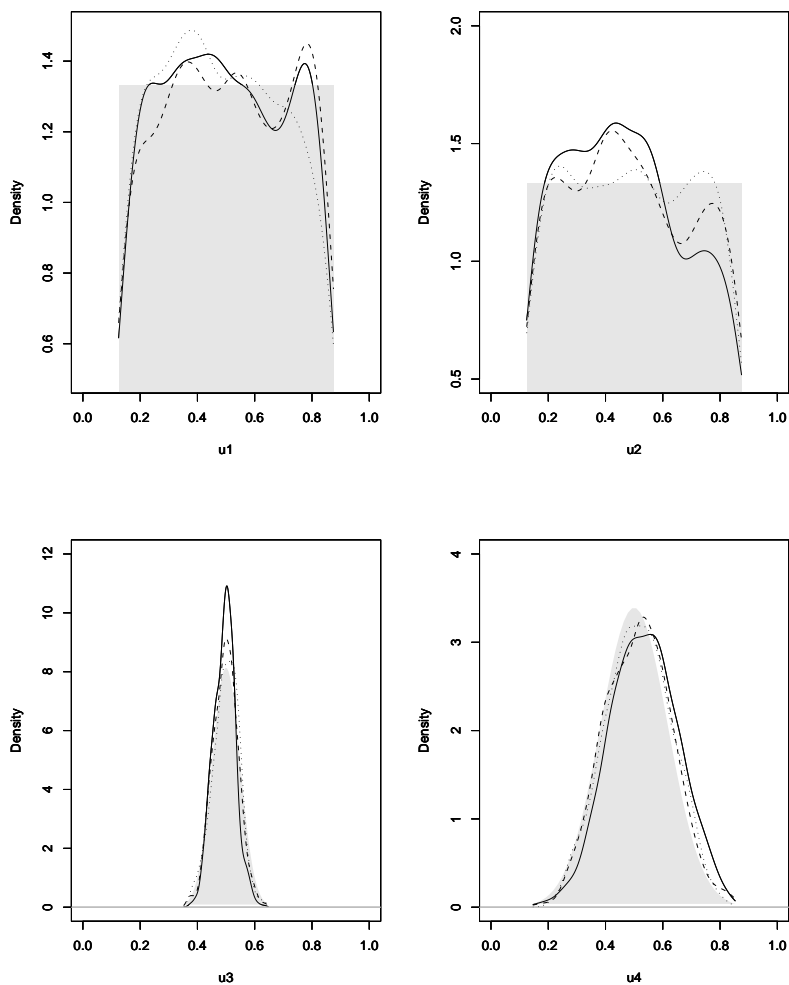


Figure 5: Marginal posterior distributions of the components of  $\mathbf{u}$ ; the solid lines correspond to the combined analysis whereas the dashed and dotted lines are obtained analyzing separately the data for site 1 and 2, respectively. The shaded areas represent the prior distributions.

With respect to the calibration strategy presented, our method is developed for the situation where only one value of the controllable input is exercised in the field experiments. This problem has been usually solved on highly problem-specific basis (see e.g. Higdon et al. (2008)) while our solution is more general and is motivated by the literature on Objective Bayes methodology and hence require no additional input from the user.

The reason why we restrict attention to a single value of the controllable input in the field experiments is simple: in the more general situation we are again confronted with specifying a multivariate Gaussian Process (GP), in this case for the bias term. To be more precise, we would

