International Journal for Uncertainty Quantification, 15(6):37–59 (2025)

BAYESIAN CALIBRATION FOR PREDICTION IN A MULTI-OUTPUT TRANSPOSITION CONTEXT

Charlie Sire,[1,2,*] Josselin Garnier,[2] Baptiste Kerleguer,[3] Céric Durantin,[3] Gilles Defaux,[3] & Guillaume Perrin[4]

[1]*Inria Saclay Centre, Palaiseau, France*

[2]*CMAP, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France*

[3]*CEA, DAM, DIF, F-91297 Arpajon, France*

[4]*COSYS, Universite Gustave Eiffel, Marne-La-Vallée, France*

*Address all correspondence to: Charlie Sire, Inria Saclay Centre, Palaiseau, France,
E-mail: sire.charlie971@gmail.com

Numerical simulations are widely used to predict the behavior of physical systems, with Bayesian approaches being particularly well suited for this purpose. However, experimental observations are necessary to calibrate certain simulator parameters for the prediction. In this work, we use a multi-output simulator to predict all its outputs, including those that have never been experimentally observed. This situation is referred to as the transposition context. To accurately quantify the discrepancy between model outputs and real data in this context, conventional methods cannot be applied, and the Bayesian calibration must be augmented by incorporating a joint model error across all outputs. To achieve this, the proposed method is to consider additional input parameters within a hierarchical Bayesian model, which includes hyperparameters for the prior distribution of the calibration variables. This approach is applied to a computer code with three outputs that models the Taylor cylinder impact test with a small number of observations. The outputs are considered as the observed variables one at a time, to work with three different transposition situations. The proposed method is compared with other approaches that embed model errors to demonstrate the significance of the hierarchical formulation.

KEY WORDS: *Bayesian inference, code transposition, hierarchical model, model error, model calibration, uncertainty quantification, Markov chain Monte carlo*

1. INTRODUCTION

Numerical simulations are increasingly used to understand and optimize complex physical systems, as they enable the prediction of behavior in previously unobserved scenarios. For accurate predictions, the input parameters of these simulations must be carefully specified. Some of the simulation inputs, known as control variables, replicate experimental conditions chosen by the experimenter and are denoted by $\mathbf{x}$, while others are unknown parameters specific to the computer code, denoted by $\boldsymbol{\lambda}$. These parameters typically have a physical significance and must be calibrated to ensure the code predictions closely match the actual system behavior, observed from limited experimental data. The parameters $\boldsymbol{\lambda}$ can also be numerical parameters that are needed to run the code. The simplest approach involves performing an optimization scheme on $\boldsymbol{\lambda}$ to minimize a given distance between observed data and model outputs, see [1–3], then using the optimal values as if they were known. However, this method assumes no uncertainty in the observed data (no measurement error) and in the code prediction (no model error), which is unrealistic in practice.

A Bayesian methodology that considers noisy observations and treats the unknown parameters as a random vector $\boldsymbol{\Lambda}$ can be adopted. The prior distribution of $\boldsymbol{\Lambda}$ and the observations are used to build the posterior distribution of $\boldsymbol{\Lambda}$, which captures the uncertainty and allows for a predictive distribution of the output. This framework is thoroughly presented in [4] and further studied in numerous works, for example [5–7].

Nevertheless, predicting the behavior of physical systems using numerical models is never perfect, even with well-identified parameter values. There is always a discrepancy between the model outputs and the observed data that

Submitted: 9/30/24; Accepted: 6/13/25; Online: 8/5/25 https://doi.org/10.1615/Int.J.UncertaintyQuantification.2025056586

must be accounted for as a model error, as highlighted in various works, see [8–11]. Typically, as detailed in [4], the model error term can be explicitly represented as an additive Gaussian process, whose distribution hyperparameters are estimated from the observed data. This approach is particularly useful but cannot be implemented for prediction in some situations due to the impossible estimation of the hyperparameters. These situations arise when the output of interest does not directly correspond to the experimental data, such as when calibration must be performed based on different experiments not directly suited to the objective. The framework introduced in this paper will be called transposition. Four cases can be identified:

1. Change of scale: The calibration domain is different from the domain of interest of the code.
2. New inputs: The code is calibrated with fewer inputs than it will use.
3. New code: The calibration parameters will be used in another code.
4. Unobserved outputs: The calibration and prediction will be carried out by observing some outputs, but not all of them.

In this paper, only the case of predicting unobserved outputs will be considered, although the presented method is general enough to be adapted to many transposition scenarios.

A collection of T different output variables of interest is identified and denoted by $\boldsymbol{y}$. It depends on control variables $\mathbf{x} \in \mathcal{X}$, with $\mathcal{X}$ a subset of $\mathbb{R}^s, s \in \mathbb{N}^\star$, such that

$$\boldsymbol{y}(\mathbf{x}) = (y_t(\mathbf{x}))_{t=1}^T \in \mathbb{R}^T.$$

Let $\boldsymbol{f}$ be a numerical code to compute this T-dimensional output. Its input variables are $\mathbf{x}$ and p unknown physical calibration parameters $\boldsymbol{\lambda} = (\lambda_i)_{i=1}^p \in \mathcal{L} \subset \mathbb{R}^p$, and

$$\boldsymbol{f}(\mathbf{x}, \boldsymbol{\lambda}) = (f_t(\mathbf{x}, \boldsymbol{\lambda}))_{t=1}^T \in \mathbb{R}^T.$$

The experimental phase is assumed to provide n observations but only for the variable t', where $1 \leq t' \leq T$. They form the vector $\boldsymbol{y}_{\boldsymbol{t'},\mathbf{obs}} = (y_{t',\text{obs},j})_{j=1}^n$, which we consider as realizations of noisy measurements of $y_{t'}$ at design points $\mathbb{X} = (\mathbf{x}_j)_{j=1}^n$ with output measurement error $\boldsymbol{E} = (E_j)_{j=1}^n$:

$$\boldsymbol{Y}_{\boldsymbol{t'},\mathbf{obs}} = (Y_{t',\text{obs},j})_{j=1}^n = (y_{t'}(\mathbf{x}_j) + E_j)_{j=1}^n \in \mathbb{R}^n. \tag{1}$$

