

Calibration and prediction of two nested computer codes

Sophie Marque-Pucheu, Guillaume Perrin, Josselin Garnier

▶ **To cite this version:**

Sophie Marque-Pucheu, Guillaume Perrin, Josselin Garnier. Calibration and prediction of two nested computer codes. 2017. hal-01566968

HAL Id: hal-01566968
https://hal.science/hal-01566968

Preprint submitted on 21 Jul 2017

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

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

Calibration and prediction of two nested computer codes

Sophie Marque-Pucheu[1,2], Guillaume Perrin[1], and Josselin Garnier[3]

[1] CEA DAM DIF, F-91297, Arpajon, France
[2] Laboratoire de Probabilités et Modèles Aléatoires, Laboratoire Jacques-Louis Lions, Université Paris Diderot, 75205 Paris Cedex 13, France
[3] Centre de Mathematiques Appliquees, Ecole Polytechnique, 91128 Palaiseau Cedex, France

Abstract

Thanks to computing power increase, risk quantification relies more and more on computer modeling. Methods of risk quantification based on a fixed computational budget exist, but computer codes are almost always considered as a single black box.
In this paper, we are interested in analyzing the behavior of a complex phenomenon, which consists of two nested computer codes. By two nested computer codes, we mean that some inputs of the second code are outputs of the first code. Each code can be approximated by a parametrized computer model.
First we propose methods to calibrate the parameters of the computer models and build a predictor of the nested phenomenon for a given set of observations. The presented methods enable to take into account observations of the first code, the second code and the nested code.
Second the choice of the observations is studied. Methods of sequential designs, that means step by step addition of observation points to the observations' set, are examined. These sequential designs aim at improving the performance of calibration or prediction while optimizing the computational budget.
We show that the independent calibration of the 2 computer models is not efficient for predicting the nested phenomenon. We propose an original method that significantly improves the prediction's performance.

Keywords

Nested computer models, metamodeling, Bayesian calibration, Gaussian process, uncertainty quantification.

Email adress : sophie.marque01@etu.univ-paris-diderot.fr (S. Marque-Pucheu)

1 Introduction

Simulation has an increasing role for the conception, the optimization and the risk analysis of complex systems. Computer codes are thus introduced. These codes depend on the system inputs that are used to characterize the system we want to model (geometry, initial conditions, boundary conditions...). However these codes can be numerically expensive and are therefore replaced by a surrogate model.
In this paper the term phenomenon refers to the complex and perfect but expensive to evaluate computer code and the term computer model refers to the surrogate model of the computer code. The computer model is characterized by the system inputs and its parameters. For a computer model to be predictive, its parameters have to be calibrated from direct and/or indirect measurements. In this work, we assume that these computer models are deterministic, relatively quick to evaluate, but imperfect, in the sense that playing on the values of the input parameters only is not sufficient to make the predictions of the calibrated computer model exactly match the outputs of the computer code. In this case the computer model is too crude for its outputs to match exactly the outputs of the complex code after calibration. In this paper we mainly think at this configuration.

In order to take into account the model error and quantify the uncertainties associated with the parameters identification, a Bayesian formalism [15] is adopted in the following. The input parameters are therefore modeled by random variables, whereas the difference between the outputs of the calibrated computer model and of the complex code is modeled by a Gaussian process. This latter hypothesis is widely used in computer sciences ([16], [17], [14], [10], [11]), as it allows a very good trade-off between error control, complexity, and efficiency. Two central issues of this approach, also called Kriging, concern the choice of the statistical properties of the Gaussian process that is used, and the choice of a prior distribution for the input parameters ([6],[12],[5]). In this work, only non-informative prior distributions will be discussed for the input parameters ([1],[17]), and, for the sake of concision, we will assume that the Gaussian process is centered and that its covariance function is known. The interested reader may refer to [2] for further details on the covariance function identification in calibration purposes.
The posterior distributions of the input parameters and of the Gaussian processes are conditioned by the set of available observation points. The optimized choice of the observation points is therefore a way to improve the performance of calibration of the parameters and of prediction of the phenomenon. Concerning the case of a unique phenomenon several methods of Design of Experiments exist and include methods of space-filling designs and criterion-based designs ([16], [17], [4], [8]).

A lot of industrial issues involve multi-physics phenomena, however, the existing methods of Kriging and designs often consider a unique phenomenon or a multi-physics phenomenon as a unique phenomenon. In this paper we will focus on the case of two nested phenomena, that means that the studied complex phenomenon can be divided in two phenomena and the output of the first phenomenon is an input of the second phenomenon.

In this paper methods to calibrate the input parameters and build a predictor of a nested phenomenon are proposed. The choice of the observations is also studied and sequential designs specific for the case of a nested phenomenon are proposed. These designs enable to improve the performance of calibration and prediction by exploiting the nested structure of the studied

phenomenon.
The outline of this paper is the following. Section 2 presents the background in Kriging and sequential designs for a unique phenomenon and introduces the formalism used for the nested phenomenon. Section 3 presents the studied methods of calibration and prediction :

1. The "parallel" method corresponds to the case where the parameters and the predictors of the phenomena 1 and 2 are first calibrated and built separately, and then a predictor of the nested phenomenon is obtained by coupling the predictors of the two phenomena.

2. The linearized methods rely on the linearization of the nested phenomenon and enable to build Gaussian predictors for all the phenomena (1, 2 and nested). The posterior distribution of the parameters is also Gaussian.

Section 4 presents the studied sequential designs in case of a nested phenomenon. Existing methods are adapted to select new observation points on the phenomena 1 and 2 alone in order to improve the performance of calibration and prediction of the nested phenomenon while optimizing the computational budget. In section 5 the presented methods are applied to two numerical examples.

2 Background in Kriging and sequential designs for a unique phenomenon and formalism used for the nested phenomenon

2.1 Kriging of a unique phenomenon

In this part, we are interested in the modeling of a unique phenomenon. Let $\mathbf{x} \in \mathcal{D}_d$ be the input vector that characterizes this phenomenon, where $\mathcal{D}_d$ is assumed to be a compact subset of $\mathbb{R}^d, d \geq 1$, and let $\boldsymbol{y} \in \mathrm{L}^2\left(\mathcal{D}_d, \mathbb{R}^n\right), n \geq 1$ be the quantity of interest that is used to analyze the studied phenomenon, where $\mathrm{L}^2\left(\mathcal{D}_d, \mathbb{R}^n\right)$ is the space of the functions with values in $\mathbb{R}^n$ which are square integrable on $\mathcal{D}_d$. This quantity of interest is supposed to be a particular realization of a Gaussian process $\boldsymbol{Y}$, such that :

$$\boldsymbol{Y}(\mathbf{x}) = \boldsymbol{f}\left(\mathbf{x}; \boldsymbol{\beta}\right) + \epsilon\left(\mathbf{x}\right), \tag{1}$$

where :

- $\boldsymbol{f}\left(\mathbf{x}; \boldsymbol{\beta}\right) : \mathcal{D}_d \times \mathbb{R}^p \rightarrow \mathbb{R}^n$ is a deterministic function, which is continuous on $\mathcal{D}_d$,
- $\boldsymbol{\beta} \in \mathbb{R}^p$ is the parameters' vector,
- $\epsilon\left(\mathbf{x}\right) \sim \mathcal{GP}\left(\mathbf{0}, \boldsymbol{C}\left(\mathbf{x}, \mathbf{x}^{'}\right)\right)$, where $\mathcal{GP}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ denotes a Gaussian process of mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ and $\boldsymbol{C}\left(\mathbf{x}, \mathbf{x}^{'}\right)$ is a $(n \times n)$-dimensional matrix-valued covariance function, which is assumed continuous on $\mathcal{D}_d \times \mathcal{D}_d$.

We suppose that N observations of the studied phenomenon are available. We denote by $\mathcal{S}_N$ the σ-algebra generated by these observations, $\mathbb{X}^{(N)} = \left(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\right)$ the vector of the inputs of the observations, $\mathbb{Y}^{(N)} = \left(\boldsymbol{y}\left(\mathbf{x}^{(1)}\right), \ldots, \boldsymbol{y}\left(\mathbf{x}^{(N)}\right)\right)$ the vector of the outputs of the observations. In this paper we will consider the case where the function $\boldsymbol{f}\left(\mathbf{x}; \boldsymbol{\beta}\right)$ can be written (or approximated thanks to a linearization) as a linear function with respect to $\boldsymbol{\beta}$: $\boldsymbol{f}\left(\mathbf{x}; \boldsymbol{\beta}\right) = \boldsymbol{h}\left(\mathbf{x}\right)^t \boldsymbol{\beta}$. So the formalism is as follows :

$$\mathcal{L}\left(\boldsymbol{Y}(\mathbf{x}) \mid \boldsymbol{\beta}\right) = \mathcal{GP}\left(\boldsymbol{h}\left(\mathbf{x}\right)^t \boldsymbol{\beta}, \boldsymbol{C}\left(\mathbf{x}, \mathbf{x}\right)\right), \tag{2}$$

where

- $\boldsymbol{h}\left(\mathbf{x}\right) : \mathbb{R}^d \rightarrow M_{p,n}\left(\mathbb{R}\right)$ is a $(p \times n)$ -dimensional matrix of deterministic functions that are continuous on $\mathcal{D}_d$,
- $\mathcal{L}\left(X \mid Y\right)$ denotes the conditional distribution of X given Y.

From Eq. (2) it can be inferred that, in case of a non informative prior for the parameters, that is the prior pdf of the parameters is proportional to 1 on its definition space ([1], [6], [5]), the posterior distributions of the parameters and of the Gaussian process are Gaussian ([1]). In what follows, the term predictor of the phenomenon corresponds to the posterior distribution of the Gaussian process associated with the phenomenon. If we denote by $\mathcal{N}\left(\ \boldsymbol{\mu}\ ,\ \boldsymbol{\Sigma}\ \right)$ the Gaussian distribution of mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, the posterior distributions of the parameters and of the Gaussian process can be written as follows :

$$\begin{aligned} \mathcal{L}\left(\boldsymbol{\beta} \mid \mathcal{S}_N\right) &= \mathcal{N}\left(\ \boldsymbol{\mu}_{\boldsymbol{\beta}}^{(N)}\ ,\ \boldsymbol{V}_{\boldsymbol{\beta}}^{(N)}\ \right), \\ \mathcal{L}\left(\boldsymbol{Y}(\mathbf{x})\ \mid \mathcal{S}_N\right) &= \mathcal{GP}\left(\ \boldsymbol{\mu}^{(N)}\left(\mathbf{x}\right)\ ,\ \boldsymbol{C}^{(N)}\left(\mathbf{x},\mathbf{x}\right)\ \right). \end{aligned} \tag{3}$$

where :

$$\begin{aligned} \boldsymbol{V}_{\boldsymbol{\beta}}^{(N)} &= \left(\left(\boldsymbol{H}^{(N)}\right)^t \left(\boldsymbol{R}^{(N)}\right)^{-1} \boldsymbol{H}^{(N)}\right)^{-1}, \\ \boldsymbol{\mu}_{\boldsymbol{\beta}}^{(N)} &= \boldsymbol{V}_{\boldsymbol{\beta}}^{(N)} \left(\boldsymbol{H}^{(N)}\right)^t \left(\boldsymbol{R}^{(N)}\right)^{-1} \mathbb{Y}^{(N)}, \\ \boldsymbol{\mu}_{\boldsymbol{\beta}}^{(N)} &\in \mathbb{R}^p, \quad \boldsymbol{V}_{\boldsymbol{\beta}}^{(N)} \in M_{p,p}\left(\mathbb{R}\right), \end{aligned}$$

$$\boldsymbol{\mu}^{(N)}\left(\mathbf{x}\right) = \boldsymbol{h}\left(\mathbf{x}\right)^t \boldsymbol{\mu}_{\boldsymbol{\beta}}^{(N)} + \boldsymbol{C}\left(\mathbf{x}, \mathbb{X}^{(N)}\right)\left(\boldsymbol{R}^N\right)^{-1}\left(\mathbb{Y}^{(N)} - \boldsymbol{H}^N \boldsymbol{\mu}_{\boldsymbol{\beta}}^{(N)}\right),$$

$$\boldsymbol{C}^{(N)}\left(\mathbf{x},\mathbf{x}'\right) = \boldsymbol{C}\left(\mathbf{x},\mathbf{x}'\right) - \boldsymbol{C}\left(\mathbf{x},\mathbb{X}^{(N)}\right)\left(\boldsymbol{R}^{(N)}\right)^{-1} \boldsymbol{C}\left(\mathbb{X}^{(N)},\mathbf{x}'\right) + \boldsymbol{u}^{(N)}\left(\mathbf{x}\right)^t \boldsymbol{V}_{\boldsymbol{\beta}}^{(N)} \boldsymbol{u}^{(N)}\left(\mathbf{x}'\right),$$

$$\boldsymbol{\mu}^{(N)}\left(\mathbf{x}\right) \in \mathbb{R}^n, \quad \boldsymbol{C}^{(N)}\left(\mathbf{x},\mathbf{x}'\right) \in M_{n,n}\left(\mathbb{R}\right),$$

$$\begin{aligned} \boldsymbol{u}^{(N)}\left(\mathbf{x}\right) &= \boldsymbol{h}\left(\mathbf{x}\right) - \left(\boldsymbol{H}^{(N)}\right)^t \left(\boldsymbol{R}^{(N)}\right)^{-1} \boldsymbol{C}\left(\mathbb{X}^{(N)},\mathbf{x}\right), \\ \boldsymbol{H}^{(N)} &= \boldsymbol{h}\left(\mathbb{X}^{(N)}\right)^t, \quad \boldsymbol{R}^{(N)} = \boldsymbol{C}\left(\mathbb{X}^{(N)},\mathbb{X}^{(N)}\right), \\ \boldsymbol{H}^{(N)} &\in M_{Nn,p}\left(\mathbb{R}\right), \quad \boldsymbol{R}^{(N)} \in M_{Nn,Nn}\left(\mathbb{R}\right). \end{aligned} \tag{4}$$

The Gaussian process modeling provides therefore a formalism to solve two problems given these observations : the calibration of the computer model and the prediction of the phenomenon.
The calibration's problem can be summarized as follows :

- The function $f\left(\mathbf{x};\boldsymbol{\beta}\right)$ denotes an available computer model, which is parametrized by an unknown vector $\boldsymbol{\beta}$.
- The centered Gaussian process $\epsilon\left(\mathbf{x}\right)$ represents the model error of the computer model compared to the phenomenon.

- We suppose that there exists a unique and true value for $\boldsymbol{\beta}$, which is denoted by $\boldsymbol{\beta}^*$. The goal of the calibration is to estimate $\boldsymbol{\beta}^*$ and to add a measure of uncertainty to this estimation, which is achieved by calculating the posterior distribution of the parameters given the observations.

- The performance of the calibration can therefore be assessed by computing the following error :

$$\epsilon_\beta^2 = \frac{\mathbb{E}\left[\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|^2 \mid \mathcal{S}_N\right]}{\|\boldsymbol{\beta}^*\|^2}. \tag{5}$$

  The numerator of the fraction above can be divided in two parts :

$$\mathbb{E}\left[\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|^2 \mid \mathcal{S}_N\right] = \underbrace{\sum_{i=1}^{p} \left(\boldsymbol{\beta}^* - \boldsymbol{\mu}_{\boldsymbol{\beta}}^{(N)}\right)_i^2}_{\text{error on the parameters' mean}} + \underbrace{\sum_{i=1}^{p} \left(\boldsymbol{V}_{\boldsymbol{\beta}}^{(N)}\right)_{ii}}_{\text{variance of the parameters}} \tag{6}$$

  The mean of the posterior distribution of the parameters is the best deterministic estimation of $\boldsymbol{\beta}^*$ in the L^2 sense, and its variance characterizes the uncertainty of this estimation. The error of calibration integrates therefore the error on the mean and the uncertainty of the mean.