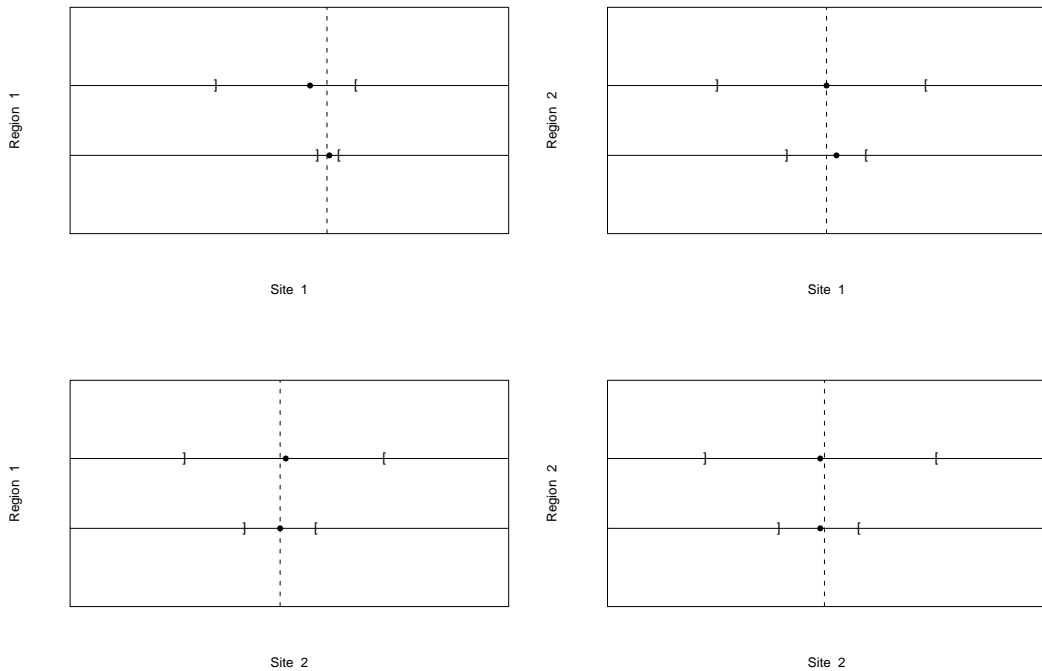


Figure 6: Bias corrected predictions of the range and associated 95% credible intervals under the combined analysis (lower line) and using only site-specific data (upper line.) The dashed vertical lines correspond to the mean of the field data.

need to be able to handle a model such as

$$\mathbf{y}^F(\mathbf{x}) = \mathbf{y}^M(\mathbf{x}, \mathbf{u}) + \mathbf{b}(\mathbf{x}) + \boldsymbol{\epsilon}$$

where  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^F)$ . Note that although we have not actually implemented it, the situation where  $\mathbf{b} \equiv \mathbf{0}$  is in principle straightforward to handle. Both Bhat et al. (2010) and Habib et al. (2007) implement this strategy. The interesting problem to consider is when  $\mathbf{b}(\mathbf{x})$  is *a priori* modeled as a multivariate Gaussian process. One can certainly use a coregionalization model and write

$$\mathbf{b}(\mathbf{x}) = \mathbf{B} \mathbf{W}^b(\mathbf{x})$$

where  $\mathbf{B}$  is a  $p \times h$  matrix and  $\mathbf{W}^b(\mathbf{x})$  is a vector of  $h$  independent Gaussian processes. This is clearly a model difficult to handle in this full generality. Higdon et al. (2008) addresses it in the situation where  $\mathbf{B}$  is a *known* matrix which arises by fully specifying  $h$  basis functions for the problem at hand using expert information. Bhat et al. (2010) also implements this strategy but with  $\mathbf{B}$  being the identity matrix of order  $p$ , that is, they consider the components of  $\mathbf{b}$  to be independent. Our goal is to find a way of estimating  $\mathbf{B}$  from the data, perhaps using arguments similar to the ones we developed for the emulation problem. This is still work in progress.

## 9. Acknowledgments

The authors would like to thank two anonymous referees for comments which have considerably improved the paper.

## Appendix A. Proof of Theorem

Recall that the posterior distribution of all unknowns is proportional to (10) multiplied by  $\pi(\mathbf{u})$ . We will omit the factor  $\pi(\mathbf{u})$  until the end of the proof as it plays no role until then. Also, to simplify the notation, we write  $\mathbf{m}$ ,  $\mathbf{V}$ , and  $\mathbf{A}$  instead of  $\widehat{\mathbf{m}}$ ,  $\widehat{\mathbf{V}}$  and  $\widehat{\mathbf{A}}$ . We next show that the marginal likelihood is finite.