The probability density function (pdf) of the measurement error is known and is denoted by $p_{\boldsymbol{E}}$. The transposition comes from the fact that only the variable $y_{t'}$ is measured in the experimental phase, while the objective is to predict $y_t(\mathbf{x}_0)$ for $1 \leq t \leq T$, and $\mathbf{x}_0 \in \mathcal{X} \setminus \mathbb{X}$. In the following, and without loss of generality, we assume that the variable y_1 is measured (i.e., $t' = 1$) considering the random vector $\boldsymbol{Y}_{\mathbf{1},\text{obs}} = (Y_{1,\text{obs},j})_{j=1}^n = (y_1(\mathbf{x}_j) + E_j)_{j=1}^n$. Note that in the proposed framework, the number of experiments n is assumed to be small (e.g., $n = 10$ in the application presented in Section 3). This limited sample size makes prediction at $\mathbf{x}_0$ more challenging. However, this constraint is often unavoidable due to the financial, material, or time-related costs associated with each physical experiment.

The common additive discrepancy to represent the model error would be introduced as follows, by modeling the output variables of interest as $\boldsymbol{Y}(\mathbf{x}) = (Y_t(\mathbf{x}))_{t=1}^T = \boldsymbol{f}(\mathbf{x}, \boldsymbol{\Lambda}) + \boldsymbol{\Delta}(\mathbf{x})$, with $\boldsymbol{\Lambda} \in \mathbb{R}^p$ a random vector of pdf $p_{\boldsymbol{\Lambda}}$ and $\boldsymbol{\Delta}$ a T-variate Gaussian process with prior information depending on some hyperparameters ϕ. In our transposition context, some of these hyperparameters are impossible to estimate. Indeed, let us consider, for instance, $\boldsymbol{\Delta} = (\Delta_t)_{t=1}^T$ pairwise independent Gaussian processes with respective mean $\mu_t(.)$ and covariance $c_t(.,.)$ that depends on hyperparameters ϕ_t. It quickly appears that for every t, experimental measurments of y_t are needed to estimate ϕ_t , but they are not available in this context.

Therefore, the objective is to propose a model error for the prediction of $y_t(\mathbf{x}_0)$ with hyperparameters that can be estimated from the observations $\boldsymbol{Y}_{\mathbf{1},\text{obs}}$. As presented in [12], embedded errors can be employed to build a mathematical relationship between the observations and the variables of interest. A similar approach is described in [13], where an embedding $Y_t(\mathbf{x}) = f_t(\mathbf{x}, \boldsymbol{\Lambda} + \boldsymbol{\Delta})$ is considered with $\boldsymbol{\Delta}$ being a random discrepancy whose parameters

require calibration, as detailed in Appendix C. In our work, an original embedding is proposed by increasing the variability in the calibration variables $\mathbf{\Lambda}$ through the inclusion of $q-p$ additional parameters to represent the model error, and then work with a q-dimensional vector $\mathbf{\Lambda}$. These $q-p$ parameters encompass all parameters that influence the relationship between the controlled inputs and all the T outputs but are not classically intended to be calibrated. These could include, for example, mesh parameters, variables related to the simulation duration, or known physical law parameters. As this study is conducted in the context of a very limited number of observations, prior assumptions play a crucial role. This is why a hierarchical approach is proposed for these additional parameters, allowing for the estimation of the hyperparameters of their prior distribution [14], thereby refining it to better account for model error.

The main contribution of this work lies in this embedded model error for prediction in transposition contexts. A key challenge is the treatment of hyperparameters within the hierarchical framework. Note that this method cannot be effective if the variations in the output due to the fluctuations of the additional parameters are not sufficiently large to capture the model error. This point will be further discussed in Section 3 and Appendix B.

The article is organized as follows: Section 2 outlines the Bayesian methodology tailored to the transposition context and introduces the surrogate model used. Section 3 describes the application test case and the performance metrics examined, followed by numerical results with a comparison of various approaches, including the method from [13]. Finally, Section 4 summarizes the work and proposes possible extensions.

2. BUILDING AN ADAPTED BAYESIAN FRAMEWORK

2.1 Bayesian Calibration

In the Bayesian framework, the idea is to obtain the posterior distribution (i.e., conditioned by the experimental data) of parameters that balances prior knowledge with observed data, see [15]. More precisely, we consider

$$\begin{cases} Y_t(\mathbf{x}) = f_t(\mathbf{x}, \mathbf{\Lambda}), & 1 \leq t \leq T,\ \mathbf{x} \in \mathcal{X}, \\ Y_{1,\text{obs},j} = Y_1(\mathbf{x}_j) + E_j, & 1 \leq j \leq n, \end{cases} \tag{2}$$

with $\mathbf{\Lambda}$ of prior pdf $p_{\mathbf{\Lambda}}$, $\boldsymbol{E}$ of known pdf $p_{\boldsymbol{E}}$, and a small n. As explained in Section 1, here we consider $\mathbf{\Lambda} \in \mathbb{R}^q$, which is made of the p unknown physical parameters and the $q-p$ additional parameters included to emulate the model error. Obviously, the following formulas are valid when considering the classical situation $\mathbf{\Lambda} \in \mathbb{R}^p$.

The goal is to predict $y_t(\mathbf{x}_0)$, which requires investigating the posterior distribution of $Y_t(\mathbf{x}_0)$, i.e., the distribution of $Y_t(\mathbf{x}_0) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}$. This involves estimating $\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ for all functions $\psi : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ is well defined. In particular, the estimations of $\mathbb{E}(Y_t(\mathbf{x}_0) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ and $\mathbb{E}(Y_t(\mathbf{x}_0)^2 \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ are required to determine the predictive mean and standard deviation. In situations where extreme events are studied, functions of the form $\Psi(y) : y \to \mathbf{1}_{[b,+\infty[}$, with $b \in \mathbb{R}$, are considered.