The prediction's problem can be summarized as follows :

- The function $f\left(\mathbf{x};\boldsymbol{\beta}\right)$ is seen as a model trend and is generally a priori chosen.

- The centered Gaussian process $\epsilon\left(\mathbf{x}\right)$ can be seen as the learning process of the phenomenon.

- The goal of the prediction is to predict the value of $\boldsymbol{y}\left(\mathbf{x}\right)$ in a non computed point $\mathbf{x}$ and to add a measure of uncertainty to this prediction, which is reached by calculating the posterior distribution of the Gaussian process $\boldsymbol{Y}(\mathbf{x})$ given the available information.

- The performance of the prediction can therefore be evaluated by the computation of this second error :

$$\boldsymbol{\epsilon}_{\boldsymbol{y}}^2 = \frac{\mathbb{E}\left[\|\boldsymbol{y} - \boldsymbol{Y}\|_{L^2}^2 \mid \mathcal{S}_N\right]}{\|\boldsymbol{y}\|_{L^2}^2}, \tag{7}$$

  where, for all $\boldsymbol{u}$ in $\mathrm{L}^2\left(\mathcal{D}_d, \mathbb{R}^n\right)$,$\|\boldsymbol{u}\|_{L^2}^2 = \int_{\mathcal{D}_d} \|\boldsymbol{u}\left(\mathbf{x}\right)\|^2 d\mathbf{x}$.

  The numerator of the fraction above can be divided in two parts :

$$\mathbb{E}\left[\|\boldsymbol{y} - \boldsymbol{Y}\|_{L^2}^2 \mid \mathcal{S}_N\right] = \underbrace{\int_{\mathcal{D}_d} \|\boldsymbol{y}\left(\mathbf{x}\right) - \boldsymbol{\mu}^{(N)}\left(\mathbf{x}\right)\|^2 d\mathbf{x}}_{\text{integrated square error on the predictor's mean}} + \underbrace{\int_{\mathcal{D}_d} \mathrm{tr}\left(\boldsymbol{C}^{(N)}\left(\mathbf{x},\mathbf{x}\right)\right) d\mathbf{x}}_{\text{integrated variance of the predictor}} \tag{8}$$

  where $\mathrm{tr}\left(\boldsymbol{M}\right)$ denotes the trace of the matrix $\boldsymbol{M}$.

  The mean of the posterior distribution of the process $\boldsymbol{Y}(\mathbf{x})$ is the best deterministic prediction of $\boldsymbol{y}\left(\mathbf{x}\right)$ and its variance allows us to characterize the uncertainty of this prediction. The error of prediction integrates therefore the error on the mean and the uncertainty of the mean.

2.2 Designs of Experiments in case of a unique phenomenon

The efficiency of prediction and calibration depends on the space filling properties of the set of the available observations of $\boldsymbol{y}$, which is generally called Design of Experiments.
Hence, several methods exist ([16], [17], [4], [8]) to choose the Design of Experiments used for the calibration or the prediction :

- methods based on the exploration of the inputs' space : space-filling Latin Hypercube Sampling, quasi-Monte-Carlo sampling...
- methods based on the minimization of a quantity associated with the variance of the calibrated parameters $\boldsymbol{V}_{\boldsymbol{\beta}}^{(N)}$,
- methods based on the minimization of a quantity associated with the variance of the predictor $\boldsymbol{C}^{(N)}(\mathbf{x}, \mathbf{x})$.

2.3 Formalism used for the nested phenomenon

In this paper, we focus on the case of two nested phenomena. Two quantities of interest, y_1 and y_2, are thus introduced to characterize these two phenomena, which are supposed to be two real-valued continuous functions on their respective definition domains $\mathcal{D}_{d_1}$ and $\mathcal{D}_{d_2} = \mathbb{R} \times \mathcal{D}_{d_{c2}}$. The sets $\mathcal{D}_{d_1}$ and $\mathcal{D}_{d_{c2}}$ are moreover supposed to be two compact subsets of $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2-1}$ respectively, where d_1 and d_2 are two positive integers. Given these two functions, the nested phenomenon is defined as follows :

$$y_c(\mathbf{x_c}) = y_2\left(\left(y_1(\mathbf{x_{c1}}), \mathbf{x_{c2}}\right)\right), \quad \mathbf{x_c} = (\mathbf{x_{c1}}, \mathbf{x_{c2}}) \in \mathcal{D}_{d_1} \times \mathcal{D}_{d_{c2}}. \tag{9}$$

In theory, the definition domains may be unbounded, but the reduction to compact sets enables the square integrability of y_c on $\mathcal{D}_{d_1} \times \mathcal{D}_{d_{c2}}$ and therefore the calculation of the performance criterion defined in Eq. (7).

In this work, we suppose that two parametrized computer models are available for the analysis of these two phenomena, which are assumed to be linear with respect to their parameters. The objective is to adapt the previously presented methods to calibrate the parameters of the two computer models and to build a predictor of the nested phenomenon. As previously, a Gaussian formalism is introduced, which consists in assuming that the deterministic functions y_1 and y_2 are the realizations of two Gaussian processes Y_1 and Y_2, such that:

$$\begin{aligned} Y_1(\mathbf{x_1}) &= \boldsymbol{h_1}(\mathbf{x_1})^t \boldsymbol{\beta_1} + \epsilon_1(\mathbf{x_1}), \\ Y_2(\mathbf{x_2}) &= \boldsymbol{h_2}(\mathbf{x_2})^t \boldsymbol{\beta_2} + \epsilon_2(\mathbf{x_2}), \end{aligned} \tag{10}$$

where :

- $\boldsymbol{\beta_1} \in \mathbb{R}^{p_1}$ and $\boldsymbol{\beta_2} \in \mathbb{R}^{p_2}$ are the parameters' vectors,
- p_1 and p_2 are positive integers that characterize the dimension of the parameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$,
- $\boldsymbol{h_1}(\mathbf{x_1}) \in \mathbb{R}^{p_1}$ and $\boldsymbol{h_2}(\mathbf{x_2}) \in \mathbb{R}^{p_2}$ are vectors of deterministic functions, which are supposed to be known, $\mathbf{x_1} \mapsto \boldsymbol{h_1}(\mathbf{x_1})$ is continuous on $\mathcal{D}_{d_1}$, $\mathbf{x_2} \mapsto \boldsymbol{h_2}(\mathbf{x_2})$ is continuous on $\mathcal{D}_{d_2}$ and $(\mathbf{x_2})_1 \mapsto \boldsymbol{h_2}(\mathbf{x_2})$ is continuously differentiable,

- ϵ_1 and ϵ_2 are real-valued centered Gaussian processes, defined by their covariance functions $C_1\left(\mathbf{x_1}, \mathbf{x_1'}\right)$ and $C_2\left(\mathbf{x_2}, \mathbf{x_2'}\right)$. C_1 is assumed to be Hölder continuous on $\mathcal{D}_{d_1} \times \mathcal{D}_{d_1}$, so that the realizations of the process Y_1 are Hölder continuous, and C_2 is assumed to be continuous on $\mathcal{D}_{d_2} \times \mathcal{D}_{d_2}$.

According to Eq. (9) and (10), the nested phenomenon $y_c\left(\mathbf{x_c}\right)$ can therefore be seen as a particular realization of the process $Y_c\left(\mathbf{x_c}\right)$, such that:

$$Y_c\left(\mathbf{x_c}\right) = Y_2\left(\left(Y_1\left(\mathbf{x_{c1}}\right), \mathbf{x_{c2}}\right)\right) \tag{11}$$

It can be inferred from Eq. (10) and (11) that there are three possible types of observations in case of a nested phenomenon :

1. Observations of the phenomenon 1. The σ-algebra generated by the observations of the inputs $\mathbb{X}_1^{(N_1)} = \left(\mathbf{x_1}^{(1)}, \ldots, \mathbf{x_1}^{(N_1)}\right)$ and the outputs $\mathbb{Y}_1^{(N_1)} = \left(y_1\left(\mathbf{x_1}^{(1)}\right), \ldots, y_1\left(\mathbf{x_1}^{(N_1)}\right)\right)$ will be denoted by $\mathcal{S}_{N_1}$,
2. Observations of the phenomenon 2. The σ-algebra generated by the observations of the inputs $\mathbb{X}_2^{(N_2)} = \left(\mathbf{x_2}^{(1)}, \ldots, \mathbf{x_2}^{(N_2)}\right)$ and the outputs $\mathbb{Y}_2^{(N_2)} = \left(y_2\left(\mathbf{x_2}^{(1)}\right), \ldots, y_2\left(\mathbf{x_2}^{(N_2)}\right)\right)$ will be denoted by $\mathcal{S}_{N_2}$,
3. Observations of the nested phenomenon. The σ-algebra generated by the observations of the inputs $\mathbb{X}_c^{(N_c)} = \left(\mathbf{x_c}^{(1)}, \ldots, \mathbf{x_c}^{(N_c)}\right)$ and the outputs $\mathbb{Y}_c^{(N_c)} = \left(y_c\left(\mathbf{x_c}^{(1)}\right), \ldots, y_c\left(\mathbf{x_c}^{(N_c)}\right)\right)$ will be denoted by $\mathcal{S}_{N_c}$,

3 Calibration and prediction of two nested phenomena

In this section, the objective is to calibrate the parameters of the two computer models and to build a predictor of a nested phenomenon for a given set of observation points, which means to determine the posterior distributions of the parameters and of the process modeling the nested phenomenon :

$$\begin{gathered} \mathcal{L}\left(\boldsymbol{\beta_c} \mid \mathcal{S}_{N_{gp}}\right), \\ \mathcal{L}\left(Y_c\left(\mathbf{x_c}\right) \mid \mathcal{S}_{N_{gp}}\right), \end{gathered} \tag{12}$$

where :

- $\mathcal{S}_{N_{gp}}$ denotes the available information, which can be defined as the σ-algebra generated by the observations of the phenomena 1 and 2 in the parallel method and by the observations of the phenomena 1, 2 and nested in the linearized methods,
- $\boldsymbol{\beta_c} = \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \end{bmatrix}$.

According to Eq. (10) and (11) the process associated with the nested phenomenon is given by :

$$Y_c(\mathbf{x_c}) = \boldsymbol{h_2}\left(\left(\boldsymbol{h_1}\left(\mathbf{x_{c1}}\right)^t \boldsymbol{\beta_1} + \epsilon_1\left(\mathbf{x_{c1}}\right), \mathbf{x_{c2}}\right)\right)^t \boldsymbol{\beta_2} + \epsilon_2\left(\left(\boldsymbol{h_1}\left(\mathbf{x_{c1}}\right)^t \boldsymbol{\beta_1} + \epsilon_1\left(\mathbf{x_{c1}}\right), \mathbf{x_{c2}}\right)\right). \tag{13}$$

It can be inferred from Eq. (13) that the distribution of the process $Y_c(\mathbf{x_c})$ given $\boldsymbol{\beta_c}$ is not Gaussian and its mean is not linear with respect to the parameters. So the posterior distributions of the parameters and of the process of Eq. (12) are not Gaussian. Their general formulation is, according to the Bayes rule and the law of total probability :

$$\begin{aligned} \mathcal{L}\left(\boldsymbol{\beta_c} \mid \mathcal{S}_{N_{gp}}\right) &\propto \mathcal{L}\left(\left(\mathbb{Y}_1^{(N_1)}, \mathbb{Y}_2^{(N_2)}, \mathbb{Y}_c^{(N_c)}\right) \mid \boldsymbol{\beta_c}\right) \mathcal{L}^{\text{prior}}\left(\boldsymbol{\beta_c}\right), \\ \mathcal{L}\left(Y_c(\mathbf{x_c}) \mid \mathcal{S}_{N_{gp}}\right) &= \int_{\boldsymbol{\beta_c}} \mathcal{L}\left(Y_c(\mathbf{x_c}) \mid \boldsymbol{\beta_c}, \mathcal{S}_{N_{gp}}\right) \mathcal{L}\left(\boldsymbol{\beta_c} \mid \mathcal{S}_{N_{gp}}\right) d\boldsymbol{\beta_c}, \end{aligned} \tag{14}$$

where $\mathcal{L}^{\text{prior}}\left(\boldsymbol{\beta_c}\right)$ is the prior distribution of $\boldsymbol{\beta_c}$ and $\mathcal{L}$ stand in this equation for the densities of the distribution by abuse of notation.

3.1 Coupling of predictors : the parallel method

According to the previous section, the most intuitive way to calibrate the parameters and to build a predictor of the nested phenomenon is to first consider the phenomena 1 and 2 separately, calibrate the parameters and build a Gaussian predictor for each of them and then to couple the predictors of the two phenomena to obtain a predictor of the nested phenomenon. So the steps of the so-called parallel method are :

1. Calibration and prediction for the phenomena 1 and 2 separately.
2. The predictors of the phenomena 1 and 2 are coupled to build a predictor of the nested phenomenon.
3. The moments of order 1 and 2 of the coupled predictor are evaluated.

3.1.1 Calibration and prediction of the phenomena 1 and 2

The phenomena 1 and 2 are considered separately. The parameters are calibrated and a predictor is built for each phenomenon independently. According to Eq. (2), (3) and (10), we have for $i \in \{1, 2\}$ and $\mathbf{x_i} \in \mathcal{D}_{d_i}$:

$$\begin{aligned} \mathcal{L}\left(\boldsymbol{\beta_i} \mid \mathcal{S}_{N_i}\right) &= \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\beta_i}}^{(N_i)}, \boldsymbol{V}_{\boldsymbol{\beta_i}}^{(N_i)}\right), \\ \mathcal{L}\left(Y_i(\mathbf{x_i}) \mid \mathcal{S}_{N_i}\right) &= \mathcal{GP}\left(\mu_i^{(N_i)}(\mathbf{x_i}), C_i^{(N_i)}(\mathbf{x_i}, \mathbf{x_i})\right). \end{aligned} \tag{15}$$

where for $i \in \{1,2\}$:

$$\begin{aligned}
\boldsymbol{V}_{\boldsymbol{\beta_i}}^{(N_i)} &= \left(\left(\boldsymbol{H}_i^{(N_i)} \right)^t \left(\boldsymbol{R}_i^{(N_i)} \right)^{-1} \boldsymbol{H}_i^{(N_i)} \right)^{-1}, \\
\boldsymbol{\mu}_{\boldsymbol{\beta_i}}^{(N_i)} &= \boldsymbol{V}_{\boldsymbol{\beta_i}}^{(N_i)} \left(\boldsymbol{H}_i^{(N_i)} \right)^t \left(\boldsymbol{R}_i^{(N_i)} \right)^{-1} \mathbb{Y}_i^{(N_i)}, \\
\boldsymbol{\mu}_{\boldsymbol{\beta_i}}^{(N_i)} &\in \mathbb{R}^{p_i}, \quad \boldsymbol{V}_{\boldsymbol{\beta_i}}^{(N_i)} \in M_{p_i,p_i}(\mathbb{R}),
\end{aligned}$$

$$\begin{aligned}
\mu_i^{(N_i)}(\mathbf{x_i}) &= \boldsymbol{h_i}(\mathbf{x_i})^t \boldsymbol{\mu}_{\boldsymbol{\beta_i}}^{(N_i)} + C_i\left(\mathbf{x_i}, \mathbb{X}_i^{(N_i)}\right) \left(\boldsymbol{R}_i^{(N_i)} \right)^{-1} \left(\mathbb{Y}_i^{(N_i)} - \boldsymbol{H}_i^{(N_i)} \boldsymbol{\mu}_{\boldsymbol{\beta_i}}^{(N_i)} \right), \\
C_i^{(N_i)}\left(\mathbf{x_i}, \mathbf{x_i'}\right) &= C_i\left(\mathbf{x_i}, \mathbf{x_i'}\right) - C_i\left(\mathbf{x_i}, \mathbb{X}_i^{(N_i)}\right) \left(\boldsymbol{R}_i^{(N_i)} \right)^{-1} C_i\left(\mathbb{X}_i^{(N_i)}, \mathbf{x_i'}\right) + \boldsymbol{u_i}^{(N_i)}(\mathbf{x_i})^t \boldsymbol{V}_{\boldsymbol{\beta_i}}^{(N_i)} \boldsymbol{u_i}^{(N_i)}\left(\mathbf{x_i'}\right), \\
\mu_i^{(N_i)}(\mathbf{x_i}) &\in \mathbb{R}, \quad C_i^{(N_i)}\left(\mathbf{x_i}, \mathbf{x_i'}\right) \in \mathbb{R},
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{u_i}^{(N_i)}(\mathbf{x_i}) &= \boldsymbol{h_i}(\mathbf{x_i}) - \left(\boldsymbol{H}_i^{(N_i)} \right)^t \left(\boldsymbol{R}_i^{(N_i)} \right)^{-1} C_i\left(\mathbb{X}_i^{(N_i)}, \mathbf{x_i}\right), \\
\boldsymbol{H}_i^{(N_i)} &= \boldsymbol{h_i}\left(\mathbb{X}_i^{(N_i)}\right)^t, \quad \boldsymbol{R}_i^{(N_i)} = C_i\left(\mathbb{X}_i^{(N_i)}, \mathbb{X}_i^{(N_i)}\right), \\
\boldsymbol{H}_i^{(N_i)} &\in M_{N_i,p_i}(\mathbb{R}), \quad \boldsymbol{R}_i^{(N_i)} \in M_{N_i,N_i}(\mathbb{R}).
\end{aligned} \tag{16}$$

At this step, the parameters are calibrated but there is no predictor of the nested phenomenon.