Pick an arbitrary set of indexes  $M \subset \{1, \dots, L\} \times \{1, \dots, p\}$ . It suffices to consider the situation in which if  $(\ell, i) \in M$ , then we integrate  $\tau_{i\ell}^2$  over  $(\sigma_{i\ell}^2/K, +\infty)$ ; if  $(\ell, i) \notin M$ , then we integrate  $\tau_{i\ell}^2$  over  $(0, \sigma_{i\ell}^2/K)$ . Recall that  $\mathbf{T}_\ell$  is diagonal, and proceed as follows: if  $(\ell, i) \in M$ , bound (up to a constant)  $N(b_{i\ell} \mid 0, \tau_{i\ell}^2)$  by  $(\tau_{i\ell}^2)^{-1/2}$  and then integrate out  $b_{i\ell}$  for all  $(\ell, i) \in M$ . For every  $\ell = 1, \dots, L$ , define  $\tilde{\mathbf{y}}_\ell^F$ ,  $\tilde{\mathbf{b}}_\ell$ ,  $\tilde{\mathbf{y}}_\ell^M$  and  $\tilde{\Sigma}_\ell^F$  as being equal to the respective quantities without the tilde but with component  $i$  missing if  $(\ell, i) \in M$  (meaning column and line  $i$  missing in the case of the covariance matrix). The resulting integrated likelihood is then, up to a constant and omitting  $\pi(\mathbf{u})$ , bounded by above by the product of

$$\prod_{\ell=1}^L \pi(\mathbf{T}_\ell \mid \Sigma_\ell^F) \left\{ N(\tilde{\mathbf{b}}_\ell \mid \tilde{\mathbf{y}}_\ell^F - \tilde{\mathbf{y}}_\ell^M, \tilde{\Sigma}_\ell^F/K) W_p(\mathbf{S}_\ell \mid \Sigma_{\ell j}^F, K-1) N_p(\mathbf{y}_\ell^M \mid \mathbf{A}_\ell \mathbf{m}_\ell(\mathbf{u}), \mathbf{A}_\ell \mathbf{V}_\ell(\mathbf{u}) \mathbf{A}_\ell') \pi(\Sigma_\ell^F) \right\}$$

and

$$\prod_{(\ell, i) \in M} (\tau_{i\ell}^2)^{-1/2} \prod_{(\ell, i) \notin M} N(b_{i\ell} \mid 0, \tau_{i\ell}^2).$$

Bound, up to a constant,  $N(\tilde{\mathbf{b}}_\ell \mid \tilde{\mathbf{y}}_\ell^F - \tilde{\mathbf{y}}_\ell^M, \tilde{\Sigma}_\ell^F/K)$  by  $|\tilde{\Sigma}_\ell^F|^{-1/2}$  and integrate out  $\mathbf{y}_\ell^M$ ,  $\ell = 1, \dots, L$ , and  $b_{i\ell}$ ,  $(\ell, i) \notin M$ , to be left with

$$\prod_{\ell=1}^L |\tilde{\Sigma}_\ell^F|^{-1/2} W_p(\mathbf{S}_\ell \mid \Sigma_\ell^F, K-1) \pi(\Sigma_\ell^F) \prod_{(\ell, i) \in M} (\tau_{i\ell}^2)^{-1/2} \prod_{\ell=1}^L \prod_{i=1}^p \frac{1}{\tau_{i\ell}^2 + \sigma_{i\ell}^2/K}$$

In the formula above,  $\tau_{i\ell}^2$  is integrated over  $(\sigma_{i\ell}^2/K, +\infty)$  if  $(\ell, i) \in M$ , and over  $(0, \sigma_{i\ell}^2/K)$  otherwise. Using Lemma 1 below, we conclude that the resulting integrated likelihood is bounded by above, up to a constant, by

$$\prod_{\ell=1}^L |\tilde{\Sigma}_\ell^F|^{-1/2} W_p(\mathbf{S}_\ell \mid \Sigma_\ell^F, K-1) \pi(\Sigma_\ell^F) \times \prod_{(\ell, i) \in M} (\sigma_{i\ell}^2)^{-1/2}. \quad (\text{A.1})$$