The expectation given the observations can be expressed as

$$\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) = \int_{\mathbb{R}^q} \psi(f_t(\mathbf{x}_0, \boldsymbol{\lambda}))p(\boldsymbol{\lambda} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})d\boldsymbol{\lambda}, \tag{3}$$

with $p(\boldsymbol{\lambda} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ the pdf of the posterior distribution of $\mathbf{\Lambda}$. Bayes' theorem, see [16], gives

$$p(\boldsymbol{\lambda} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) = \frac{p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda})p_{\mathbf{\Lambda}}(\boldsymbol{\lambda})}{\int_{\mathbb{R}^q} p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda}')p_{\mathbf{\Lambda}}(\boldsymbol{\lambda}')d\boldsymbol{\lambda}'},$$

where $p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda}) = p_{\boldsymbol{E}}\big(\boldsymbol{y}_{\mathbf{1},\text{obs}} - (f_1(\mathbf{x}_j, \boldsymbol{\lambda}))_{j=1}^n\big)$ is the likelihood of $\boldsymbol{\lambda}$. A consistent estimator of $\mathbb{E}(\psi(Y_t(\mathbf{x}_0) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ can be

$$\hat{E}_M\Big(\psi(Y_t(\mathbf{x}_0))\Big) = \frac{1}{M}\sum_{k=1}^{M} \psi(f_t(\mathbf{x}_0, \mathbf{\Lambda}_k)), \tag{4}$$

where $(\mathbf{\Lambda}_k)_{k=1}^M$ is sampled with pdf $p(\boldsymbol{\lambda} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) \propto p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda})p_{\mathbf{\Lambda}}(\boldsymbol{\lambda})$. However, the normalizing term $\int_{\mathbb{R}^q} p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda}')p_{\mathbf{\Lambda}}(\boldsymbol{\lambda}')d\boldsymbol{\lambda}'$ is known to be intractable in practice, making it impossible to sample directly from $p(\boldsymbol{\lambda} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ with

classical Monte Carlo schemes. Markov chain Monte Carlo (MCMC) methods are then required here to work with posterior samples $(\mathbf{\Lambda}_k)_{k=1}^M$ with pdf proportional to $p(\boldsymbol{y_1}_{,\text{obs}} \mid \cdot)p_{\mathbf{\Lambda}}$. In our study, these MCMC samples are obtained with the delayed rejection adaptive metropolis (DRAM) algorithm detailed in [17], but other algorithms such as Hamiltonian Monte Carlo (HMC, [18]) could be used, particularly if q is large.

Note that here $p_{\boldsymbol{E}}$ will be considered known in all the methods we implement to ensure a fair comparison between them and focus on the model error. However, in some situations, inferring the parameters of $p_{\boldsymbol{E}}$ is crucial. For instance, if $\boldsymbol{E} \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$ with unknown σ_ε, then a solution is to consider σ_ε as a hyperparameter to be inferred together with $\boldsymbol{\lambda}$, and then work with the augmented set $\tilde{\boldsymbol{\lambda}} = (\boldsymbol{\lambda}, \sigma_\varepsilon)$ of likelihood $p(\tilde{\boldsymbol{\lambda}} \mid \boldsymbol{y_1}_{,\text{obs}}) = p_{\boldsymbol{E}(\sigma_\varepsilon)}\big(\boldsymbol{y_1}_{,\text{obs}} - (f_1(\mathbf{x}_j, \boldsymbol{\lambda}))_{j=1}^n\big)$, with $p_{\boldsymbol{E}(\sigma_\varepsilon)}$ the density of $\boldsymbol{E}(\sigma_\varepsilon) \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$.

2.2 Hierarchical Model

In the following, $\mathcal{L}_{\text{cal}} \subset \mathbb{R}^p$ is the domain of physical parameters to calibrate, and $\mathcal{L}_{\text{err}} \subset \mathbb{R}^{q-p}$ is the domain of the additional parameters considered to emulate the model error. Then $\boldsymbol{\lambda} \in \mathcal{L}$ is introduced with $\mathcal{L} = \mathcal{L}_{\text{cal}} \times \mathcal{L}_{\text{err}}$. In our study, the $q - p$ parameters in $\mathcal{L}_{\text{err}}$ are integrated with a hierarchical description, see [19] and [20], to better target their correct distribution and account for the model error. This is particularly effective even in contexts where measurement noise is significant. The hierarchical description assumes that the *a priori* distribution of the additional parameters depends on hyperparameters $\boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{R}^r$. The pdf of $\mathbf{\Lambda}$ given $\boldsymbol{\alpha}$ is denoted $p_{\mathbf{\Lambda}}(. \mid \boldsymbol{\alpha})$, and each $\boldsymbol{\alpha}$ drives these parameters for the model error. In a hierarchical description, the objective is to perform Bayesian inference on the hyperparameters $\boldsymbol{\alpha}$ as well and then consider them as a random variable $\mathbf{A} \in \mathcal{A}$ with prior distribution $p_{\mathbf{A}}$. Note that to avoid possible confusion, $\mathbf{A}$ is simply the uppercase notation of $\boldsymbol{\alpha}$ here, as lowercase letters represent deterministic variables, while the uppercase letters are associated with random variables throughout this article. The same convention applies to $\boldsymbol{\lambda}$ and $\mathbf{\Lambda}$, for instance.

As for Eq. (3), Bayes' theorem provides that $p(\boldsymbol{\alpha} \mid \boldsymbol{y_1}_{,\text{obs}})$, the posterior pdf of $\mathbf{A}$ evaluated at $\boldsymbol{\alpha}$, is proportional to $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha})$, where $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ is the likelihood of $\boldsymbol{\alpha}$. This likelihood can be expressed as

$$\begin{aligned} p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}) &= \int_{\mathbb{R}^q} p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\lambda}, \boldsymbol{\alpha}) p_{\mathbf{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}) d\boldsymbol{\lambda} \\ &= \int_{\mathbb{R}^q} p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\lambda}) p_{\mathbf{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}) d\boldsymbol{\lambda} \quad \text{as given } \boldsymbol{\lambda}, \boldsymbol{y_1}_{,\text{obs}} \text{ does not depend on } \alpha \\ &= \mathbb{E}[p(\boldsymbol{y_1}_{,\text{obs}} \mid \mathbf{\Lambda}) \mid \boldsymbol{\alpha}], \end{aligned} \tag{5}$$