3.1.2 Coupling of the predictors of the phenomena 1 and 2

According to Eq. (10) and (15), a predictor of the nested phenomenon is obtained by coupling the predictors of the phenomena 1 and 2. For a given $\mathbf{x_c} = (\mathbf{x_{c1}}, \mathbf{x_{c2}}) \in \mathcal{D}_{d_c}$, the coupled predictor can be written :

$$\begin{aligned}
Y_2^{(N_2)}\left(\left(Y_1^{(N_1)}(\mathbf{x_{c1}}), \mathbf{x_{c2}}\right)\right) &= \mu_2^{(N_2)}\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}}) + \sigma_1^{(N_1)}(\mathbf{x_{c1}})\, u, \mathbf{x_{c2}}\right)\right) \\
&\quad + \sigma_2^{(N_2)}\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}}) + \sigma_1^{(N_1)}(\mathbf{x_{c1}})\, u, \mathbf{x_{c2}}\right)\right) v,
\end{aligned} \tag{17}$$

where :

- $Y_i^{(N_i)}(\mathbf{x_i})$ denotes a process which has the conditional distribution of $Y_i(\mathbf{x_i})$ given $\mathcal{S}_{N_i}$,
- $\sigma_i^{(N_i)}(\mathbf{x_i}) = \sqrt{C_i^{(N_i)}(\mathbf{x_i}, \mathbf{x_i})}$,
- u and v are independent standard Gaussian random variables.

According to Eq. (17), the predictor of the nested phenomenon is non Gaussian. However, we will focus on the mean and the variance of this predictor, according to the following considerations :

1. the mean of the nested predictor is the best prediction of the nested phenomenon in the L^2 sense,

2. the variance of the nested predictor is an interesting indicator of the uncertainty of this prediction .

3.1.3 Estimation of the moments of order 1 and 2

The general formulation of the moments of order 1 and 2 of the nested predictor can be inferred from Eq. (17), according to the standard normality and the independence of u and v :

$$\begin{aligned}\mathbb{E}\left[Y_2\left(Y_1\left(\mathbf{x_{c1}}\right),\mathbf{x_{c2}}\right)\mid\mathcal{S}_{N_{gp}}\right] &= \mathbb{E}\left[Y_2^{(N_2)}\left(\left(Y_1^{(N_1)}\left(\mathbf{x_{c1}}\right),\mathbf{x_{c2}}\right)\right)\right]\\ &= \mathbb{E}_u\left[\mu_2^{(N_2)}\left(\left(\mu_1^{(N_1)}\left(\mathbf{x_{c1}}\right)+\sigma_1^{(N_1)}\left(\mathbf{x_{c1}}\right)u,\mathbf{x_{c2}}\right)\right)\right],\end{aligned} \tag{18}$$

$$\begin{aligned}\mathbb{E}\left[Y_2\left(Y_1\left(\mathbf{x_{c1}}\right),\mathbf{x_{c2}}\right)^2\mid\mathcal{S}_{N_{gp}}\right] &= \mathbb{E}\left[Y_2^{(N_2)}\left(\left(Y_1^{(N_1)}\left(\mathbf{x_{c1}}\right),\mathbf{x_{c2}}\right)\right)^2\right]\\ &= \mathbb{E}_u\left[\mu_2^{(N_2)}\left(\left(\mu_1^{(N_1)}\left(\mathbf{x_{c1}}\right)+\sigma_1^{(N_1)}\left(\mathbf{x_{c1}}\right)u,\mathbf{x_{c2}}\right)\right)^2\right]\\ &+\mathbb{E}_u\left[\sigma_2^{(N_2)}\left(\left(\mu_1^{(N_1)}\left(\mathbf{x_{c1}}\right)+\sigma_1^{(N_1)}\left(\mathbf{x_{c1}}\right)u,\mathbf{x_{c2}}\right)\right)^2\right].\end{aligned} \tag{19}$$

According to Eq. (18) and (19), the estimates of the moment of order 1 and 2 are the calculations of two one-dimensional integrals with a Gaussian measure. This can be done by quadrature rules or by Monte-Carlo methods ([3]).

Theorem 1. *If the following assumptions are verified :*

1. *The computer model of the phenomenon 2 can be written as a linear combination of products of a polynomial by an exponential of order less than 2, that is :*

$$\left(\boldsymbol{h_2}\left(\mathbf{x_2}\right)\right)_i = g\left(\left(\mathbf{x_2}\right)_1;j_i,a_i,b_i\right)m_i\left(\left(\mathbf{x_2}\right)_{-1}\right), \tag{20}$$

   where

   - $i\in\{1\dots p_2\}$,
   - m_i *is a deterministic function,*
   - $g\left(x_2;j,a,b\right)=x_2^j\exp\left(ax_2+bx_2^2\right)$,
   - $\mathbf{x_2}=\left(\left(\mathbf{x_2}\right)_1,\left(\mathbf{x_2}\right)_{-1}\right)$,

2. *The covariance function C_2 is a tensor product of a Gaussian covariance for $\left(\mathbf{x_2}\right)_1$ by any covariance function for $\left(\mathbf{x_2}\right)_{-1}$, that is :*

$$C_2\left(\mathbf{x_2},\mathbf{x_2}^{'}\right)=\exp\left(-\left(\frac{\left(\mathbf{x_2}\right)_1-\left(\mathbf{x_2}^{'}\right)_1}{l_2}\right)^2\right)C_{\bar{2}}\left(\left(\mathbf{x_2}\right)_{-1},\left(\mathbf{x_2}^{'}\right)_{-1}\right), \tag{21}$$

   where l_2 is the correlation length associated with $\left(\mathbf{x_2}\right)_1$ and $C_{\bar{2}}$ is the covariance function associated with $\left(\mathbf{x_2}\right)_{-1}$,

3. *The condition $b_i<\frac{1}{2\sigma_1^{(N_1)}\left(\mathbf{x_{c1}}\right)^2}$, $\quad i\in\{1\dots p_2\}$ is verified,*

then the conditional moments of order 1 and 2 of $Y_2\left(\left(Y_1\left(\mathbf{x_{c1}}\right), \mathbf{x_{c2}}\right)\right)$ *can be calculated analytically.*

In other words, if the Gaussian process modeling the phenomenon 2 has a trend which is a linear combination of products of polynomials by exponentials of order 2, and a covariance function which is a tensor product of a Gaussian covariance for $\left(\mathbf{x_2}\right)_1$ and some covariance function for $\left(\mathbf{x_2}\right)_{-1}$, then conditionally to some integration criteria on the trend, the moments of order 1 and 2 of the coupling of the predictors of the phenomena 1 and 2 can be computed analytically.

Hence, two methods for obtaining the moments of order 1 and 2 of the predictor of the nested phenomenon can be used : the Monte-Carlo one and the analytical one.
In the Monte-Carlo method the computational cost can be high, especially for the moment of order 2 but the approach can be generalized to the coupling of more than two phenomena.
The computational cost of the analytical method is almost zero but the approach cannot be generalized to the coupling of more than two phenomena and the covariance function of the Gaussian process modeling the phenomenon 2 has to present a Gaussian expression with respect to $\left(\mathbf{x_2}\right)_1$.

3.2 Linearized method

In this section we propose another approach to calibrate the parameters and build a predictor of the nested phenomenon by taking into account the observation points of the three phenomena (1, 2 and nested).

3.2.1 Linearization of the nested model

According to Eq. (13) and the following assumptions :

1. $\boldsymbol{\beta_1} - \bar{\boldsymbol{\beta}}_{\mathbf{1}} = O\left(\delta\right)$,
2. $\epsilon_i = O\left(\delta\right), i \in \{1, 2\}$,
3. δ is small,

we can get a linear approximation of the process modeling the nested phenomenon by a Taylor expansion of order 1 with δ:

$$\begin{aligned}
Y_c(\mathbf{x_c}) = &\ \boldsymbol{h_2}\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\boldsymbol{\beta_1}+\epsilon_1(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right)^t\boldsymbol{\beta_2}+\epsilon_2\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\boldsymbol{\beta_1}+\epsilon_1(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right),\\
= &\ \boldsymbol{h_2}\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1+\boldsymbol{h_1}(\mathbf{x_{c1}})^t\left(\boldsymbol{\beta_1}-\bar{\boldsymbol{\beta}}_1\right)+\epsilon_1(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right)^t\boldsymbol{\beta_2}\\
&+\epsilon_2\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right),\\
= &\ \boldsymbol{h_2}\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right)^t\boldsymbol{\beta_2}+\epsilon_2\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right)\\
&+\tfrac{\partial\boldsymbol{h_2}}{\partial(x_2)_1}\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1\right)^t\bar{\boldsymbol{\beta}}_2\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\left(\boldsymbol{\beta_1}-\bar{\boldsymbol{\beta}}_1\right)+\epsilon_1(\mathbf{x_{c1}})\right),\\
= &\ \boldsymbol{h_2}\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right)^t\boldsymbol{\beta_2}\\
&+\tfrac{\partial\boldsymbol{h_2}}{\partial(x_2)_1}\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right)^t\bar{\boldsymbol{\beta}}_2\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\left(\boldsymbol{\beta_1}-\bar{\boldsymbol{\beta}}_1\right)\right)\\
&+\tfrac{\partial\boldsymbol{h_2}}{\partial(x_2)_1}\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right)^t\bar{\boldsymbol{\beta}}_2\epsilon_1(\mathbf{x_{c1}})+\epsilon_2\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right)
\end{aligned} \tag{22}$$

So the a priori linearized model of the process modeling the nested phenomenon can be written as follows :

$$Y_{c_{lb}}(\mathbf{x_c}) = \boldsymbol{h_{c_{lb}}}(\mathbf{x_c})^t\boldsymbol{\beta}_c+\epsilon_{c_{lb}}(\mathbf{x_c}), \tag{23}$$

where $\epsilon_{c_{lb}}(\mathbf{x_c})$ is the proposed model error :

$$\epsilon_{c_{lb}}(\mathbf{x_c}) = \Delta F_c\left(\mathbf{x_c},\bar{\boldsymbol{\beta}}_{\boldsymbol{c}}\right)\epsilon_1(\mathbf{x_{c1}})+\epsilon_2\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right), \tag{24}$$

and :

$$Y_{c_{lb}}(\mathbf{x_c}) = Y_c(\mathbf{x_c})+\boldsymbol{h_{c_{lb\cdot 1}}}(\mathbf{x_c})^t\bar{\boldsymbol{\beta}}_1, \tag{25}$$

$$\bar{\boldsymbol{\beta}}_{\boldsymbol{c}} = \begin{bmatrix}\bar{\boldsymbol{\beta}}_{\boldsymbol{1}}\\ \bar{\boldsymbol{\beta}}_{\boldsymbol{2}}\end{bmatrix}, \tag{26}$$

$$\Delta F_c\left(\mathbf{x_c},\bar{\boldsymbol{\beta}}_{\boldsymbol{c}}\right) = \frac{\partial\boldsymbol{h_2}}{\partial(x_2)_1}\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right)^t\bar{\boldsymbol{\beta}}_2, \tag{27}$$

$$\begin{aligned}
\boldsymbol{h_{c_{lb}}}(\mathbf{x_c}) &= \begin{bmatrix}\boldsymbol{h_{c_{lb\cdot 1}}}(\mathbf{x_c})\\ \boldsymbol{h_{c_{lb\cdot 2}}}(\mathbf{x_c})\end{bmatrix},\\
\boldsymbol{h_{c_{lb\cdot 1}}}(\mathbf{x_c}) &= \Delta F_c\left(\mathbf{x_c},\bar{\boldsymbol{\beta}}_{\boldsymbol{c}}\right)\boldsymbol{h_1}(\mathbf{x_{c1}}),\\
\boldsymbol{h_{c_{lb\cdot 2}}}(\mathbf{x_c}) &= \boldsymbol{h_2}\left(\left(\boldsymbol{h_1}(\mathbf{x_{c1}})^t\bar{\boldsymbol{\beta}}_1,\mathbf{x_{c2}}\right)\right).
\end{aligned} \tag{28}$$

Thanks to the Gaussianity and independence of ϵ_1 and ϵ_2 we thus have a Gaussian process model for the nested phenomenon :

$$\mathcal{L}\left(Y_{c_{lb}}(\mathbf{x_c})\mid\boldsymbol{\beta}_c\right) = \mathcal{GP}\left(\boldsymbol{h_{c_{lb}}}(\mathbf{x_c})^t\boldsymbol{\beta}_c,\ C_{c_{lb}}\left(\mathbf{x_c},\mathbf{x_c'}\right)\right), \tag{29}$$

where the covariance function of the Gaussian process is defined as follows :

$$\begin{aligned} C_{c_{lb}}\left(\mathbf{x_c}, \mathbf{x_c^{'}}\right) = \; & \Delta F_c\left(\mathbf{x_c}, \bar{\boldsymbol{\beta}}_{\boldsymbol{c}}\right) C_1\left(\mathbf{x_{c1}}, \mathbf{x_{c1}^{'}}\right) \Delta F_c\left(\mathbf{x_c^{'}}, \bar{\boldsymbol{\beta}}_{\boldsymbol{c}}\right) \\ & + C_2\left(\left(\boldsymbol{h_1}\left(\mathbf{x_{c1}}\right)^t \bar{\boldsymbol{\beta}}_1, \mathbf{x_{c2}}\right), \left(\boldsymbol{h_1}\left(\mathbf{x_{c1}^{'}}\right)^t \bar{\boldsymbol{\beta}}_1, \mathbf{x_{c2}^{'}}\right)\right). \end{aligned} \tag{30}$$