Note that the Hadamard product of positive definite matrices is positive definite, and that  $|\mathbf{I} + \mathbf{D}| \geq 1$  for all  $\mathbf{D}$  positive definite, so that it's clear that  $\pi(\boldsymbol{\Sigma}_{\ell j}^F) \leq |\boldsymbol{\Sigma}_{\ell j}^F|^{-(p+1)/2}$ . We can rearrange (A.1) to obtain that it's bounded by above, up to a constant, by

$$\prod_{\ell=1}^L \prod_{i:(\ell,i) \in M} (\sigma_{i\ell}^2)^{-1/2} |\tilde{\boldsymbol{\Sigma}}_{\ell}^F|^{-1/2} |\boldsymbol{\Sigma}_{\ell}^F|^{-(p+1)/2} W_p(\mathbf{S}_{\ell} | \boldsymbol{\Sigma}_{\ell}^F, K-1).$$

The quantity  $\prod_{i:(\ell,i) \in M} \sigma_{i\ell}^2$  is the product of the diagonal elements of the lines/columns that have been removed from  $\boldsymbol{\Sigma}_{\ell}^F$  to produce  $\tilde{\boldsymbol{\Sigma}}_{\ell}^F$ . For that reason, Theorem 13.3.8 of Harville (1997) can be used to prove that

$$|\boldsymbol{\Sigma}_{\ell}^F| \leq |\tilde{\boldsymbol{\Sigma}}_{\ell}^F| \prod_{i:(\ell,i) \in M} \sigma_{i\ell}^2$$

so that, up to a constant, we are left with

$$\prod_{\ell=1}^L |\boldsymbol{\Sigma}_{\ell}^F|^{-(p+2)/2} W_p(\mathbf{S}_{\ell} | \boldsymbol{\Sigma}_{\ell}^F, K-1) \propto \text{IW}_p(\boldsymbol{\Sigma}_{\ell}^F | \mathbf{S}_{\ell}, K)$$

where we are denoting by  $\text{IW}_p(\mathbf{B}, \nu)$  the distribution of the inverse of a  $W_p(\mathbf{B}, \nu)$ -distributed random matrix. Since  $(\mathbf{u})$  is assigned a proper prior, the proof is complete, except for the following lemma:

**Lemma 1.** *For  $n$  positive integer and  $a > 0$ , there are  $c_1$  and  $c_2$ , positive constants, such that*

$$\int_a^{+\infty} x^{-n/2} \frac{1}{x+a} dx \leq c_1 a^{-n/2}$$

$$\int_0^a \frac{1}{x+a} dx = c_2$$

Proof: The second statement is trivial. The first will be shown to be true separately for  $n = 1$ ,  $n = 2$  and  $n > 2$ . If  $n = 1$ , we use the fact that  $\int x^{-1/2}/(x+a) dx = 2a^{-1/2} \arctan(ax)^{1/2}$  (e.g. Abramowitz and Stegun (1972), formula 3.3.29). If  $n = 2$ , we use the fact that, in the present circumstances,  $\int [x(x+a)]^{-1} dx = -a^{-1} \ln[(x+a)/x]$  (e.g. Abramowitz and Stegun (1972), formula 3.3.20). Finally, if  $n > 2$ , we start by noting that

$$\int_a^{+\infty} \frac{1}{x^{n/2}} \frac{1}{x+a} dx < \frac{1}{2a} \int_a^{+\infty} \frac{1}{x^{n/2}} dx$$

and the result follows easily.

## Appendix B. Additional details of the sampling algorithm

To simplify the notation, we write  $\mathbf{m}$ ,  $\mathbf{V}$ , and  $\mathbf{A}$  instead of  $\widehat{\mathbf{m}}$ ,  $\widehat{\mathbf{V}}$  and  $\widehat{\mathbf{A}}$ .