which is the expectation of $p(\boldsymbol{y_1}_{,\text{obs}} \mid \mathbf{\Lambda})$ when $\mathbf{\Lambda}$ has pdf $p_{\mathbf{\Lambda}}(. \mid \boldsymbol{\alpha})$. Thus, the most natural way to estimate $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ for a given $\boldsymbol{\alpha}$ is to compute the Monte Carlo estimator,

$$\hat{P}_L(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}) = \frac{1}{L} \sum_{k=1}^{L} p(\boldsymbol{y_1}_{,\text{obs}} \mid \mathbf{\Lambda}'_k), \tag{6}$$

with $(\mathbf{\Lambda}'_k)_{k=1}^L$ i.i.d. with pdf $p_{\mathbf{\Lambda}}(. \mid \boldsymbol{\alpha})$. With this approach, it would lead to $L \times n$ runs of the computer code for the estimation of the likelihood of a single $\boldsymbol{\alpha}$, which is not feasible in practice. To save significant computation time, the idea here is to use an importance sampling formulation, presented in [21], based on the following result:

$$\begin{aligned} &\forall(\boldsymbol{\alpha}, \boldsymbol{\alpha}^\star) \text{ such that } \text{supp}(p_{\mathbf{\Lambda}}(. \mid \boldsymbol{\alpha})) \subset \text{supp}(p_{\mathbf{\Lambda}}(. \mid \boldsymbol{\alpha}^\star)), \\ &\mathbb{E}[p(\boldsymbol{y_1}_{,\text{obs}} \mid \mathbf{\Lambda}) \mid \boldsymbol{\alpha}] = \mathbb{E}\left[p(\boldsymbol{y_1}_{,\text{obs}} \mid \mathbf{\Lambda}) \frac{p_{\mathbf{\Lambda}}(\mathbf{\Lambda} \mid \boldsymbol{\alpha})}{p_{\mathbf{\Lambda}}(\mathbf{\Lambda} \mid \boldsymbol{\alpha}^\star)} \mid \boldsymbol{\alpha}^\star\right]. \end{aligned}$$

Then, for all $\boldsymbol{\alpha}$, we can use the following importance sampling estimator of $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$, based on sampling associated with a vector of hyperparameters $\boldsymbol{\alpha}^\star$:

$$\hat{P}_L^{\boldsymbol{\alpha}^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}) = \frac{1}{L} \sum_{k=1}^{L} p(\boldsymbol{y_1}_{,\text{obs}} \mid \mathbf{\Lambda}'_k) \frac{p_{\mathbf{\Lambda}}(\mathbf{\Lambda}'_k \mid \boldsymbol{\alpha})}{p_{\mathbf{\Lambda}}(\mathbf{\Lambda}'_k \mid \boldsymbol{\alpha}^\star)}, \tag{7}$$

with $(\mathbf{\Lambda}'_k)_{k=1}^L$ i.i.d. with pdf $p_{\mathbf{\Lambda}}(. \mid \boldsymbol{\alpha}^\star)$. This estimator is consistent and asymptotically normal, as detailed in Section 2.3. Here, the same samples are used for all $\boldsymbol{\alpha}$. Then, with only $L \times n$ runs of the computer code we can estimate the likelihood $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ for all $\boldsymbol{\alpha}$, rather than just for a single $\boldsymbol{\alpha}$ with the Monte Carlo estimation. This allows us to conduct a plug-in strategy by maximizing the posterior density of $\mathbf{A}$ and working with the maximum *a posteriori* $\boldsymbol{\alpha}_{\text{MAP}} = \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\text{argmax}}\, p(\boldsymbol{\alpha} \mid \boldsymbol{y_1}_{,\text{obs}})$. Another strategy is to adopt a full-Bayesian approach and consider the entire posterior distribution of $\mathbf{A}$ with MCMC methods.

2.3 Investigating $\boldsymbol{\alpha}_{\text{MAP}}$

This section details the estimation of $\boldsymbol{\alpha}_{\text{MAP}}$ that can then be used as a plug-in value for the hyperparameters to compute $\hat{E}_M\Big(\psi(Y_t(\mathbf{x}_0))\Big)$ with the prior distribution $p_{\mathbf{\Lambda}}(. \mid \boldsymbol{\alpha}_{\text{MAP}})$ for $\mathbf{\Lambda}$. The idea is to consider $\hat{P}_L^{\boldsymbol{\alpha}^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ as the estimator of the likelihood $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ of $\boldsymbol{\alpha}$, for a relevant $\boldsymbol{\alpha}^\star$, as detailed in the following.

2.3.1 Estimation of $\boldsymbol{\alpha}_{MAP}$

Theoretically, the maximization of the posterior density on $\mathcal{A}$ can be performed by selecting a given $\boldsymbol{\alpha}^\star$ and evaluating $\hat{P}_L^{\boldsymbol{\alpha}^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha})$ for every $\boldsymbol{\alpha}$. However, the ratios $(p_{\mathbf{\Lambda}}(\mathbf{\Lambda}'_k \mid \boldsymbol{\alpha}))/(p_{\mathbf{\Lambda}}(\mathbf{\Lambda}'_k \mid \boldsymbol{\alpha}^\star))$ can be highly fluctuating when $\boldsymbol{\alpha}$ is far from $\boldsymbol{\alpha}^\star$, leading to estimators with a high variance. To overcome this issue, the proposed solution is worked with $\boldsymbol{\alpha}^\star$ close to $\boldsymbol{\alpha}_{\text{MAP}}$, using an iterative strategy in $\boldsymbol{\alpha}^\star$ until reaching a fixed point, as detailed in Algorithm 1. The inputs of the algorithm are $\mathbf{A}_0^\star$, the initial value of $\boldsymbol{\alpha}^\star$ and τ, the stop threshold. The value of τ depends on the interpretation of $\boldsymbol{\alpha}$ and will be briefly discussed in Section 3.4. Note that in this work, the optimization $\mathbf{A}_{\ell+1}^\star = \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\text{argmax}}\, \hat{P}_L^{\mathbf{A}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha})$ is performed with the L-BFGS-B optimizer [22]. In the following we will work with the optimal vector $\boldsymbol{\alpha}_\ell^\star$ which is the approximation of $\boldsymbol{\alpha}_{\text{MAP}}$ obtained at the end of the algorithm.