Given Eq. (10) and (23) and the independence of ϵ_1 and ϵ_2 a joint model for the three phenomena (1, 2 and nested) is proposed :

$$\mathcal{L}\left(\boldsymbol{Y_{lb}}\left(\mathbf{x_{lb}}\right) \mid \boldsymbol{\beta_c}\right) = \mathcal{GP}\left(\, \boldsymbol{h_{lb}}\left(\mathbf{x_{lb}}\right)^t \boldsymbol{\beta}_c \,,\, \boldsymbol{C}_{lb}\left(\mathbf{x_{lb}}, \mathbf{x_{lb}}\right)\right), \tag{31}$$

where :

$$\mathbf{x_{lb}} = \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \mathbf{x_c} \end{bmatrix}, \qquad \boldsymbol{Y_{lb}}\left(\mathbf{x_{lb}}\right) = \begin{bmatrix} Y_1\left(\mathbf{x_1}\right) \\ Y_2\left(\mathbf{x_2}\right) \\ Y_{c_{lb}}\left(\mathbf{x_c}\right) \end{bmatrix}, \tag{32}$$

$$\boldsymbol{h_{lb}}\left(\mathbf{x_{lb}}\right) = \begin{bmatrix} \boldsymbol{h_1}\left(\mathbf{x_1}\right) & 0 & \boldsymbol{h}_{\boldsymbol{c_{lb\cdot 1}}}\left(\mathbf{x_c}\right) \\ 0 & \boldsymbol{h_2}\left(\mathbf{x_2}\right) & \boldsymbol{h}_{\boldsymbol{c_{lb\cdot 2}}}\left(\mathbf{x_c}\right) \end{bmatrix}, \tag{33}$$

$$\boldsymbol{C}_{lb}\left(\mathbf{x_{lb}}, \mathbf{x_{lb}^{'}}\right) =$$

$$\begin{bmatrix} C_1\left(\mathbf{x_1}, \mathbf{x_1^{'}}\right) & 0 & C_1\left(\mathbf{x_1}, \mathbf{x_{c1}^{'}}\right) \Delta F_c\left(\mathbf{x_c^{'}}, \bar{\boldsymbol{\beta}}_{\boldsymbol{c}}\right) \\ 0 & C_2\left(\mathbf{x_2}, \mathbf{x_2^{'}}\right) & C_2\left(\mathbf{x_2}, \left(\boldsymbol{h_1}\left(\mathbf{x_{c1}^{'}}\right)^t \bar{\boldsymbol{\beta}}_1, \mathbf{x_{c2}^{'}}\right)\right) \\ \Delta F_c\left(\mathbf{x_c}, \bar{\boldsymbol{\beta}}_{\boldsymbol{c}}\right) C_1\left(\mathbf{x_{c1}}, \mathbf{x_1^{'}}\right) & C_2\left(\left(\boldsymbol{h_1}\left(\mathbf{x_{c1}}\right)^t \bar{\boldsymbol{\beta}}_1, \mathbf{x_{c2}}\right), \mathbf{x_2^{'}}\right) & C_c\left(\mathbf{x_c}, \mathbf{x_c^{'}}\right) \end{bmatrix}. \tag{34}$$

3.2.2 Calibration of the computer models' parameters and posterior predictive distributions

We suppose that we have observation points of the studied phenomena. Given Eq. (31), we can infer from Eq. (2) and (3) that, in case of a non informative uniform prior over $\mathbb{R}^{p_1+p_2}$ for the parameters $\boldsymbol{\beta_c}$, the posterior distributions of the parameters and of the Gaussian processes associated with the three phenomena are Gaussian. We have :

$$\begin{aligned} \mathcal{L}\left(\boldsymbol{\beta_c} \mid \mathcal{S}_{N_1} \cup \mathcal{S}_{N_2} \cup \mathcal{S}_{N_c}\right) &= \mathcal{N}\left(\, \boldsymbol{\mu}_{\boldsymbol{\beta_c}}^{(N_{lb})} \,,\, \boldsymbol{V}_{\boldsymbol{\beta_c}}^{(N_{lb})} \right), \\ \mathcal{L}\left(\boldsymbol{Y_{lb}}\left(\mathbf{x_{lb}}\right) \mid \mathcal{S}_{N_1} \cup \mathcal{S}_{N_2} \cup \mathcal{S}_{N_c}\right) &= \mathcal{GP}\left(\, \boldsymbol{\mu_{lb}}^{(N_{lb})}\left(\mathbf{x_{lb}}\right) \,,\, \boldsymbol{C}_{lb}^{(N_{lb})}\left(\mathbf{x_{lb}}, \mathbf{x_{lb}}\right)\right), \end{aligned} \tag{35}$$

where :

$$\boldsymbol{V}_{\boldsymbol{\beta_c}}^{(N_{lb})} = \left(\left(\boldsymbol{H}_{lb}^{(N_{lb})} \right)^t \left(\boldsymbol{R}_{lb}^{(N_{lb})} \right)^{-1} \boldsymbol{H}_{lb}^{(N_{lb})} \right)^{-1},$$
$$\boldsymbol{\mu}_{\boldsymbol{\beta_c}}^{(N_{lb})} = \boldsymbol{V}_{\boldsymbol{\beta_c}}^{(N_{lb})} \left(\boldsymbol{H}_{lb}^{(N_{lb})} \right)^t \left(\boldsymbol{R}_{lb}^{(N_{lb})} \right)^{-1} \mathbb{Y}_{lb}^{(N_{lb})},$$

$$\begin{aligned}
\boldsymbol{\mu_{lb}}^{(N_{lb})}(\mathbf{x_{lb}}) = \;& \boldsymbol{h_{lb}}(\mathbf{x_{lb}})^t \boldsymbol{\mu}_{\boldsymbol{\beta_c}}^{(N_{lb})} + \boldsymbol{C}_{lb}\left(\mathbf{x_{lb}}, \mathbb{X}_{lb}^{(N_{lb})}\right) \left(\boldsymbol{R}_{lb}^{(N_{lb})} \right)^{-1} \left(\mathbb{Y}_{lb}^{(N_{lb})} - \boldsymbol{H}_{lb}^{(N_{lb})} \boldsymbol{\mu}_{\boldsymbol{\beta_c}}^{(N_{lb})} \right), \\
\boldsymbol{C}_{lb}^{(N_{lb})}(\mathbf{x_{lb}}, \mathbf{x_{lb}}') = \;& \boldsymbol{C}_{lb}(\mathbf{x_{lb}}, \mathbf{x_{lb}}') - \boldsymbol{C}_{lb}\left(\mathbf{x_{lb}}, \mathbb{X}_{lb}^{(N_{lb})}\right) \left(\boldsymbol{R}_{lb}^{(N_{lb})} \right)^{-1} \boldsymbol{C}_{lb}\left(\mathbb{X}_{lb}^{(N_{lb})}, \mathbf{x_{lb}}'\right) \\
& + \left(\boldsymbol{u_{lb}}^{(N_{lb})}(\mathbf{x_{lb}}) \right)^t \boldsymbol{V}_{\boldsymbol{\beta_c}}^{(N_{lb})} \boldsymbol{u_{lb}}^{(N_{lb})}(\mathbf{x_{lb}}'),
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{u_{lb}}^{(N_{lb})}(\mathbf{x_{lb}}) &= \boldsymbol{h_{lb}}(\mathbf{x_{lb}}) - \left(\boldsymbol{H}_{lb}^{(N_{lb})} \right)^t \left(\boldsymbol{R}_{lb}^{(N_{lb})} \right)^{-1} \boldsymbol{C}_{lb}\left(\mathbb{X}_{lb}^{(N_{lb})}, \mathbf{x_{lb}}\right), \\
\boldsymbol{H}_{lb}^{(N_{lb})} &= \boldsymbol{h_{lb}}\left(\mathbb{X}_{lb}^{(N_{lb})}\right)^t, \quad \boldsymbol{R}_{lb}^{(N_{lb})} = \boldsymbol{C}_{lb}\left(\mathbb{X}_{lb}^{(N_{lb})}, \mathbb{X}_{lb}^{(N_{lb})}\right).
\end{aligned} \tag{36}$$

This method thus enables to obtain Gaussian posterior distributions for the parameters and for the Gaussian processes associated with all the phenomena. However, the linearization assumption has to be validated, which can be done by cross-validation techniques ([7],[13]).

3.3 A posteriori linearized method

This method of calibration and prediction is also based on a linearization of the nested phenomenon, but instead of considering the initial Gaussian process modeling the phenomenon 1, we consider the predictor of the phenomenon 1, whose error is smaller. So this enables to reduce the risks posed by the linearization of the process modeling the nested phenomenon. Moreover, because the posterior distributions of the parameters and of the Gaussian process modeling the phenomenon 1 are estimated separately, the risk of error compensation on the parameters $\boldsymbol{\beta_1}$ is also reduced.
The steps of the method can be summarized as follows :

1. Calibration of the parameters of the first computer model and construction of a predictor of the phenomenon 1 given the observation points of this phenomenon.

2. The process modeling the nested phenomenon is linearized by replacing the phenomenon 1 by its predictor, this enables to write the joint distribution of the processes modeling the phenomena 2 and nested as a multi-dimensional Gaussian process.

3. Calibration of the parameters of the second computer model and construction of a predictor for the phenomena 2 and nested given the observation points of these phenomena.

3.3.1 Linearization of the nested model

We suppose that we have observation points of the phenomenon 1, which are denoted by $\mathcal{S}_{N_1}$ and that the prior distribution of $\boldsymbol{\beta_1}$ is uniform over $\mathbb{R}^{p_1}$. According to the method presented in

section 2 the posterior distributions of the parameters and of the Gaussian process modeling the phenomenon 1 are Gaussian.
Let $\epsilon_1^{(N_1)}(\mathbf{x_1})$ be a process with distribution $\mathcal{GP}\left(0\ ,\ C_1^{(N_1)}(\mathbf{x_1},\mathbf{x_1})\right)$, then the process $\mu_1^{(N_1)}(\mathbf{x_1})+\epsilon_1^{(N_1)}(\mathbf{x_1})$ has the conditional distribution of $Y_1(\mathbf{x_1})$ given $\mathcal{S}_{N_1}$. Furthermore, the conditional distribution of $Y_c(\mathbf{x_c})$ given $\mathcal{S}_{N_1}$ is the distribution of $Y_2\left(\left(Y_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right)$. This process has the form :

$$\begin{aligned} Y_2\left(\left(Y_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right) &= \boldsymbol{h_2}\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}})+\epsilon_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right)^t\boldsymbol{\beta_2} \\ &\quad +\epsilon_2\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}})+\epsilon_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right). \end{aligned} \tag{37}$$

Assuming that $\epsilon_1^{(N_1)}(\mathbf{x_1})$ is small, we can linearize the previous equation :

$$\begin{aligned} Y_2\left(\left(Y_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right) &= \boldsymbol{h_2}\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right)^t\boldsymbol{\beta_2} \\ &\quad +\tfrac{\partial \boldsymbol{h_2}}{\partial (x_2)_1}\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right)^t\boldsymbol{\beta_2}\,\epsilon_1^{(N_1)}(\mathbf{x_{c1}}) \\ &\quad +\epsilon_2\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right). \end{aligned} \tag{38}$$

Thanks to the Gaussianity and independence of $\epsilon_1^{(N_1)}(\mathbf{x_1})$ and $\epsilon_2(\mathbf{x_2})$, the process modeling the nested phenomenon is Gaussian, so the joint distribution of the processes modeling the phenomena 2 and nested can be written :

$$\mathcal{L}\left(\boldsymbol{Y_{la}}(\mathbf{x_{la}})\ |\ \boldsymbol{\beta_2},\mathcal{S}_{N_1}\right)=\mathcal{GP}\left(\boldsymbol{h_{la}}(\mathbf{x_{la}})^t\boldsymbol{\beta}_2\ ,\ \boldsymbol{C}_{la}(\mathbf{x_{la}},\mathbf{x_{la}}')\right), \tag{39}$$

where :

$$\mathbf{x_{la}}=\begin{bmatrix}\mathbf{x_2}\\ \mathbf{x_c}\end{bmatrix},\qquad \boldsymbol{Y_{la}}(\mathbf{x_{la}})=\begin{bmatrix}Y_2(\mathbf{x_2})\\ Y_c(\mathbf{x_c})\end{bmatrix}, \tag{40}$$

$$\boldsymbol{h_{la}}(\mathbf{x_{la}})=\begin{bmatrix}\boldsymbol{h_2}(\mathbf{x_2}) & \boldsymbol{h_2}\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right)\right)\end{bmatrix}. \tag{41}$$

The errors $\epsilon_1^{(N)}(\mathbf{x_1})$ and $\epsilon_2(\mathbf{x_2})$ are independent centered Gaussian processes, so the covariance function $\boldsymbol{C}_{la}(\mathbf{x_{la}},\mathbf{x_{la}}')$ is defined by :

$$\boldsymbol{C}_{la}(\mathbf{x_{la}},\mathbf{x_{la}}')=\begin{bmatrix} C_2\left(\mathbf{x_2},\mathbf{x_2'}\right) & C_2\left(\mathbf{x_2},\left(\mu_1^{(N_1)}\left(\mathbf{x_{c1}'}\right),\mathbf{x_{c2}'}\right)\right) \\ C_2\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right),\mathbf{x_2'}\right) & C_2\left(\left(\mu_1^{(N_1)}(\mathbf{x_{c1}}),\mathbf{x_{c2}}\right),\left(\mu_1^{(N_1)}\left(\mathbf{x_{c1}'}\right),\mathbf{x_{c2}'}\right)\right)\end{bmatrix}. \tag{42}$$

3.3.2 Calibration of the parameters of the second computer model and posterior predictive distribution

We suppose that we have observation points of the phenomena 2 and nested, which are denoted by $\mathcal{S}_{N_2}\cup\mathcal{S}_{N_c}$ and that the prior distribution of $\boldsymbol{\beta_2}$ is uniform over $\mathbb{R}^{p_2}$.