Appendix B.1. Means and variances of the full conditionals of  $\mathbf{y}^M$  and  $\mathbf{b}$

We have

$$\Sigma_\ell^1 = \left\{ \left( \frac{\Sigma_\ell^F}{K} \right)^{-1} + [\mathbf{A}_\ell \mathbf{V}_\ell(\mathbf{u}) \mathbf{A}_\ell']^{-1} \right\}^{-1}$$

and

$$\mathbf{m}_\ell^1 = \Sigma_\ell^1 \left\{ (\Sigma_\ell^F/K)^{-1} (\bar{\mathbf{y}}_\ell^F - \mathbf{b}_\ell) + [\mathbf{A}_\ell \mathbf{V}_\ell(\mathbf{u}) \mathbf{A}_\ell']^{-1} \mathbf{A}_\ell \mathbf{m}_\ell(\mathbf{u}) \right\}.$$

Also,

$$\Sigma_\ell^2 = \left\{ (\mathbf{T}_\ell)^{-1} + [\Sigma_\ell^F/K + \mathbf{A}_\ell \mathbf{V}_\ell(\mathbf{u}) \mathbf{A}_\ell']^{-1} \right\}^{-1}$$

and

$$\mathbf{m}_\ell^2 = \Sigma_\ell^2 [\Sigma_\ell^F/K + \mathbf{A}_\ell \mathbf{V}_\ell(\mathbf{u}) \mathbf{A}_\ell']^{-1} [\bar{\mathbf{y}}_\ell^F - \mathbf{A}_\ell \mathbf{m}_\ell(\mathbf{u})].$$

Appendix B.2. Sampling from  $\pi(\boldsymbol{\tau}^2 \mid \Sigma^F, \mathbf{u}, \mathbf{D})$

First notice that

$$\begin{aligned} \pi(\boldsymbol{\tau}^2 \mid \Sigma^F, \mathbf{u}, \mathbf{D}) &\propto & (B.1) \\ &\prod_{\ell=1}^L \text{N}_p(\bar{\mathbf{y}}_\ell^F \mid \mathbf{A}_\ell \mathbf{m}_\ell(\mathbf{u}), \Sigma_\ell^F/K + \mathbf{A}_\ell \mathbf{V}_\ell(\mathbf{u}) \mathbf{A}_\ell' + \mathbf{T}_\ell) \\ &\prod_{\ell=1}^L \prod_{i=1}^p \frac{1}{\tau_{i\ell}^2 + \sigma_{i\ell}^2/K}; \end{aligned}$$

so that, for all  $\ell = 1, \dots, L$ ,  $\tau_\ell^2$  is conditionally independent of  $(\tau_1^2, \dots, \tau_{\ell-1}^2, \tau_{\ell+1}^2, \tau_L^2)$ , with conditional density given by

$$\pi(\tau_\ell^2 \mid \Sigma^F, \mathbf{u}, \mathbf{D}) \propto \text{N}_p(\bar{\mathbf{y}}_\ell^F \mid \mathbf{A}_\ell \mathbf{m}_\ell(\mathbf{u}), \Sigma_\ell^F/K + \mathbf{A}_\ell \mathbf{V}_\ell(\mathbf{u}) \mathbf{A}_\ell' + \mathbf{T}_\ell) \prod_{i=1}^p \frac{1}{\tau_{i\ell}^2 + \sigma_{i\ell}^2/K}.$$

We sample each  $\tau_\ell^2$  independently using the the following Metropolis-Hastings scheme. Denoting by  $(\tau_\ell^2)^{\text{old}}$  the current value of  $\tau_\ell^2$ , we propose the next state  $(\tau_\ell^2)^{\text{prop}}$  by sampling from

$$g(\tau_\ell^2 \mid (\tau_\ell^2)^{\text{old}}) \propto \prod_{i=1}^p \frac{1}{\tau_{i\ell}^2} I_{(e^{-E(\tau_{i\ell}^2)^{\text{old}}}, e^{E(\tau_{i\ell}^2)^{\text{old}}})}(\tau_{i\ell}^2),$$

where  $E = 0.7$  and  $I_A(\cdot)$  stands for the indicator function of the set  $A$ . Next, compute

$$p = \frac{\pi((\tau_\ell^2)^{\text{prop}} \mid \Sigma^F, \mathbf{u}, \mathbf{D})}{\pi((\tau_\ell^2)^{\text{old}} \mid \Sigma^F, \mathbf{u}, \mathbf{D})} \frac{g((\tau_\ell^2)^{\text{prop}} \mid (\tau_\ell^2)^{\text{old}})}{g((\tau_\ell^2)^{\text{old}} \mid (\tau_\ell^2)^{\text{prop}})}.$$

Accept the proposal, that is let  $(\tau_\ell^2)^{\text{new}} = (\tau_\ell^2)^{\text{prop}}$ , with probability  $\min\{1, p\}$  and let  $(\tau_\ell^2)^{\text{new}} = (\tau_\ell^2)^{\text{old}}$  with probability  $1 - \min\{1, p\}$ .