2.3.2 Confidence in the Estimation

Although this is not required in the proposed method, it is possible to check that $p(\boldsymbol{\alpha}_\ell^\star \mid \boldsymbol{y_1}_{,\text{obs}})$, the posterior pdf at $\boldsymbol{\alpha}_\ell^\star$, is close to the maximum *a posteriori* density across all $\boldsymbol{\alpha} \in \mathcal{A}$. As detailed in Appendix A, with a statistical test, we propose a confidence level $\gamma(\boldsymbol{\alpha})$ associated with $p(\boldsymbol{\alpha} \mid \boldsymbol{y_1}_{,\text{obs}}) < \beta p(\boldsymbol{\alpha}_\ell^\star \mid \boldsymbol{y_1}_{,\text{obs}})$ for a given $\boldsymbol{\alpha}$, where β represents an acceptable margin of error for this estimation, set by the user. For example, in the application, $\beta = 1.05$ is used, see Section 3.5, corresponding to a tolerance of 5% on the posterior density. This confidence level will be computed for $\boldsymbol{\alpha} \in \mathcal{A}$, as illustrated in Fig. D1(b). A threshold ζ can be introduced to ensure that $\forall \boldsymbol{\alpha} \in \mathcal{A}, \gamma(\boldsymbol{\alpha}) \geq \zeta$, with, for instance, $\zeta = 0.95$. To estimate these confidence levels and similarly to the approach detailed in Section 2.4, we use $\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ as the estimator of $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$. More precisely, for this specific study, a number L' of samples is considered, with L' very large, to work with $\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ and investigate the convergence, when L' goes to $+\infty$, of

Algorithm 1: Iterative estimation of $\boldsymbol{\alpha}_{\text{MAP}}$

Input: $\mathbf{A}_0^\star$, τ
$\ell \leftarrow 0$, $\nu \leftarrow +\infty$
while $\nu > \tau$ **do**
  Sample $(\mathbf{\Lambda}'_k)_{k=1}^L$ i.i.d. with pdf $p_{\mathbf{\Lambda}}(. \mid \mathbf{A}_\ell^\star)$
  $\mathbf{A}_{\ell+1}^\star \leftarrow \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\text{argmax}}\, \hat{P}_L^{\mathbf{A}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha})$
  $\nu \leftarrow \|\mathbf{A}_\ell^\star - \mathbf{A}_{\ell+1}^\star\|$
  $\ell \leftarrow \ell + 1$
end

$$\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta \hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star) p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)$$
$$= \frac{1}{L'} \sum_{k=1}^{L'} \left(p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\Lambda}_k') \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k' \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k' \mid \boldsymbol{\alpha}_\ell^\star) p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k' \mid \boldsymbol{\alpha}_\ell^\star)} \right), \tag{8}$$

with $(\boldsymbol{\Lambda}_k')_{k=1}^{L'}$ i.i.d. with pdf $p_{\boldsymbol{\Lambda}}(. \mid \boldsymbol{\alpha}_\ell^\star)$.

We introduce the estimator

$$S_{L'}(\boldsymbol{\alpha})^2 = \frac{1}{L'-1} \sum_{k=1}^{L'} \left(p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\Lambda}_k') \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k' \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k' \mid \boldsymbol{\alpha}_\ell^\star) p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k' \mid \boldsymbol{\alpha}_\ell^\star)} \right.$$
$$\left. - \left(\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta \hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star) p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star) \right) \right)^2.$$

As detailed in Appendix A, with the central limit theorem, we have the following asymptotic confidence level:

$$\gamma(\boldsymbol{\alpha}) = \Phi \left(\frac{\sqrt{L'} \left(\beta \hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)) p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star) - \hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha}) \right)}{s_{L'}(\boldsymbol{\alpha})} \right), \tag{9}$$

where Φ is the cumulative distribution function (cdf) of a Gaussian distribution with mean 0 and variance 1, and $s_{L'}(\boldsymbol{\alpha})$, $\hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)$ and $\hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha})$ are the obtained realizations of the associated estimators.

2.4 Sampling from the Posterior Distribution of $\mathbf{A}$

Instead of working only with $\boldsymbol{\alpha}_{\text{MAP}}$ as in the plug-in strategy, we can conduct a full-Bayesian strategy and work with the entire posterior distribution $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ estimated by

$$\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) = \frac{\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha})}{\int_{\mathcal{A}} \hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}') p_{\mathbf{A}}(\boldsymbol{\alpha}') d\boldsymbol{\alpha}'}. \tag{10}$$

The method is summarized in Algorithm 2. Note that the first step of this procedure is solely used to estimate the likelihood for all values of $\boldsymbol{\alpha}$, with the estimator $\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$. Since $\|\boldsymbol{\alpha}_\ell^\star - \boldsymbol{\alpha}_{\ell-1}^\star\| \leq \tau$ with small τ, in practice $\hat{P}_L^{\boldsymbol{\alpha}_{\ell-1}^\star}(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$, obtained from Algorithm 1, can be used to estimate $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$. Therefore, no additionnal call to the simulator is required compared to the plug-in method. To simplify the notations and ensure everything depends only on $\boldsymbol{\alpha}_\ell^\star$, the method is presented with $\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ as the estimator.