Given Eq. (39), we can infer from Eq. (2) and (3) that the posterior distributions of $\boldsymbol{\beta_2}$ and of the process modeling the phenomena 2 and nested are Gaussian and are defined as follows :

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\beta_2} \mid \mathcal{S}_{N_2} \cup \mathcal{S}_{N_c}, \mathcal{S}_{N_1}\right) &= \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\beta_2}}^{(N_{la})}, \boldsymbol{V}_{\boldsymbol{\beta_2}}^{(N_{la})}\right), \\ \mathcal{L}\left(\boldsymbol{Y_{la}}\left(\mathbf{x_{la}}\right) \mid \mathcal{S}_{N_2} \cup \mathcal{S}_{N_c}, \mathcal{S}_{N_1}\right) &= \mathcal{GP}\left(\boldsymbol{\mu_{la}}^{(N_{la})}\left(\mathbf{x_{la}}\right), \boldsymbol{C}_{la}^{(N_{la})}\left(\mathbf{x_{la}}, \mathbf{x_{la}}\right)\right),\end{aligned} \tag{43}$$

where :

$$\begin{aligned}\boldsymbol{V}_{\boldsymbol{\beta_2}}^{(N_{la})} &= \left(\left(\boldsymbol{H}_{la}^{(N_{la})}\right)^t \left(\boldsymbol{R}_{la}^{(N_{la})}\right)^{-1} \boldsymbol{H}_{la}^{(N_{la})}\right)^{-1}, \\ \boldsymbol{\mu}_{\boldsymbol{\beta_2}}^{(N_{la})} &= \boldsymbol{V}_{\boldsymbol{\beta_2}}^{(N_{la})} \left(\boldsymbol{H}_{la}^{(N_{la})}\right)^t \left(\boldsymbol{R}_{la}^{(N_{la})}\right)^{-1} \mathbb{Y}_{la}^{(N_{la})},\end{aligned}$$

$$\begin{aligned}\boldsymbol{\mu_{la}}^{(N_{la})}\left(\mathbf{x_{la}}\right) = &\ \boldsymbol{h_{la}}\left(\mathbf{x_{la}}\right)^t \boldsymbol{\mu}_{\boldsymbol{\beta_2}}^{(N_{la})} + \boldsymbol{C}_{la}\left(\mathbf{x_{la}}, \mathbb{X}_{la}^{(N_{la})}\right)\left(\boldsymbol{R}_{la}^{(N_{la})}\right)^{-1}\left(\mathbb{Y}_{la}^{(N_{la})} - \boldsymbol{H}_{la}^{(N_{la})}\boldsymbol{\mu}_{\boldsymbol{\beta_2}}^{(N_{la})}\right), \\ \boldsymbol{C}_{la}^{(N_{la})}\left(\mathbf{x_{la}}, \mathbf{x_{la}}'\right) = &\ \boldsymbol{C}_{la}\left(\mathbf{x_{la}}, \mathbf{x_{la}}'\right) - \boldsymbol{C}_{la}\left(\mathbf{x_{la}}, \mathbb{X}_{la}^{(N_{la})}\right)\left(\boldsymbol{R}_{la}^{(N_{la})}\right)^{-1}\boldsymbol{C}_{la}\left(\mathbb{X}_{la}^{(N_{la})}, \mathbf{x_{la}}'\right) \\ &+ \left(\boldsymbol{u_{la}}^{(N_{la})}\left(\mathbf{x_{la}}\right)\right)^t \boldsymbol{V}_{\boldsymbol{\beta_2}}^{(N_{la})} \boldsymbol{u_{la}}^{(N_{la})}\left(\mathbf{x_{la}}'\right),\end{aligned}$$

$$\begin{aligned}\boldsymbol{u_{la}}^{(N_{la})}\left(\mathbf{x_{la}}\right) &= \boldsymbol{h_{la}}\left(\mathbf{x_{la}}\right) - \left(\boldsymbol{H}_{la}^{(N_{la})}\right)^t \left(\boldsymbol{R}_{la}^{(N_{la})}\right)^{-1} \boldsymbol{C}_{la}\left(\mathbb{X}_{la}^{(N_{la})}, \mathbf{x_{la}}\right), \\ \boldsymbol{H}_{la}^{(N_{la})} &= \boldsymbol{h_{la}}\left(\mathbb{X}_{la}^{(N_{la})}\right)^t, \quad \boldsymbol{R}_{la}^{(N_{la})} = \boldsymbol{C}_{la}\left(\mathbb{X}_{la}^{(N_{la})}, \mathbb{X}_{la}^{(N_{la})}\right).\end{aligned} \tag{44}$$

As mentioned earlier, the posterior distributions of the parameters $\boldsymbol{\beta_1}$ and of process modeling the phenomenon 1 are Gaussian (Eq. (2), (3) and (10)), so this method enables to obtain Gaussian posterior distributions for the parameters and Gaussian predictors for all the phenomena.

4 Adaptive sequential design

In the previous section, methods of calibration and prediction of a nested phenomenon have been presented. These methods require that observation points of the phenomena 1, 2 and/or nested are available. These methods have the following interesting properties :

1. They enable to take into account the partial information, that is the observation points of the phenomena 1 or 2 alone.

2. The moments of order 1 and 2 of the posterior distributions of the parameters $\boldsymbol{\beta_c}$ and of the process modeling the nested phenomenon $Y_c\left(\mathbf{x_c}\right)$ given the observation points can be computed analytically.

The performance of calibration and prediction is therefore conditioned by the Design of Experiments. In this section methods of DoE suited for the case of a nested phenomenon are studied.

4.1 Definition of a reference Design of Experiments

A reference method to choose the observation points in the case of a nested phenomenon is to build a maximin Latin Hypercube Design (LHS) on the input space $\mathcal{D}_{d_c}$ and then to generate the observations of the three phenomena (1, 2 and nested) :

$$\mathbf{x_c} \mapsto \begin{cases} y_1\left(\mathbf{x_{c1}}\right) \\ y_2\left(\left(y_1\left(\mathbf{x_{c1}}\right), \mathbf{x_{c2}}\right)\right) \\ y_c\left(\mathbf{x_c}\right) \end{cases} .$$

where $y_2\left(\left(y_1\left(\mathbf{x_{c1}}\right), \mathbf{x_{c2}}\right)\right) = y_c\left(\mathbf{x_c}\right)$ according to Eq. (9).
So if the Latin Hypercube Design on $\mathcal{D}_{d_c}$ contains N samples, the set of available observations is defined as follows :

1. $\mathbb{X}_1^{(N)} = \left(\mathbf{x_{c1}}^{(1)}, \ldots, \mathbf{x_{c1}}^{(N)}\right)$ and $\mathbb{Y}_1^{(N)} = \left(y_1\left(\mathbf{x_{c1}}^{(1)}\right), \ldots, y_1\left(\mathbf{x_{c1}}^{(N)}\right)\right)$,
2. $\mathbb{X}_2^{(N)} = \left(\left(y_1\left(\mathbf{x_{c1}}^{(1)}\right), \mathbf{x_{c2}}^{(1)}\right), \ldots, \left(y_1\left(\mathbf{x_{c1}}^{(N)}\right), \mathbf{x_{c2}}^{(N)}\right)\right)$ and
   $\mathbb{Y}_2^{(N)} = \left(y_2\left(y_1\left(\mathbf{x_{c1}}^{(1)}\right), \mathbf{x_{c2}}^{(1)}\right), \ldots, y_2\left(y_1\left(\mathbf{x_{c1}}^{(N)}\right), \mathbf{x_{c2}}^{(N)}\right)\right)$,
3. $\mathbb{X}_c^{(N)} = \left(\mathbf{x_c}^{(1)}, \ldots, \mathbf{x_c}^{(N)}\right)$ and $\mathbb{Y}_c^{(N)} = \left(y_c\left(\mathbf{x_c}^{(1)}\right), \ldots, y_c\left(\mathbf{x_c}^{(N)}\right)\right)$.

Throughout this paper, this method of Design of Experiments will be referred as the chained Latin Hypercube Design.

4.2 Sequential designs in the case of a nested phenomenon

In this section we will study methods of sequential (step by step) choice of additional observation points. In order to improve the computational performance, the new observation points are chosen in one of the phenomena 1 and 2. There are two main difficulties to choose a new observation point:

1. on which phenomenon to add a new observation point,
2. for a given phenomenon, which new observation point to choose.

If we denote by $\Omega_i^{N_i} \subset \mathcal{D}_{d_i}$ the set of candidates in the input space of the phenomenon y_i when there are already N_i observations of the phenomenon y_i, the new candidates can be chosen according to the following criteria :

- for the methods based on the covariance matrix of the posterior distribution of the parameters :
  - the partial A-Optimal criterion is an alternate method where each new observation point is alternately chosen on the phenomenon 1 or 2. The new observation point is selected in order to minimize the trace of the covariance matrix of the posterior distribution of the parameters of the first computer model (respectively second computer

model) given the existing observation points of the phenomenon 1 (respectively 2) plus the new candidate for the phenomenon 1 (respectively 2) :

$$\underset{\mathbb{x}_{\boldsymbol{i}} \subset \Omega_i^{N_i},\ i=(n_{iter}\%2)+1}{\operatorname{argmin}} \quad \operatorname{tr}\left(\mathbb{V}\left(\boldsymbol{\beta_i} \mid \mathcal{S}_{N_i} \cup \mathbb{x}_{\boldsymbol{i}}\right)\right), \tag{45}$$

where $n\%2$ is equal to 0 if n is even and to 1 otherwise,

– the A-Optimal criterion is a global method (in contrast to an alternate method) which selects the best candidate among the candidates of the phenomena 1 and 2 in order to minimize the trace of the covariance matrix of the posterior joint distribution of the parameters of the two computer models :

$$\underset{\mathbb{x}_{\boldsymbol{i}} \subset \Omega_i^{N_i},\ i\in\{1,2\}}{\operatorname{argmin}} \quad \operatorname{tr}\left(\mathbb{V}\left(\boldsymbol{\beta_c} \mid \mathcal{S}_{N_{seq}}\left(\mathbb{x}_{\boldsymbol{i}}\right)\right)\right), \tag{46}$$

- for the methods based on the variance of the predictor of the nested phenomenon :

  – the partial I-Optimal criterion is an alternate method where each new observation point is alternately chosen on the phenomenon 1 or 2. The new observation point is selected in order to minimize the integrated variance of the predictor of the phenomenon 1 (respectively 2) given the existing observation points of the phenomenon 1 (respectively 2) plus the new candidate for the phenomenon 1 (respectively 2) :

$$\underset{\mathbb{x}_{\boldsymbol{i}} \subset \Omega_i^{N_i},\ i=(n_{iter}\%2)+1}{\operatorname{argmin}} \quad \int_{\mathbf{x_c}\in\mathcal{D}_{d_c}} \mathbb{V}\left(Y_i\left(\mathbf{x_i}\right) \mid \mathcal{S}_{N_i} \cup \mathbb{x}_{\boldsymbol{i}}\right) d\mathbf{x_i}, \tag{47}$$

  – the I-Optimal criterion is a global method which selects the best among the candidates of the phenomena 1 and 2 in order to minimize the integrated variance of the predictor of the nested phenomenon :

$$\underset{\mathbb{x}_{\boldsymbol{i}} \subset \Omega_i^{N_i},\ i\in\{1,2\}}{\operatorname{argmin}} \quad \int_{\mathbf{x_c}\in\mathcal{D}_{d_c}} \mathbb{V}\left(Y_c\left(\mathbf{x_c}\right) \mid \mathcal{S}_{N_{seq}}\left(\mathbb{x}_{\boldsymbol{i}}\right)\right) d\mathbf{x_c}, \tag{48}$$

where :

- n_{iter} is the iteration of the sequential design,
- $\mathcal{S}_{N_{seq}}\left(\mathbb{x}_{\boldsymbol{i}}\right)$ denotes the available information including the new candidate. Table 1 summarizes the way to take into account the available information depending on the method of calibration and prediction and the design criterion. In particular, some cases require to have an estimate of $y_1\left(\mathbb{x}_{\mathbf{1}}\right)$. In this work, inspired by the Kriging Believer strategy proposed in [9], we will replace the unknown response of the phenomenon $y_1\left(\mathbb{x}_{\mathbf{1}}\right)$ by the expectation of its kriging predictor $\mu_1^{(N_1)}\left(\mathbb{x}_{\mathbf{1}}\right)$.

5 Applications

In this section, the methods presented in the sections 3 and 4 are illustrated on two examples : the first one is analytical and one-dimensional and the second one is an hydrodynamic example, which is the coupling of two codes, a CFD code and a balistic code, and is multidimensional.

Method	Candidate on phenomenon 1	Candidate on phenomenon 2
Parallel, A-optimal	$\mathcal{S}_{N_1} \cup \mathbb{x}_{\mathbf{1}} \cup \mathcal{S}_{N_2}$	$\mathcal{S}_{N_1} \cup \mathcal{S}_{N_2} \cup \mathbb{x}_{\mathbf{2}}$
Parallel, I-optimal	$\mathcal{S}_{N_2}$, $\mathcal{S}_{N_1} \cup \{\mathbb{x}_{\mathbf{1}}, \mu_1^{(N_1)}(\mathbb{x}_{\mathbf{1}})\}$	$\mathcal{S}_{N_2} \cup \mathbb{x}_{\mathbf{2}}$, $\mathcal{S}_{N_1}$
A priori linearized, A-optimal or I-optimal	$\mathcal{S}_{N_1} \cup \mathbb{x}_{\mathbf{1}} \cup \mathcal{S}_{N_2} \cup \mathcal{S}_{N_c}$	$\mathcal{S}_{N_1} \cup \mathcal{S}_{N_2} \cup \mathbb{x}_{\mathbf{2}} \cup \mathcal{S}_{N_c}$
A posteriori linearized, A-optimal or I-optimal	$\mathcal{S}_{N_2} \cup \mathcal{S}_{N_c}$, $\mathcal{S}_{N_1} \cup \{\mathbb{x}_{\mathbf{1}}, \mu_1^{(N_1)}(\mathbb{x}_{\mathbf{1}})\}$	$\mathcal{S}_{N_2} \cup \mathbb{x}_{\mathbf{2}} \cup \mathcal{S}_{N_c}$, $\mathcal{S}_{N_1}$

Table 1: Definition of the information $\mathcal{S}_{N_{seq}}(\mathbb{x}_{\boldsymbol{i}})$ depending on the calibration and prediction method and the design criterion

5.1 Definition of the reference method for the calibration and prediction

For the calibration and the prediction the black-box method will be considered as a reference. This method corresponds to the case where the nested phenomenon is considered as a unique phenomenon. The nested computer model is linearized in the vicinity of $\bar{\boldsymbol{\beta}}_{\boldsymbol{c}}$. The process obtained has the same mean than in the a priori linearized method, but the model error is different. According to Eq. (23) and (25), in the black-box method, the process $Y_c(\mathbf{x_c})$ is defined as follows :

$$Y_c(\mathbf{x_c}) + \boldsymbol{h}_{\boldsymbol{c_{lb}}\cdot\mathbf{1}}(\mathbf{x_c})^t \bar{\boldsymbol{\beta}}_{\mathbf{1}} = \boldsymbol{h}_{\boldsymbol{c_{lb}}}(\mathbf{x_c})^t \boldsymbol{\beta}_{\boldsymbol{c}} + \epsilon_c(\mathbf{x_c}), \tag{49}$$

where $\epsilon_c(\mathbf{x_c})$ is a Gaussian process whose covariance function is supposed to be stationary.
According to Eq. (2) and (49), the calibration and the prediction are done according to the method presented in section 2.

5.2 Observation points choice and associated computational cost

In the reference design, the observation points are drawn according to chained Latin Hypercube Designs (see section 4.1) of increasing sizes on $\mathcal{D}_{d_c}$.
In the sequential designs, an initial small dimensional chained Latin Hypercube Design is chosen and new observation points are added one by one on the phenomena 1 or 2.
Each chained Latin Hypercube Design is repeated 50 times and the figures 3, 4, 6 and 7 present therefore the median of the results obtained for the 50 repetitions of the design.

In the presented examples the computational cost is considered to be 1 for the phenomena 1 and 2 such that it is equal to 2 for the nested phenomenon. The cost of a chained Latin Hypercube Design of n observation points is therefore 2n.

5.3 Covariance function

The covariance functions of ϵ_1, ϵ_2 and ϵ_c are chosen in a parametric family. Except for the figure 4, the covariance functions are Matérn $\frac{5}{2}$. In figure 4 the covariance functions of ϵ_1 and ϵ_2 are Gaussian.
Identifying the covariance parameters is a key issue when using the Gaussian process regression. However, to focus on the interest of exploiting the nested structure of the codes in terms of

prediction and calibration, we decided to fix the parameters of the covariance functions from a preliminary study. In the following examples, the covariance functions are thus supposed to be known.