*Appendix B.3. Sampling from  $\pi(\mathbf{u} \mid \boldsymbol{\tau}^2, \boldsymbol{\Sigma}^F, \mathbf{D})$*

We detail this sampling algorithm in the situation where the prior for  $(\mathbf{u})$  has the same structure as the one used in the application of Section 7. To be more precise, it is the product of uniform and normal distributions restricted to ranges given by the experts. In what follows, we denote by  $(a_i, b_i)$  the support of the prior for the  $i$ -th component of  $\mathbf{u}$ . In the application of Section 7,  $\mathbf{u}$  has 4 components.

In order to sample from the density

$$\pi(\mathbf{u} \mid \boldsymbol{\tau}^2, \boldsymbol{\Sigma}^F, \mathbf{D}) \propto \prod_{\ell=1}^L \text{N}_p(\bar{\mathbf{y}}_\ell^F \mid \mathbf{A}_\ell \mathbf{m}_\ell(\mathbf{u}), \boldsymbol{\Sigma}_\ell^F / K + \mathbf{A}_\ell \mathbf{V}_\ell(\mathbf{u}) \mathbf{A}_\ell' + \mathbf{T}_\ell) \pi(\mathbf{u})$$

we use a Metropolis-Hastings algorithm. Given the current value of the chain, denoted by  $\mathbf{u}^{\text{old}}$ , draw  $\mathbf{u}^{\text{prop}}$  from the mixture

$$g(\mathbf{u} \mid \mathbf{u}^{\text{old}}) = \prod_{i=1}^4 \left[ 0.5 \pi(u_i \mid a_i, b_i) + 0.5 \text{U} \left( u_i \mid \max\{a_i, u_i^{\text{old}} - 0.05\}, \min\{b_i, u_i^{\text{old}} + 0.05\} \right) \right]$$

where  $\text{U}(\cdot \mid a, b)$  represents the density of a uniform distribution in the interval  $(a, b)$  and  $\pi(u_i \mid a_i, b_i)$  denotes the prior density of  $u_i$ , which has support  $(a_i, b_i)$ . Compute

$$p = \frac{\pi(\mathbf{u}^{\text{prop}} \mid \boldsymbol{\tau}^2, \boldsymbol{\Sigma}^F, \mathbf{D}) g(\mathbf{u}^{\text{old}} \mid \mathbf{u}^{\text{prop}})}{\pi(\mathbf{u}^{\text{old}} \mid \boldsymbol{\tau}^2, \boldsymbol{\Sigma}^F, \mathbf{D}) g(\mathbf{u}^{\text{prop}} \mid \mathbf{u}^{\text{old}})}$$

Accept the proposed value, that is, set  $\mathbf{u}^{\text{new}} = \mathbf{u}^{\text{prop}}$  with probability  $\min\{1, p\}$  and let  $\mathbf{u}^{\text{new}} = \mathbf{u}^{\text{old}}$  with probability  $1 - \min\{1, p\}$ .

*Appendix B.4. Sampling from  $\tilde{\pi}(\boldsymbol{\Sigma}^F \mid \mathbf{D})$*

Note that, for  $\ell = 1, \dots, L$ , the  $\boldsymbol{\Sigma}_\ell^F$  are independent under the density in (12), with density

$$\tilde{\pi}(\boldsymbol{\Sigma}_\ell^F \mid \mathbf{D}) \propto \text{W}_p(\mathbf{S}_\ell \mid \boldsymbol{\Sigma}_\ell^F, K - 1) \pi(\boldsymbol{\Sigma}_\ell^F).$$