Algorithm 2: Full-Bayesian estimation of $\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$

Input: $\boldsymbol{\alpha}_\ell^\star$
Sample $(\boldsymbol{\Lambda}_k')_{k=1}^L$ i.i.d. with pdf $p_{\boldsymbol{\Lambda}}(. \mid \boldsymbol{\alpha}_\ell^\star)$.
Sample $(\mathbf{A}_i)_{i=1}^N$ with pdf $\propto \hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}}|.) p_{\mathbf{A}}$ by MCMC, with $\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}}|.)$ given by Eq. (7).
Sample $(\boldsymbol{\Lambda}_k)_{k=1}^M$ with pdf $\propto p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda}) p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}_\ell^\star)$ by MCMC.
Compute

$$\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^\star}\Big(\psi(Y_t(\mathbf{x}_0))\Big) = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{k=1}^M \psi(f_t(\mathbf{x}_0, \boldsymbol{\Lambda}_k)) \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\mathbf{A}_i)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\boldsymbol{\alpha}_\ell^\star)}}{\sum_{k=1}^M \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\mathbf{A}_i)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\boldsymbol{\alpha}_\ell^\star)}}.$$

We have

$$\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) = \int_{\mathcal{A}} \mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{\alpha}, \boldsymbol{y}_{\mathbf{1},\text{obs}}) p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) d\boldsymbol{\alpha}. \tag{11}$$

As shown in the supplementary material [23], under additionnal assumptions on $\boldsymbol{\lambda} \mapsto p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda})$, $p_{\boldsymbol{\Lambda}}(. \mid \boldsymbol{\alpha})$ and $p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda} \mid .)$, $\int_{\mathcal{A}} \mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{\alpha}, \boldsymbol{y}_{\mathbf{1},\text{obs}}) \hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) d\boldsymbol{\alpha}$ is a consistent estimator of $\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$.

Therefore, with MCMC methods, we sample $(\mathbf{A}_i)_{i=1}^N$ with pdf proportional to $\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}}|.) p_{\mathbf{A}}$ and investigate

$$\tilde{\mathbb{E}}_N^{\boldsymbol{\alpha}_\ell^\star}(\psi(Y_t(\mathbf{x}_0))) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \mathbf{A}_i, \boldsymbol{y}_{\mathbf{1},\text{obs}}), \tag{12}$$

which approximates $\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ but is not directly an estimator of $\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$, as $\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \mathbf{A}_i, \boldsymbol{y}_{\mathbf{1},\text{obs}})$ is unknown.

Then $\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \mathbf{A}_i, \boldsymbol{y}_{\mathbf{1},\text{obs}})$ needs to be estimated. As explained in Section 2.2, importance sampling estimation can be employed to reduce the computation cost. Indeed, from Eq. (3) we have $\forall \boldsymbol{\alpha}$ s.t. $\text{supp}(p_{\boldsymbol{\Lambda}}(. \mid \boldsymbol{\alpha})) \subset \text{supp}(p_{\boldsymbol{\Lambda}}(. \mid \boldsymbol{\alpha}_\ell^\star))$:

$$\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{\alpha}, \boldsymbol{y}_{\mathbf{1},\text{obs}}) = \frac{\int_{\mathbb{R}^q} \psi(f_t(\mathbf{x}_0, \boldsymbol{\lambda})) p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda}) \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}|\boldsymbol{\alpha})}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}|\boldsymbol{\alpha}_\ell^\star)} p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}_\ell^\star) d\boldsymbol{\lambda}}{\int_{\mathbb{R}^q} p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda}) \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}|\boldsymbol{\alpha})}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}|\boldsymbol{\alpha}_\ell^\star)} p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}_\ell^\star) d\boldsymbol{\lambda}}. \tag{13}$$

This equation gives a full-Bayesian estimator of $\mathbb{E}(\psi(Y_t(\mathbf{x}_0)) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$:

$$\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^\star}\Big(\psi(Y_t(\mathbf{x}_0))\Big) = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{k=1}^M \psi(f_t(\mathbf{x}_0, \boldsymbol{\Lambda}_k)) \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\mathbf{A}_i)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\boldsymbol{\alpha}_\ell^\star)}}{\sum_{k=1}^M \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\mathbf{A}_i)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\boldsymbol{\alpha}_\ell^\star)}}, \tag{14}$$

with $(\mathbf{A}_i)_{i=1}^N$ sampled with pdf proportional to $\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha})$ and $(\boldsymbol{\Lambda}_k)_{k=1}^M$ sampled with pdf proportional to $p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda}) p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}_\ell^\star)$. To ensure for all $\mathbf{A}_i$, the inclusion $\text{supp}(p_{\boldsymbol{\Lambda}}(. \mid \mathbf{A}_i)) \subset \text{supp}(p_{\boldsymbol{\Lambda}}(. \mid \boldsymbol{\alpha}_\ell^\star))$, $\sum_{k=1}^M (p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k \mid \mathbf{A}_i)/p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k \mid \boldsymbol{\alpha}_\ell^\star)) > 0$, and for the consistency of the estimators (see Supplementary Material [23]), in the following we impose:

$$\exists \text{ compact set } \mathcal{K}, \forall \boldsymbol{\alpha} \in \mathcal{A}, \text{supp}(p_{\boldsymbol{\Lambda}}(. \mid \boldsymbol{\alpha})) = \mathcal{K}. \tag{15}$$

If we consider $N = 1$ and $p_{\mathbf{A}} = \delta_{\boldsymbol{\alpha}_\ell^\star}$ in Eq. (14) and $p_{\boldsymbol{\Lambda}} = p_{\boldsymbol{\Lambda}}(. \mid \boldsymbol{\alpha}^\star)$ in Eq. (4), then $\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^\star}\Big(\psi(Y_t(\mathbf{x}_0))\Big)$ becomes equivalent to $\hat{E}_M\Big(\psi(Y_t(\mathbf{x}_0))\Big)$. The full-Bayesian estimator is a generalization of the estimator of Eq. (4). Consequently, all equations related to $\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^\star}\Big(\psi(Y_t(\mathbf{x}_0))\Big)$ in the following remain valid for $\hat{E}_M\Big(\psi(Y_t(\mathbf{x}_0))\Big)$.