5.4 Characteristics of the examples

5.4.1 Analytic example

For the analytic example, the properties of the computer models and of the phenomena are :

$$\boldsymbol{h_1}(x) = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}, \qquad \boldsymbol{\beta_1^*} = \begin{bmatrix} -2 \\ 0.25 \\ 0.0625 \end{bmatrix}, \qquad y_1(x) = \boldsymbol{h_1}(x)^t \boldsymbol{\beta_1^*} - 0.25\cos(2\pi x), \quad (50)$$

$$\boldsymbol{h_2}(x) = \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix}, \qquad \boldsymbol{\beta_2^*} = \begin{bmatrix} 6 \\ -5 \\ -2 \\ 1 \end{bmatrix}, \qquad y_2(x) = \boldsymbol{h_2}(x)^t \boldsymbol{\beta_2^*} - 0.25\cos(2\pi x), \quad (51)$$

$$\mathcal{D}_{d_1} = \mathcal{D}_{d_c} = [-7, 7], \qquad \mathcal{D}_{d_2} = [-2, 4]. \quad (52)$$

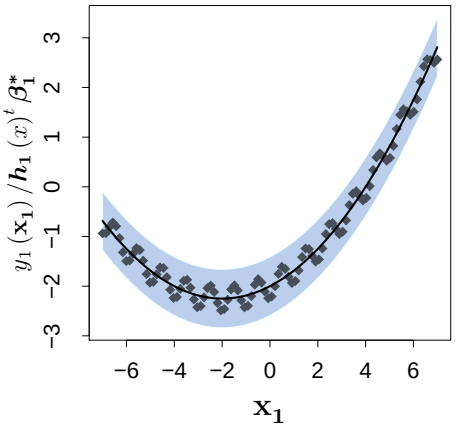
Figure 1 presents for the phenomena 1, 2 and nested, the computer model and its error and the prediction and its error (which are characterized by the conditional mean and variance of the process modeling the phenomenon). It can be seen how the predictor, which has the conditional distribution of the process modeling the phenomenon, enables to reduce the uncertainty of prediction of a phenomenon. It can be seen that the errors of the computer models for the phenomena 1 and 2 are regular, the assumption of stationarity is correct. The error for the computer model of the nested phenomenon is more chaotic and the hypothesis of stationarity does not seem to be valid.

5.4.2 Hydrodynamical example

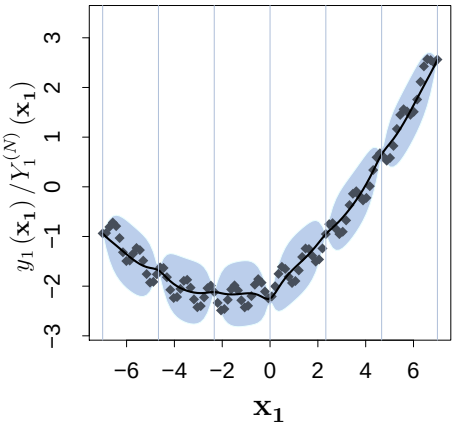
This example consists in the coupling of two computer codes. The objective is to determine the impact point of a conical projectile.
In what follows the term phenomenon refers to a computer code and the term computer model refers to the analytic approximation of a computer code.

The phenomenon 1 computes the drag coefficient of a cone divided by the height of the cone. The dimension of the inputs $\mathbf{x_1}$ is 2. These inputs are the height and the half-angle of the cone.

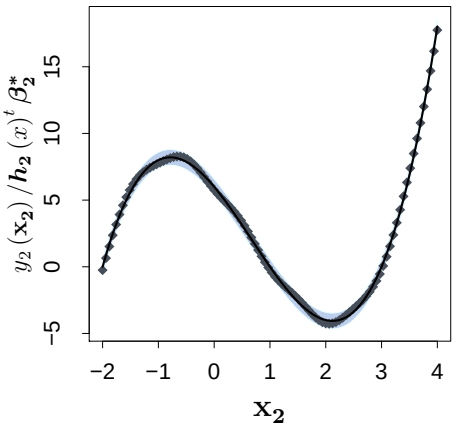
The phenomenon 2 is the distance between the beginning and the end of the ballistic trajectory of the cone. Figure 2 presents an example of ballistic trajectory. The dimension of the inputs $\mathbf{x_2}$ is 3. These inputs are the output of the phenomenon 1, the initial velocity and the initial angle of the ballistic trajectory of the cone.



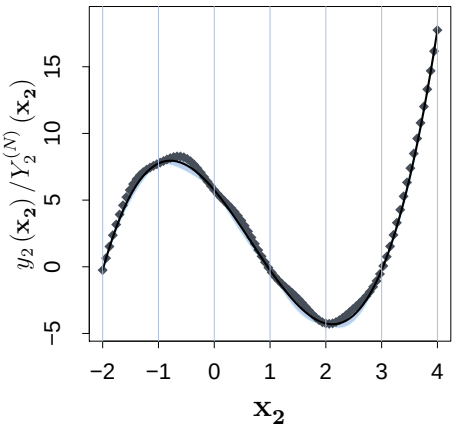
(a) Phenomenon 1 : computer model (initial process)



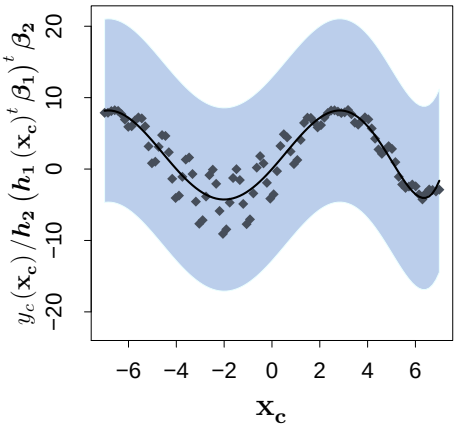
(b) Phenomenon 1 : estimator (conditional process)



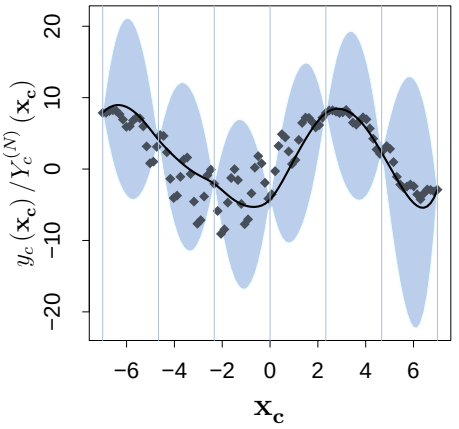
(c) Phenomenon 2 : computer model (initial process)



(d) Phenomenon 2 : estimator (conditional process)

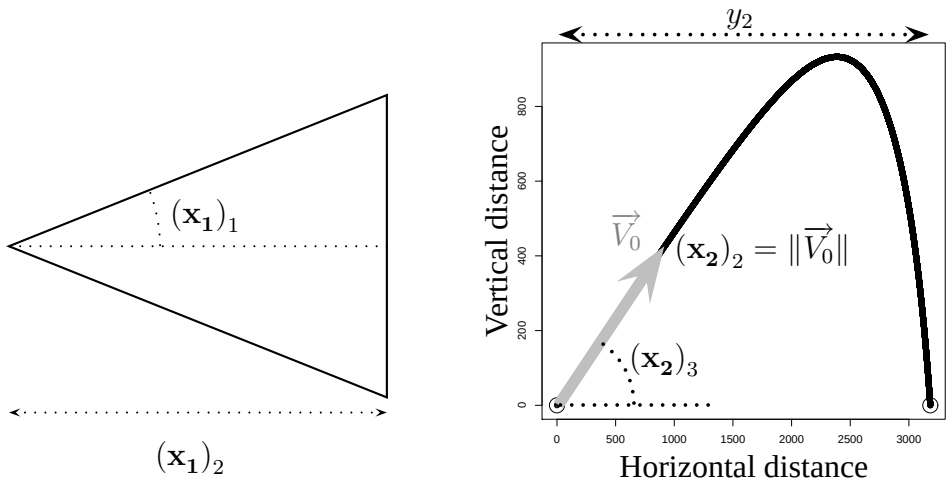


(e) Nested phenomenon : computer model (initial process)



(f) Nested phenomenon : estimator (conditional process)

Figure 1: The figures concern the analytical example. On the figures in the left-hand column the lines represent the computer models alone and the grey area the 95% prediction intervals associated with their errors ϵ_1, ϵ_2 or ϵ_c. On the figures in the right-hand column the lines represent the predictions of the studied phenomena given 7 equally spaced observation points on $\mathcal{D}_{d_1}$, $\mathcal{D}_{d_2}$ and $\mathcal{D}_{d_c}$ and their errors. In both columns the dots represent observations of the true phenomena that were not used to build the predictor. The covariance functions of the model errors are Matérn $\frac{5}{2}$.



(a) Phenomenon 1 : drag coefficient / height of the cone

(b) Phenomenon 2 : distance of a ballistic trajectory

Figure 2: Characteristics of the two phenomena in the hydrodynamic example

The input space is defined as follows :

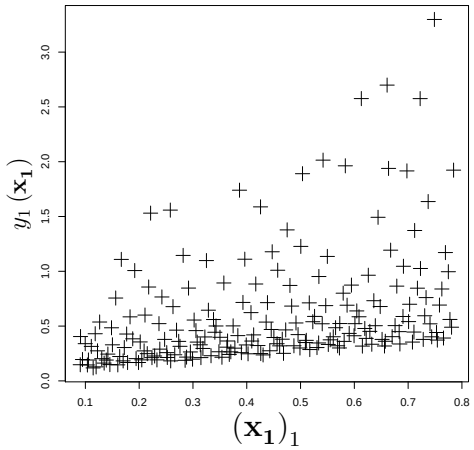
$$\begin{aligned} \mathcal{D}_{d_1} &= [5°, 45°] \times [0.2, 2]\,, \\ \mathcal{D}_{d_{c2}} &= [1500, 3000] \times [15°, 35°]\,, \\ \mathcal{D}_{d_2} &= y_1\left(\mathcal{D}_{d_1}\right) \times \mathcal{D}_{d_{c2}}, \\ \mathcal{D}_{d_c} &= \mathcal{D}_{d_1} \times \mathcal{D}_{d_{c2}}, \end{aligned} \tag{53}$$

Figures 3 presents the evolution of the outputs of the phenomena 1 and 2 with respect to their inputs. The sensitivity of the inputs is variable. In particular the output of the phenomenon 2 is very sensitive to its first input.

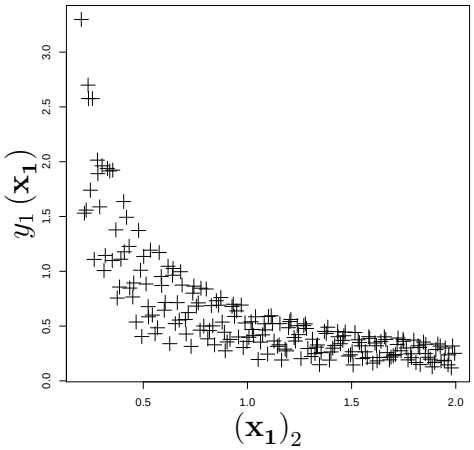
The two phenomena, which are non-linear computer codes, are approximated by analytic functions. The analytic approximations of the phenomena 1 and 2 are defined by their basis functions $\boldsymbol{h_1}\left(\mathbf{x_1}\right)$ and $\boldsymbol{h_2}\left(\mathbf{x_2}\right)$ and the associated optimal values of the parameters $\boldsymbol{\beta_1^*}$ and $\boldsymbol{\beta_2^*}$. On the figure 3 the graphs enable to define $\boldsymbol{h_1}\left(\mathbf{x_1}\right)$ and $\boldsymbol{h_2}\left(\mathbf{x_2}\right)$. The optimal values $\boldsymbol{\beta_1^*}$ and $\boldsymbol{\beta_2^*}$ of the parameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are obtained by maximum likelihood estimation on two sets of 200 observation points drawn according to Latin Hypercube Designs on $\mathcal{D}_{d_1}$ and $\mathcal{D}_{d_2}$, considering a nugget covariance function. Table 2 summarizes the inputs of the two phenomena and of the analytic approximations of these phenomena.

5.5 Results

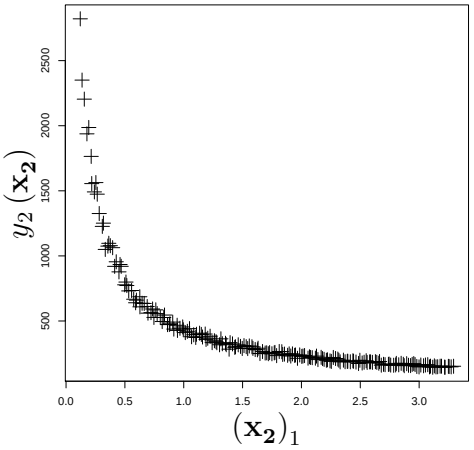
Figure 4 compares the parallel methods with the linearized methods in the case of observation points drawn according to chained LHD. Regarding the calibration's performance the parallel



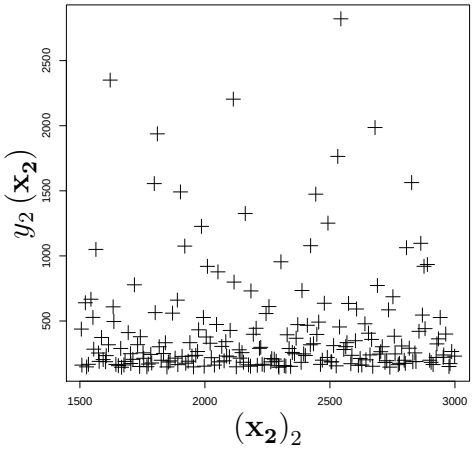
(a) Phenomenon 1 : half-angle of the cone



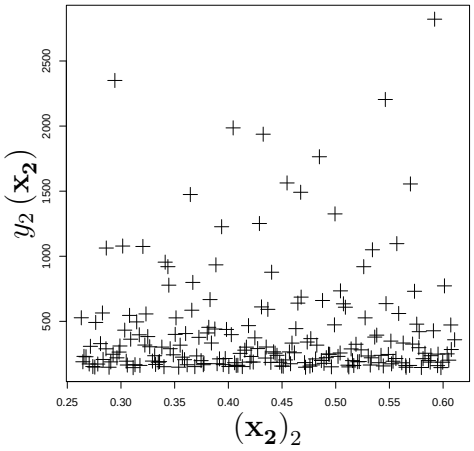
(b) Phenomenon 1 : height of the cone



(c) Phenomenon 2 : output of the phenomenon 1



(d) Phenomenon 2 : initial velocity

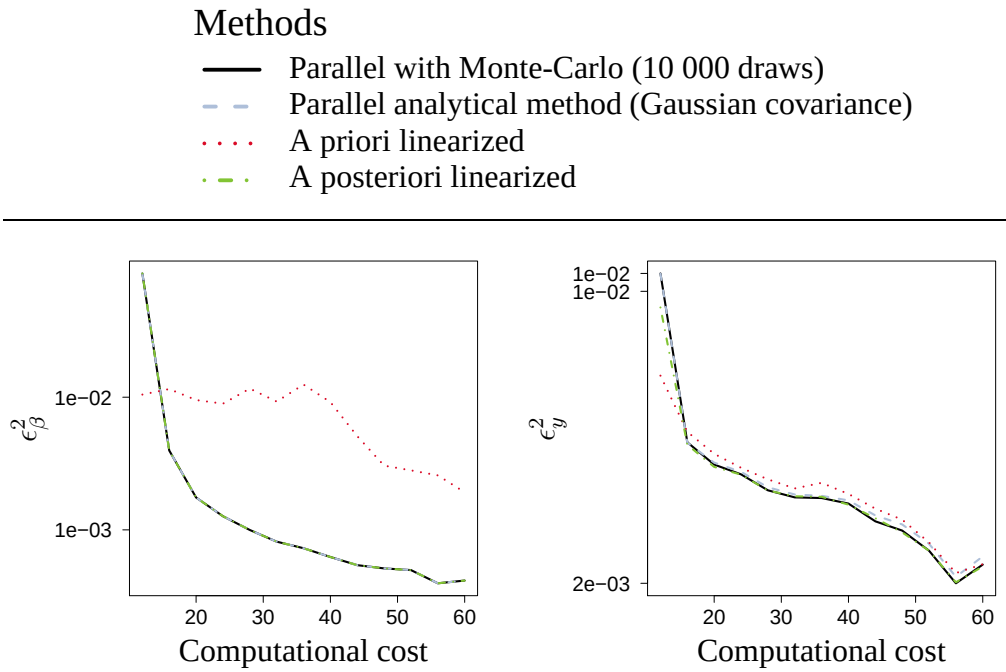


(e) Phenomenon 2 : initial angle from the horizontal

Figure 3: The figures represent the variation of the outputs of the phenomena 1 and 2 with respect to their inputs for sets of observation points drawn according to two Latin Hypercube Designs of 200 points on $\mathcal{D}_{d_1}$ and $\mathcal{D}_{d_2}$.