It can easily be shown that, with the prior (7),

$$\tilde{\pi}(\boldsymbol{\Sigma}_\ell^F \mid \mathbf{D}) \propto \text{IW}_p(\boldsymbol{\Sigma}_\ell^F \mid \mathbf{S}_\ell, K - 1) |\mathbf{I} + \boldsymbol{\Sigma}_\ell^F \circ (\boldsymbol{\Sigma}_\ell^F)^{-1}|^{-(p+1)/2}.$$

We have had good results sampling from the density above using the following Metropolis-Hastings step. Assume  $\boldsymbol{\Sigma}_\ell^{\text{old}}$  is the current value and sample  $(\boldsymbol{\Sigma}_\ell^F)^{\text{prop}}$  from the density  $\text{IW}_p(\boldsymbol{\Sigma}_\ell^F \mid \mathbf{S}_\ell, K - 1)$ . Accept the sampled value as the next state of the chain with probability

$$\min \left\{ 1, \frac{|\mathbf{I} + (\boldsymbol{\Sigma}_\ell^F)^{\text{prop}} \circ ((\boldsymbol{\Sigma}_\ell^F)^{\text{prop}})^{-1}|^{-(p+1)/2}}{|\mathbf{I} + (\boldsymbol{\Sigma}_\ell^F)^{\text{old}} \circ ((\boldsymbol{\Sigma}_\ell^F)^{\text{old}})^{-1}|^{-(p+1)/2}} \right\}$$

and stay at  $\boldsymbol{\Sigma}_\ell^{\text{old}}$  otherwise.



- Abramowitz, M. and Stegun, I. (1972), *Handbook of Mathematical Functions*, New York: Dover Publications, Inc.
- Bayarri, M., Berger, J., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R., Paulo, R., Sacks, J., and Walsh, D. (2007a), “Computer Model Validation with Functional Output,” *Annals of Statistics*, 35, 1874–1906.
- Bayarri, M. J., Berger, J. O., Kennedy, M. C., Kottas, A., Paulo, R., Sacks, J., Cafeo, J. A., Lin, C. H., and Tu, J. (2009), “Predicting vehicle crashworthiness: validation of computer models for functional and hierarchical data,” *Journal of the American Statistical Association*, to appear.
- Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C. H., and Tu, J. (2007b), “A Framework for Validation of Computer Models,” *Technometrics*, 49, 138–154.
- Bhat, K., Haran, M., and Goes, M. (2010), “Computer Model Calibration with Multivariate Spatial Output: A Case Study,” in *Frontiers of Statistical Decision Making and Bayesian Analysis*, eds. Chen, M.-H., Muller, P., Sun, D., Ye, K., and Dey, D., New York: Springer-Verlag, pp. 168–184.
- Chang, T. and Eaves, D. (1990), “Reference priors for the orbit in a group model,” *Annals of Statistics*, 18, 1595–1614.
- Conti, S. and O’Hagan, A. (2010), “Bayesian emulation of complex multi-output and dynamic computer models,” *Journal of Statistical Planning and Inference*, 140, 640–651.
- Craig, P., Goldstein, M., Seheult, A., and Smith, J. (1996), “Bayes linear strategies for history matching of hydrocarbon reservoirs,” in *Bayesian Statistics 5*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., Heckerman, D., and Smith, A. F. M., Oxford University Press: London, (with discussion).
- Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991), “Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments,” *Journal of the American Statistical Association*, 86, 953–963.
- Dancik, G. M. (2011), *mlepp: Maximum Likelihood Estimates of Gaussian Processes*, R package version 3.1.2.
- Fricker, T., Oakley, J., and Urban, N. (2010), “Multivariate emulators with nonseparable covariance structures,” *Working paper MUCM*.
- Gelfand, A., Schmidt, A., Banerjee, S., and Sirmans, C. (2004), “Nonstationary Multivariate Gaussian process modeling through spatially varying coregionalization (with discussion),” *Test*, 13, 263–312.