More generally, we will denote

$$\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^\star}\Big(h(\boldsymbol{\Lambda})\Big) = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{k=1}^M h(\boldsymbol{\Lambda}_k) \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\mathbf{A}_i)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\boldsymbol{\alpha}_\ell^\star)}}{\sum_{k=1}^M \frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\mathbf{A}_i)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k|\boldsymbol{\alpha}_\ell^\star)}}, \tag{16}$$

for every bounded and continuous function h on $\mathcal{K}$, with $(\mathbf{A}_i)_{i=1}^N$ sampled with pdf proportional to $\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha})$ and $(\boldsymbol{\Lambda}_k)_{k=1}^M$ sampled with pdf proportional to $p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda}) p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}_\ell^\star)$. It is the full-Bayesian estimator of $\mathbb{E}(h(\boldsymbol{\Lambda}) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ and it will be used with different h functions. From Sections 2 and 3 of [23], we have the following:

- $\int_{\mathcal{A}} \mathbb{E}(h(\boldsymbol{\Lambda}) \mid \boldsymbol{\alpha}, \boldsymbol{y}_{\mathbf{1},\text{obs}}) \hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) d\boldsymbol{\alpha}$ is a consistent estimator of $\mathbb{E}(h(\boldsymbol{\Lambda}) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$.
- Given $\hat{P}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) = \hat{p}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$, $\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^\star}\Big(h(\boldsymbol{\Lambda})\Big)$ is a consistent estimator of $\int_{\mathcal{A}} \mathbb{E}(h(\boldsymbol{\Lambda}) \mid \boldsymbol{\alpha}, \boldsymbol{y}_{\mathbf{1},\text{obs}})$ $\hat{p}_L^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) d\boldsymbol{\alpha}$.

2.5 Surrogate Modeling

In all of the above, we considered that we can perform runs of the simulator $\boldsymbol{f}(\mathbf{x},\boldsymbol{\lambda})$ to compute the estimators $\hat{E}_M\Big(\psi(Y_t(\mathbf{x}_0))\Big)$, $\hat{P}_L(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ or $\hat{E}^{\boldsymbol{\alpha}^\star_\ell}_{N,M}\Big(\psi(Y_t(\mathbf{x}_0))\Big)$. It is important to note that in the approach presented here, the computation of the likelihood $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\lambda}) = p_{\boldsymbol{E}}\big(\boldsymbol{y_1}_{,\text{obs}} - (f_1(\mathbf{x}_j,\boldsymbol{\lambda}))_{j=1}^n\big)$ requires n runs for a single $\boldsymbol{\lambda}$. It leads to the following numbers of calls to the simulator:

- $L \times n \times \ell$ runs for the estimation of $\boldsymbol{\alpha}_{\text{MAP}}$, where ℓ is the number of iterations before convergence of Algorithm 1.
- $L \times n + M \times (n+1)$ runs for the estimation of $\hat{E}^{\boldsymbol{\alpha}^\star_\ell}_{N,M}\Big(\psi(Y_t(\mathbf{x}_0))\Big)$. The factor $n+1$ comes from the fact that simulations must be run at $(\mathbf{x}_j)_{j=1}^n$ and at the new point $\mathbf{x}_0$. If $\hat{p}_L^{\boldsymbol{\alpha}^\star_{\ell-1}}(\boldsymbol{\alpha} \mid \boldsymbol{y_1}_{,\text{obs}})$ is used for the likelihood approximation, then only $M \times (n+1)$ runs are needed.

In practice, these computations are not feasible when the physical simulations are costly (for instance, several hours long each) and then surrogate models are required. First, surrogate models are required for the approximation of $(f_1(\mathbf{x}_j,\boldsymbol{\lambda}))_{j=1}^n$ to compute the likelihood $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\lambda}) = p_{\boldsymbol{E}}\big(\boldsymbol{y_1}_{,\text{obs}} - (f_1(\mathbf{x}_j,\boldsymbol{\lambda}))_{j=1}^n\big)$ for the MCMC sampling in the posterior distribution of $\boldsymbol{\Lambda}$ and the investigation of $\boldsymbol{\alpha}^\star_\ell$ (see Section 1). Additionally, a surrogate model is needed for $f_t(\mathbf{x}_0,\boldsymbol{\lambda})$ to compute the prediction estimator at the new point $\mathbf{x}_0$ for the unobserved output f_t. Therefore, for $\mathbf{x} \in \mathbb{X} \cup \{\mathbf{x}_0\}$ and $1 \le t \le T$, we propose to build a surrogate model for $f_{t,\mathbf{x}} \colon \boldsymbol{\lambda} \longrightarrow f_t(\mathbf{x},\boldsymbol{\lambda})$.

Gaussian processes (GP) are used here as a surrogate model as they make it possible to quantify the uncertainties, as presented in [24]. Let us assume that $f_{t,\mathbf{x}}$ is a realization of a Gaussian process $Z_{t,\mathbf{x}}(\boldsymbol{\lambda})$ with prior mean $\mu_{t,\mathbf{x}}(\boldsymbol{\lambda})$ and prior covariance kernel $k_{t,\mathbf{x}}(\boldsymbol{\lambda},\boldsymbol{\lambda}')$. $Z_{t,\mathbf{x}}$ and $Z_{t',\mathbf{x}'}$ are considered independent if $(t,\mathbf{x}) \neq (t',\mathbf{x}')$. The kernel models the correlation structure between $f_{t,\mathbf{x}}(\boldsymbol{\lambda})$ and $f_{t,\mathbf{x}}(\boldsymbol{\lambda}')$. In our study, the stationary Matérn 5/2 kernel, see [25], is used. This choice of independency between the surrogate models implies that the Gaussian process used for the prediction at $\mathbf{x}_0$ is independent of the processes associated with the observation points $\mathbf{x}_j$.