Phenomenon	*Input variables*	*Analytic approximation*
Phenomenon 1	• $(\mathbf{x_1})_1$: Half-angle of the cone • $(\mathbf{x_1})_2$: Height of the cone	• $\boldsymbol{h_1}(\mathbf{x_1}) = \begin{bmatrix} 1 \\ (\mathbf{x_1})_1 \\ \frac{1}{(\mathbf{x_1})_2} \end{bmatrix}$, • $\boldsymbol{\beta_1^*} = \begin{bmatrix} -0.3738757 \\ 0.9035255 \\ 0.4554474 \end{bmatrix}$.
Phenomenon 2	• $(\mathbf{x_2})_1$: Output of the phenomenon 1 • $(\mathbf{x_2})_2$: Initial velocity • $(\mathbf{x_2})_3$: Initial angle to the horizontal	• $\boldsymbol{h_2}(\mathbf{x_2}) = \begin{bmatrix} 1 \\ \frac{1}{\max\left((\mathbf{x_2})_1, y_{1_0}\right)} \end{bmatrix}$ where y_{1_0} is the minimal value of the set of observations of the phenomenon 1, • $\boldsymbol{\beta_2^*} = \begin{bmatrix} 62.99668 \\ 352.94021 \end{bmatrix}$.

Table 2: Input of the two studied phenomena and polynomial approximations of these phenomena



(a) Analytic example : Calibration

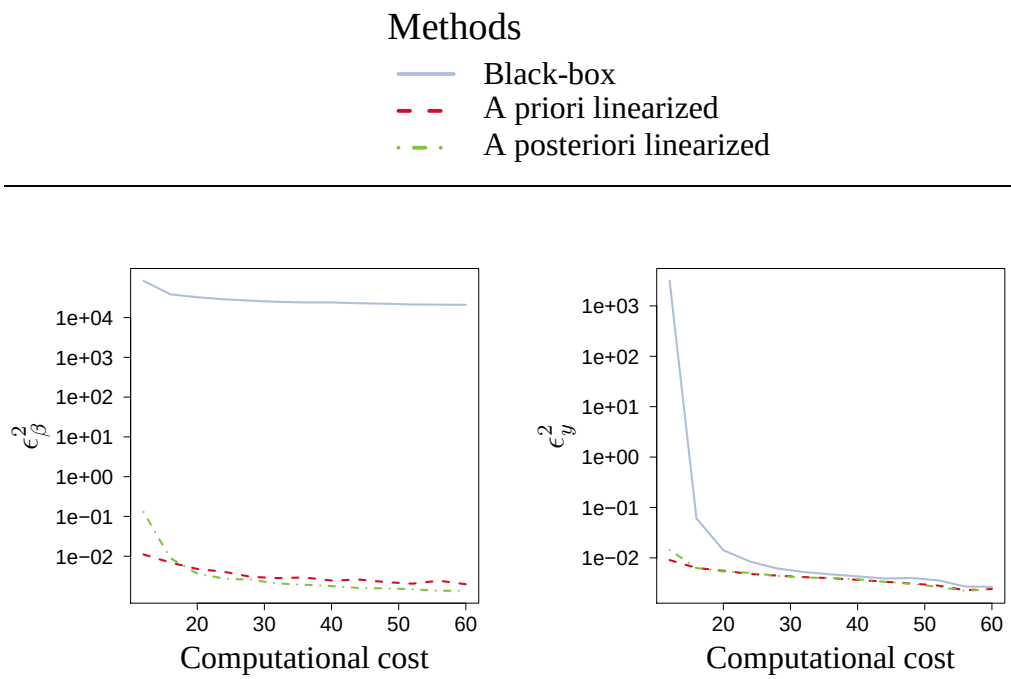
(b) Analytic example : Prediction

Figure 4: The figures represent the performance of calibration and prediction for the previously presented approaches (parallel with Monte-Carlo, parallel with analytical formula, a priori and a posteriori linearized). The observations are drawn according to chained Latin Hypercube designs (LHS) of increasing size. The observations' choice is repeated 50 times. The covariance functions of ϵ_1 and ϵ_2 are Gaussian.

methods and the a posteriori linearized method give very similar results, which are very good when the number of observations increases. In contrast with the other methods, the calibration's error of the a priori linearized method slowly decreases with the number of observations. Concerning the prediction's performance the different methods give similar results. Once again the evolution is slightly different for the a priori linearized method.

Figure 5 presents the performance of calibration and prediction for the linearized methods and the black-box method in the two studied examples. Regarding the calibration, the black-box method performs poorly compared to the a posteriori linearized method. In the analytic example the calibration's performance of the two linearized methods are similar, whereas in the hydrodynamic example the a priori linearized method performs poorly to calibrate the parameters. This emphasizes the interest of the a posteriori linearization. In both examples the prediction's performance of the linearized methods is relatively similar and the black-box method performs poorly when the number of observations is low and tends to the linearized methods when the number of observations increases.

Figure 6 presents an example of predictors of the nested phenomenon of the analytical example for the black-box and the a posteriori linearized methods with 8 and 20 observations . Concerning the mean of the predictor :



(a) Analytic example : Calibration

(b) Analytic example : Prediction

(c) Hydrodynamic example : Calibration

(d) Hydrodynamic example : Prediction

Figure 5: The figures represent the performance of calibration and prediction for the linearized approaches and the existing reference (black-box). The observation points are drawn according to chained Latin Hypercube Designs of increasing size. The observation points' draw is repeated 50 times. The black-box method takes into account only the observations of the nested phenomenon. The covariance functions of ϵ_1, ϵ_2 and ϵ_c are Matérn $\frac{5}{2}$.

- the mean values are very close for the black-box and the a posteriori linearized method,
- for both methods the mean is more accurate when the number of observations increases.

It can be seen that because of the stationarity of the Gaussian process modeling the nested phenomenon in the black-box method, which means that the covariance function of the process depends only on the distance between two inputs, the 95% prediction interval is the same between each regularly spaced observation points. On the contrary, with the a posteriori linearized method the 95% prediction interval between each regularly spaced observation points depends on the value of $\mathbf{x_c}$ and appears to be more suitable to the variations of the nested phenomenon. The 95% prediction interval obtained with the a posteriori linearized method is more accurate than the one obtained with the black-box method. The hypothesis of stationarity of the nested phenomenon is not valid, so in the black-box method the 95% prediction interval is too small or too high depending on the value of $\mathbf{x_c}$. This is particularly true when looking at the figures associated with the 20 observations.

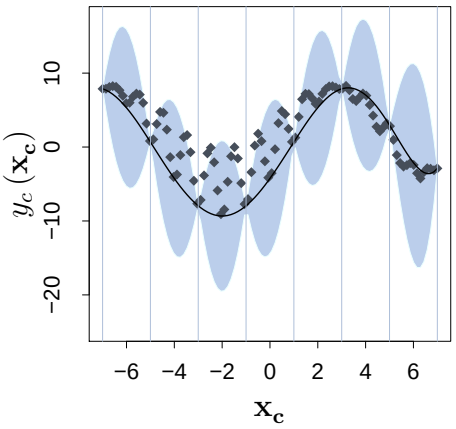
According to the previously presented results the a posteriori linearized method is very efficient to calibrate the parameters and build a predictor of the nested phenomenon. In what follows we will now study methods of sequential designs for the a posteriori linearized method.

Figure 7 presents the performance of calibration and prediction for the chained Latin Hypercube Design (LHS) and the sequential designs : partial A-optimal, partial I-optimal, A-optimal and I-optimal. The A-optimal sequential design is particularly efficient to calibrate the parameters in both examples. The I-optimal sequential design performs the best to build a predictor of the nested phenomenon in both examples. In both cases the total criteria (A-optimal and I-optimal) perform better than the partial criteria (partial A-optimal and partial I-optimal). The total criteria are not adapted to the parallel working but they enable to optimize the computational budget by choosing adequately on which phenomenon to add a new observation point.

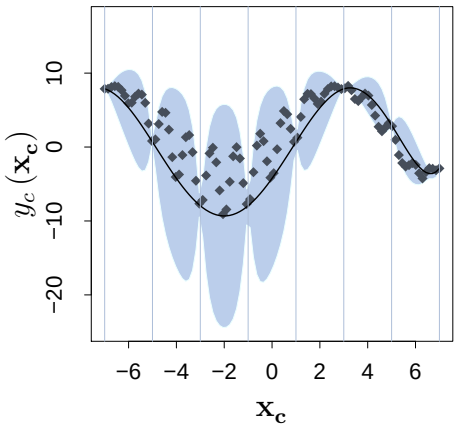
Figure 8 presents the number of observations of the phenomena 1 and 2 at a given computational cost for the different studied designs. It can be seen that these numbers are very different for the sequential design adapted for calibration (A-optimal) and the one for prediction (I-optimal). In the analytic example, the sequential I-optimal design leads to a design where there are ten times more new observations of the phenomenon 1 than new observations of the phenomenon 2. The sequential A-optimal design leads to a design where there are almost as many new observations of the phenomenon 1 as new observations of the phenomenon 2.
In the hydrodynamic example, the sequential I-optimal design leads to a design where there are five times more new observations of the phenomenon 1 than new observations of the phenomenon 2. The sequential A-optimal design leads to a design where there are new observations of the phenomenon 2 only.

6 Conclusions and further work

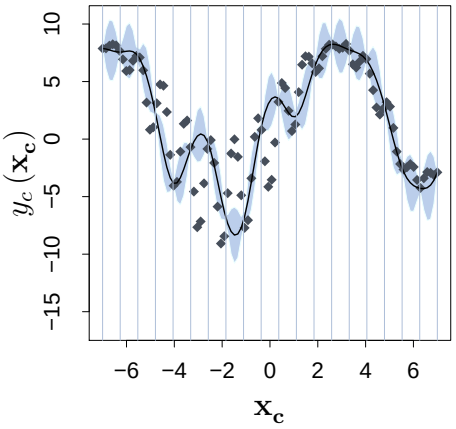
In this paper, methods have been proposed to calibrate the parameters and build a predictor of a nested phenomenon whatever the type of available observations among the phenomena 1, 2 and nested. The proposed methods enable to estimate analytically the moments of order 1 and 2 of



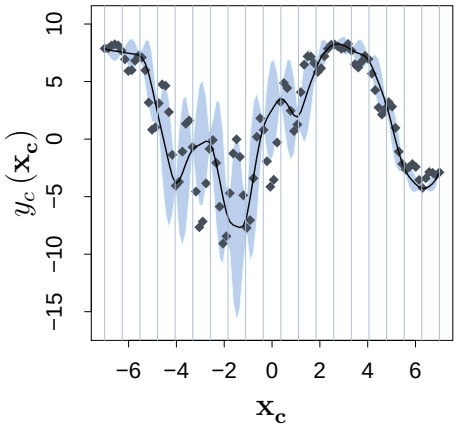
(a) 8 observations : Black-box



(b) 8 observations : A posteriori linearized

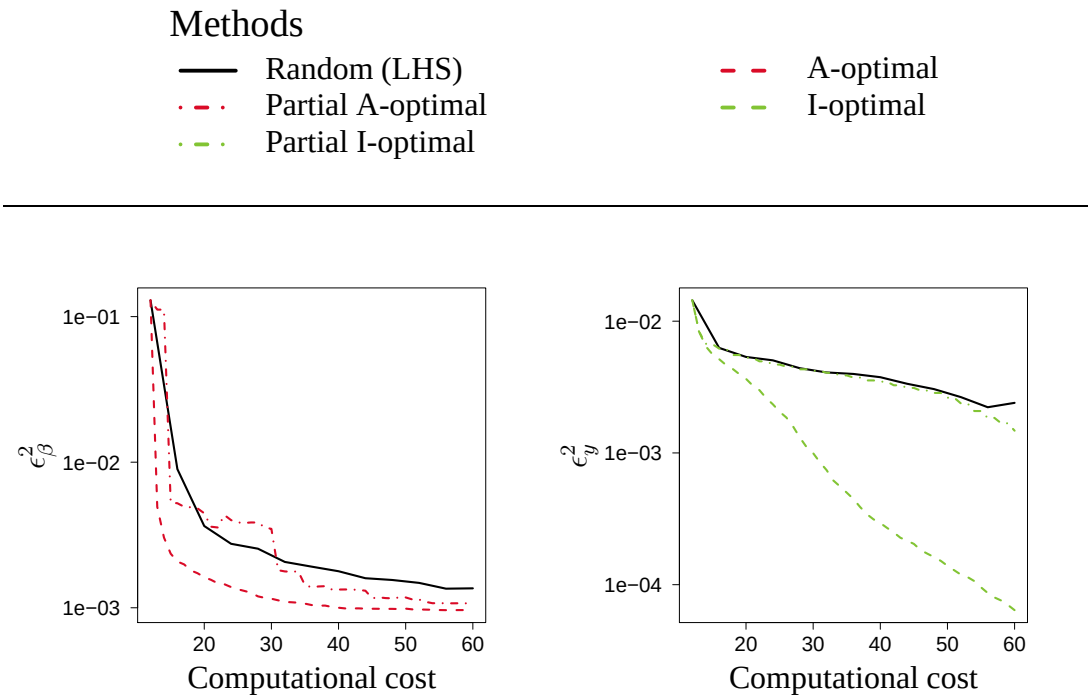


(c) 20 observations : Black-box



(d) 20 observations : A posteriori linearized

Figure 6: The figures present examples of the predictor of the nested phenomenon of the analytic example for the black-box and the a posteriori linearized method. The predictors are built with 8 and 20 equally spaced observation points on $\mathcal{D}_{d_c}$. The observation points used to build the predictors are the same for the two methods and are represented by vertical lines. The dots represent the real values of the nested phenomenon. The black line represents the mean of the predictor. The grey area represents the 95% prediction interval of the predictor. The covariance functions of ϵ_1 and ϵ_2 for the a posteriori linearized method, ϵ_c for the black-box method are Matérn $\frac{5}{2}$.