- Goldstein, M. (2010), “External Bayesian analysis of computer simulators,” in *Bayesian Statistics 9*, eds. Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., Oxford University Press: London, (with discussion).
- Gramacy, R. and Lee, H. (2008a), “Bayesian treed Gaussian process models with an application to computer modeling,” *Journal of the American Statistical Association*, 103, 1119–1130.
- (2008b), “Gaussian processes and limiting linear models,” *Computational Statistics and Data Analysis*, 53, 123–136.
- (2011), “Cases for the nugget in modeling computer experiments,” *Statistics and Computing*, to appear.
- Habib, S., Heitmann, K., Higdon, D., Nakhleh, C., and Williams, B. (2007), “Cosmic calibration: constraints from the matter power spectrum and the cosmic microwave background,” *Physical Review D*, 76.
- Harville, D. A. (1997), *Matrix Algebra from a Statistician’s Perspective*, Springer-Verlag Inc.
- Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008), “Computer Model Calibration Using High-Dimensional Output,” *Journal of the American Statistical Association*, 103, 570–583.
- Higdon, D., Heitmann, K., Nakhleh, C., and Habib, S. (2010), “Combining simulations and physical observations to estimate cosmological parameters,” in *The Oxford Handbook of Applied Bayesian Analysis*, eds. O’Hagan, A. and West, M., Oxford University Press.
- Higdon, D., Kennedy, M. C., Cavendish, J., Cafoe, J., and Ryne, R. D. (2004), “Combining field data and computer simulations for calibration and prediction,” *SIAM Journal on Scientific Computing*, 26, 448–466.
- Joseph, V. (2006), “Limit kriging,” *Technometrics*, 48, 458–466.
- Kennedy, M. C. and O’Hagan, A. (2001), “Bayesian calibration of computer models (with discussion),” *Journal of the Royal Statistical Society B*, 63, 425–464.
- Kennedy, M. C., O’Hagan, A., and Higgins, N. (2002), “Bayesian analysis of computer code outputs,” in *Quantitative Methods for Current Environmental Issues.*, eds. Anderson, C. W., Barnett, V., Chatwin, P. C., and El-Shaarawi, A. H., Springer-Verlag: London, pp. 227–243.
- Liu, F., Bayarri, M., and Berger, J. (2009), “Modularization in Bayesian analysis, with an emphasis on analysis of computer models,” *Bayesian Analysis*, 4, 119–150.
- Matheron, G. (1982), “Pour une analyse krigéante des données régionalisées,” Tech. rep., Ecole Nationale Supérieure des Mines de Paris.

- McFarland, J., Mahadevan, S., Romero, V., and Swiler, L. (2008), “Calibration and Uncertainty Analysis for Computer Simulations with Multivariate Output,” *AIAA Journal*, 46, 1253–1265.
- O’Hagan, A. (1978), “Curve Fitting and Optimal Design for Prediction (with discussion),” *Journal of the Royal Statistical Society B*, 40 (1), 1–42.
- Pepelyshev, A. (2010), “The role of the nugget term in the Gaussian process method,” in *mODa 9 Advances in Model-Oriented Design and Analysis. Contributions to Statistics*, eds. Giovagnoli, A., Atkinson, A. C., Torsney, B., and May, C., Physica-Verlag HD, pp. 149–156.
- Reichert, P., White, G., Bayarri, M., and Pitman, E. (2011), “Mechanism-based emulation of dynamic simulation models: concept and application in hydrology,” *Computational Statistics and Data Analysis*, 55, 1638–1655.
- Rencher, A. E. (1998), *Multivariate Statistical Inference and Applications*, Series in Probability and Statistics, Wiley.
- Rougier, J. (2008), “Efficient Emulators for Multivariate Deterministic Functions,” *Journal of Computational and Graphical Statistics*, 17(4), 827–843.
- Roustant, O., Ginsbourger, D., and Deville, Y. (2011), *DiceKriging: Kriging methods for computer experiments*, R package version 1.3.2.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), “Design and analysis of computer experiments (C/R: p423-435),” *Statistical Science*, 4, 409–423.
- Santner, T., Williams, B., and Notz, W. (2003), *The Design and Analysis of Computer Experiments*, Springer-Verlag.
- Wilkinson, R. (2010), “Computational Methods for Large-Scale Inverse Problems and Quantification of Uncertainty,” in *Large-Scale Inverse Problems and Quantification of Uncertainty*, ed. at al., L. B., John Wiley & Sons, New York.