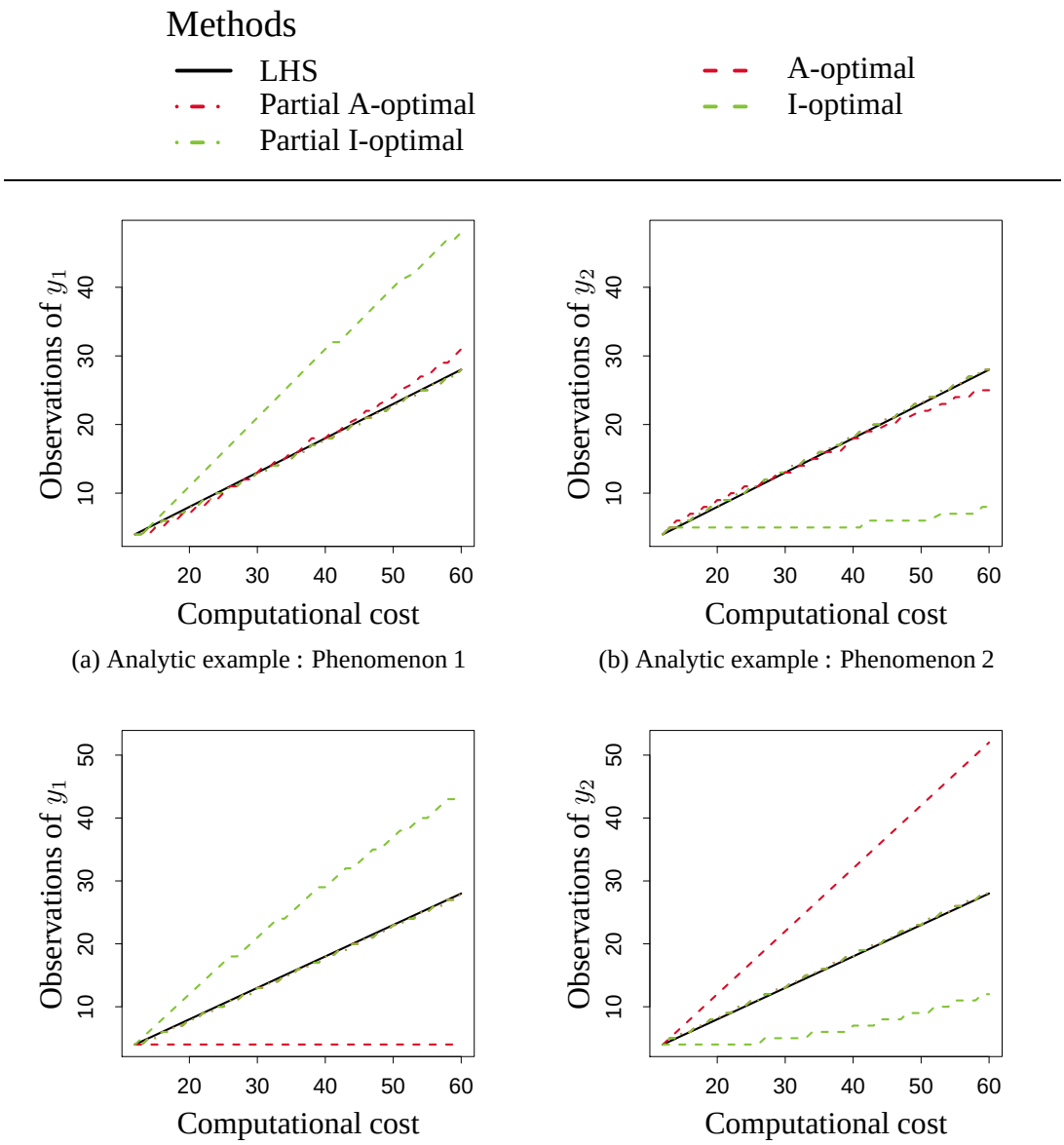
(a) Analytic example : Calibration

(b) Analytic example : Prediction

(c) Hydrodynamic example : Calibration

(d) Hydrodynamic example : Prediction

Figure 7: The figures represent the performance of calibration and prediction for the a posteriori linearized method with different types of designs. The initial designs of the sequential approaches are drawn according to chained LHD. The observation points' choice is repeated 50 times. The covariance functions are Matérn $\frac{5}{2}$.



(a) Analytic example : Phenomenon 1

(b) Analytic example : Phenomenon 2

(c) Hydrodynamic example : Phenomenon 1

(d) Hydrodynamic example : Phenomenon 2

Figure 8: The figures represent the number of observations of the phenomena 1 and 2 for the different studied designs.

all the phenomena.
The methods have been compared to the case where the nested phenomenon is considered as a unique phenomenon. The results obtained in the examples demonstrate the interest of using all the available information, that means not only the observations of the nested phenomenon but also the observations of the phenomena 1 or 2 alone. This is particularly significant for the calibration of the parameters and for the prediction when there are few observation points.
Among the presented methods, the a posteriori linearized method brings the best results for a given set of observations.

The choice of the observation points impacts the performance of calibration and prediction. Thus, methods to enrich step by step the set of observations have been studied. The sequential designs presented are criterion-based. They can be based on the performance of calibration or of prediction. They are compared to a reference design which is a maximin LHD on the input space of the nested phenomenon. In all cases the global criterion-based designs perform better than the reference design. The fact of taking into account the joint distribution of the Gaussian processes modeling the nested phenomenon and the phenomena 1 and 2 enables to optimize the computational budget by adding new observation points on the phenomena 1 and 2 alone while improving the performance of calibration or of prediction of the nested phenomenon.

The presented calibration and prediction methods take into account a non informative prior distribution for the parameters. However these methods can easily be extended to the case of a Gaussian prior distribution for the parameters.

In this paper, the case of two phenomena coupled by a scalar has been studied. However the case of a coupling by a functional seems promising for future work.

A Proof of the theorem

Proof. According to Eq. (4), the mean of the predictor of the phenomenon 2 can be written :

$$\mu_2^{(N_2)}\left(\mathbf{x_2}\right) = \boldsymbol{h_2}\left(\mathbf{x_2}\right)^t \boldsymbol{\mu}_{\boldsymbol{\beta_2}}^{(N_2)} + \boldsymbol{r_2}^{(N_2)}\left(\mathbf{x_2}\right)^t \boldsymbol{v}_2, \tag{54}$$

where :

- $\boldsymbol{r_2}^{(N_2)}\left(\mathbf{x_2}\right) = C_2\left(\mathbb{X}_2^{(N_2)}, \mathbf{x_2}\right)$,
- $\boldsymbol{v_2} = \mathbb{Y}_2^{(N_2)} - \boldsymbol{h_2}\left(\mathbb{X}_2^{(N_2)}\right)^t \boldsymbol{\mu}_{\boldsymbol{\beta_2}}^{(N_2)}$.

According to Eq. (20) and (21), we have :

$$\begin{aligned}
\left(\boldsymbol{r_2}^{(N_2)}\left(\mathbf{x_2}\right)\right)_j &= C_2\left(\mathbf{x_2}^{(j)}, \mathbf{x_2}\right) \\
&= g\left(\left(\mathbf{x_2}\right)_1; 0, \frac{2\left(\mathbf{x_2}^{(j)}\right)_1}{l_2^2}, -\frac{1}{l_2^2}\right) \exp\left(-\left(\frac{\left(\mathbf{x_2}^{(j)}\right)_1}{l_2}\right)^2\right) C_{\bar{2}}\left(\left(\mathbf{x_2}^{(j)}\right)_{-1}, \left(\mathbf{x_2}\right)_{-1}\right), \\
\left(\boldsymbol{h_2}\left(\mathbf{x_2}\right)\right)_i &= g\left(\left(\mathbf{x_2}\right)_1; j_i, a_i, b_i\right) m_i\left(\left(\mathbf{x_2}\right)_{-1}\right),
\end{aligned} \tag{55}$$

We can infer from Eq. (54) and (55) that $\mu_2^{(N_2)}\left(\mathbf{x_2}\right)$ can be written as a linear combination of functions g :

$$\mu_2^{(N_2)}\left(\mathbf{x_2}\right) = \sum_{i=1}^{N_2+p_2} g\left(\left(\mathbf{x_2}\right)_1; j_i, a_i, b_i\right) c_i\left(\left(\mathbf{x_2}\right)_{-1}\right) \tag{56}$$

where j_i, a_i, b_i are deterministic known coefficients and c_i deterministic functions.
According to Eq. (4), it can be written :

$$\begin{aligned}
\mu_2^{(N_2)}\left(\mathbf{x_2}\right)^2 + \sigma_2^{(N_2)}\left(\mathbf{x_2}\right)^2 = \; & C_2\left(\mathbf{x_2}, \mathbf{x_2}\right) + \boldsymbol{h_2}\left(\mathbf{x_2}\right)^t \boldsymbol{M_2} \boldsymbol{h_2}\left(\mathbf{x_2}\right), \\
& + 2\boldsymbol{h_2}\left(\mathbf{x_2}\right)^t \boldsymbol{N_2} \boldsymbol{r_2}^{(N_2)}\left(\mathbf{x_2}\right) + \boldsymbol{r_2}^{(N_2)}\left(\mathbf{x_2}\right)^t \boldsymbol{P_2} \boldsymbol{r_2}^{(N_2)}\left(\mathbf{x_2}\right),
\end{aligned} \tag{57}$$

where :

- $\boldsymbol{M_2} = \boldsymbol{\mu}_{\boldsymbol{\beta_2}}^{(N_2)}\left(\boldsymbol{\mu}_{\boldsymbol{\beta_2}}^{(N_2)}\right)^t + \boldsymbol{V}_{\boldsymbol{\beta_2}}^{(N_2)}$,
- $\boldsymbol{N_2} = \boldsymbol{\mu}_{\boldsymbol{\beta_2}}^{(N_2)} \boldsymbol{v_2}^t - \boldsymbol{V}_{\boldsymbol{\beta_2}}^{(N_2)} \boldsymbol{h_2}\left(\mathbb{X}_2^{(N_2)}\right)^t C_2\left(\mathbb{X}_2^{(N_2)}, \mathbb{X}_2^{(N_2)}\right)^{-1}$,
- $\begin{aligned}\boldsymbol{P_2} = \; & \boldsymbol{v_2}\boldsymbol{v_2}^t - C_2\left(\mathbb{X}_2^{(N_2)}, \mathbb{X}_2^{(N_2)}\right)^{-1} \\ & + C_2\left(\mathbb{X}_2^{(N_2)}, \mathbb{X}_2^{(N_2)}\right)^{-1} \boldsymbol{h_2}\left(\mathbb{X}_2^{(N_2)}\right) \boldsymbol{V}_{\boldsymbol{\beta_2}}^{(N_2)} \boldsymbol{h_2}\left(\mathbb{X}_2^{(N_2)}\right)^t C_2\left(\mathbb{X}_2^{(N_2)}, \mathbb{X}_2^{(N_2)}\right)^{-1}.\end{aligned}$

From Eq. (57) and (55) and the fact that for all $x, j_m, j_n, a_m, a_n, b_m, b_n,$ in $\mathbb{R}$

$$g\left(x; j_m, a_m, b_m\right) g\left(x; j_n, a_n, b_n\right) = g\left(x; j_m + j_n, a_m + a_n, b_m + b_n\right)$$

it can be inferred that $\mu_2^{(N_2)}\left(\mathbf{x_2}\right)^2 + \sigma_2^{(N_2)}\left(\mathbf{x_2}\right)^2$ can be written as a linear combination of functions g :

$$\mu_2^{(N_2)}\left(\mathbf{x_2}\right)^2 + \sigma_2^{(N_2)}\left(\mathbf{x_2}\right)^2 = \sum_{k=1}^{(N_2+p_2)^2+1} g\left(\left(\mathbf{x_2}\right)_1; j_k, a_k, b_k\right) c_k\left(\left(\mathbf{x_2}\right)_{-1}\right) \tag{58}$$

where j_k, a_k, b_k are deterministic known coefficients and c_k are deterministic functions. Furthermore, if $(\mathbf{x_2})_1 \sim \mathcal{N}\left(\mu_1^{(N_1)}(\mathbf{x_1}),\ \sigma_1^{(N_1)}(\mathbf{x_1})^2\right)$, the mean of $g((\mathbf{x_2})_1; j, a, b)$ can be calculated analytically conditionally to an integration criterion :

$$\mathbb{E}_u\left[g\left(\mu_1^{(N_1)}(\mathbf{x_1}) + \sigma_1^{(N_1)}(\mathbf{x_1})\, u; j, a, b\right)\right] = \exp\left(a\mu_1^{(N_1)}(\mathbf{x_1}) + b\mu_1^{(N_1)}(\mathbf{x_1})^2 + \frac{1}{2}\frac{\mathbb{E}[X_g]^2}{\mathbb{V}[X_g]}\right) \sum_{k=0}^{j} \binom{j}{k} \mu_1^{(N_1)}(\mathbf{x_1})^{j-k}\, \sigma_1^{(N_1)}(\mathbf{x_1})^k\, \mathbb{E}\left[X_g^k\right], \tag{59}$$

where

- $X_g \sim \mathcal{N}\left(\frac{\left(a+2b\mu_1^{(N_1)}(\mathbf{x_1})\right)\sigma_1^{(N_1)}(\mathbf{x_1})}{1-2b\sigma_1^{(N_1)}(\mathbf{x_1})^2},\ \frac{1}{1-2b\sigma_1^{(N_1)}(\mathbf{x_1})^2}\right)$,
- under the condition of integrability $b < \frac{1}{2\sigma_1^{(N_1)}(\mathbf{x_1})^2}$ that has to be verified.

It can be inferred from Eq. (18), (19), (56), (58) and (59) that, if the process modeling the phenomenon 2 has a covariance function which presents a Gaussian expression with respect to $(\mathbf{x_2})_1$ and has a trend which is a linear combination of products of polynomials by exponentials of order 1 and 2, then the moments of order 1 and 2 of the coupled predictor can be calculated analytically conditionally to some integration criteria.

■

REFERENCES

[1] F. Bachoc. Calibration and improved prediction of computer models by universal kriging. *Nuclear Science and Engineering*, 176:81–97, 2014.

[2] François Bachoc. *Parametric estimation of covariance function in Gaussian-process based Kriging models. Application to uncertainty quantification for computer experiments*. PhD thesis, Université Paris-Diderot - Paris VII, 2013.

[3] Christopher T. H. Baker. *The numerical treatment of integral equations*. Clarendon Press, Oxford, 1977.

[4] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Stat. Comput.*, 22:773–793, 2012.

[5] James O. Berger, José M. Bernardo, and DongChu Sun. The formal definition of reference priors. *The Annals of Statistics*, 37(2):905–938, 2009.

[6] James O. Berger, Victor De Oliveira, and Bruno Sansó. Objective bayesian analysis of spatially corellated data. *Journal of the American Statistical Association*, 96(456):1361–1374, 2001.

[7] Geraud Blatman and Bruno Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230:2345–2367, 2011.

[8] B. Echard, N. Gayton, and M. Lemaire. AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Struct. Saf.*, 33:145–154, 2011.

[9] David Ginsbourger, Rodolphe Le Riche, and Laurent Carraro. *Computational Intelligence in Expensive Optimization Problems*, volume 2 of *Adaptation Learning and Optimization*, chapter Kriging Is Well-Suited to Parallelize Optimization, pages 131–162. Springer Berlin Heidelberg, 2010.

[10] Marc C Kennedy and Anthony O'Hagan. Predicting the output from a complex computer code when fast approximations are avalaible. *Biometrika*, 87:1–13, 2000.

[11] Marc C Kennedy and Anthony O'Hagan. Bayesian Calibration of Computer Models. *J. R. Stat. Soc. . Ser. B (Statistical Methodol.*, 63(3):425–464, 2001.

[12] Rui Paulo. Default priors for gaussian processes. *Annals of Statistics*, 33(2):556–582, 2005.

[13] Guillaume Perrin, Christian Soize, Denis Duhamel, and Christine Funfschilling. A posteriori error and optimal reduced basis for stochatisc processes defined by a finite set of realizations. *SIAM ASA Journal of Uncertainty Quantification*, 2:745–762, 2014.

[14] Carl Edward Rasmussen and Christopher K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, 2006.

[15] Christian Robert. *The Bayesian Choice*. Springer-Verlag New York, New York, 2007.

[16] J. Sacks, W. Welch, T. J. Mitchell, and H. P. Wynn. Design and Analysis of Computer Experiments. *Stat. Sci.*, 4:409–435, 1989.

[17] Thomas J. Santner, Brian J. Williams, and William Notz. *The design and analysis of computer experiments*. Springer series in statistics. Springer, New York, 2003.