We consider a design of experiments $\mathbb{L}_{\text{train}} = (\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_{n_{\text{train}}})$ and the observations $Z^{\text{train}}_{t,\mathbf{x}} = (f_{t,\mathbf{x}}(\boldsymbol{\lambda}_i))_{i\in I_{\text{train}}}$ with $I_{\text{train}} = \{1,\ldots,n_{\text{train}}\}$. $f_{t,\mathbf{x}}$ can be approximated using the distribution of $Z_{t,\mathbf{x}}(\boldsymbol{\lambda}) \mid Z(\mathbb{L}_{\text{train}}) = Z^{\text{train}}_{t,\mathbf{x}}$, which is a Gaussian process with the following mean and covariance:

$$\begin{cases} \hat{f}_t(\mathbf{x},\boldsymbol{\lambda}) = \mu_{t,\mathbf{x}}(\boldsymbol{\lambda}) + k_{t,\mathbf{x}}(\boldsymbol{\lambda},\mathbb{L}_{\text{train}})k_{t,\mathbf{x}}(\mathbb{L}_{\text{train}},\mathbb{L}_{\text{train}})^{-1}(Z^{\text{train}}_{t,\mathbf{x}} - \mu_{t,\mathbf{x}}(\mathbb{L}_{\text{train}})) \\ c_t(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\lambda}') = k_{t,\mathbf{x}}(\boldsymbol{\lambda},\boldsymbol{\lambda}') - k_{t,\mathbf{x}}(\boldsymbol{\lambda},\mathbb{L}_{\text{train}})k_{t,\mathbf{x}}(\mathbb{L}_{\text{train}},\mathbb{L}_{\text{train}})^{-1}k_{t,\mathbf{x}}(\boldsymbol{\lambda}',\mathbb{L}_{\text{train}})^T \end{cases},$$

where $k(\mathbb{L}_{\text{train}},\mathbb{L}_{\text{train}}) = [k_{t,\mathbf{x}}(\boldsymbol{\lambda}_i,\boldsymbol{\lambda}_j)]_{i,j\in I_{\text{train}}}$ is the covariance matrix and $k_{t,\mathbf{x}}(\boldsymbol{\lambda},\mathbb{L}_{\text{train}}) = [k_{t,\mathbf{x}}(\boldsymbol{\lambda},\boldsymbol{\lambda}_i)]_{i\in I_{\text{train}}}$.

The variance is denoted $v_t(\mathbf{x},\boldsymbol{\lambda}) = c_t(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\lambda})$. The model presented in Eq. (2) then becomes

$$\begin{cases} Y_t(\mathbf{x}) = \hat{f}_t(\mathbf{x},\boldsymbol{\Lambda}) + \sqrt{v_t(\mathbf{x},\boldsymbol{\Lambda})}E^f(\mathbf{x}), & 1 \le t \le T,\ \mathbf{x} \in \mathcal{X} \\ Y_{1,\text{obs},j} = Y_1(\mathbf{x}_j) + E_j, & 1 \le j \le n \end{cases}, \tag{17}$$

with $E^f(\mathbf{x}) \sim \mathcal{N}(0,1)$, $E^f(\mathbf{x}) \perp\!\!\!\perp E^f(\mathbf{x}')$ if $\mathbf{x} \neq \mathbf{x}'$, and $\boldsymbol{E} = (E_j)_{j=1}^n$ of pdf $p_{\boldsymbol{E}}$. The pdf of the likelihood of $\boldsymbol{\lambda}$ becomes $p_{E^{\text{tot}}(\boldsymbol{\lambda})}\Big((y_{1,\text{obs},j} - \hat{f}_1(\mathbf{x}_j,\boldsymbol{\lambda}))_{j=1}^n\Big)$ with

$$E^{\text{tot}}(\boldsymbol{\lambda}) = \left(E_j + \sqrt{v_1(\mathbf{x}_j,\boldsymbol{\lambda})}E^f(\mathbf{x}_j)\right)_{j=1}^n. \tag{18}$$

Incorporating the variance of the Gaussian process into the likelihood computation ensures that its uncertainty is accounted for during sampling. Particularly, in our study we consider that $p_{\boldsymbol{E}}$ is the pdf of a Gaussian white noise vector, i.e., $\boldsymbol{E} \sim \mathcal{N}(0,\sigma_\varepsilon^2 I)$. This gives $E^{\text{tot}}(\boldsymbol{\lambda}) \sim \mathcal{N}\left(0,\sigma_\varepsilon^2 I + \text{diag}\Big(v_1(\mathbf{x}_j,\boldsymbol{\lambda})\Big)_{j=1}^n\right)$, and the estimators of the prediction mean and variance (from the law of total variance, see [26]) are

$$\begin{aligned} \hat{E}^{\boldsymbol{\alpha}_\ell^\star}_{N,M}\Big(Y_t(\mathbf{x}_0)\Big) &= \hat{E}^{\boldsymbol{\alpha}_\ell^\star}_{N,M}\Big(\hat{f}_t(\mathbf{x}_0,\boldsymbol{\Lambda})\Big), \\ \hat{E}^{\boldsymbol{\alpha}_\ell^\star}_{N,M}\Big(Y_t(\mathbf{x}_0)^2\Big) - \hat{E}^{\boldsymbol{\alpha}_\ell^\star}_{N,M}\Big(Y_t(\mathbf{x}_0)\Big)^2 &= \hat{E}^{\boldsymbol{\alpha}_\ell^\star}_{N,M}\Big(\hat{f}_t(\mathbf{x}_0,\boldsymbol{\Lambda})^2 + \nu_t(\mathbf{x}_0,\boldsymbol{\Lambda})\Big) - \hat{E}^{\boldsymbol{\alpha}_\ell^\star}_{N,M}\Big(\hat{f}_t(\mathbf{x}_0,\boldsymbol{\Lambda})\Big)^2, \end{aligned} \tag{19}$$

This approach, which uses a separate Gaussian process (GP) for each $\mathbf{x}$, is well suited to the current context, as it focuses on situations with a very limited number of observations n and a small number of predictions. Note that it is particularly appropriate in other cases where the control variables $\mathbf{x}_j$ are categorical instead of numerical. However, other implementations are possible, especially with higher values of n, by considering a Gaussian process $Z_t(\mathbf{x},\boldsymbol{\lambda})$ in the joint space $\mathcal{X}\times\mathcal{L}$, with covariance $c_t(\mathbf{x},\mathbf{x}',\boldsymbol{\lambda},\boldsymbol{\lambda}')$. In this case, the computation of the likelihood of $\boldsymbol{\lambda}$ would be modified, as the distribution of $E^{\text{tot}}(\boldsymbol{\lambda})$ is now given by $E^{\text{tot}}(\boldsymbol{\lambda}) \sim \mathcal{N}\Big(0, \sigma_\varepsilon^2 I + \big(c_t(\mathbf{x}_{j_1},\mathbf{x}_{j_2},\boldsymbol{\lambda},\boldsymbol{\lambda})\big)_{j_1,j_2=1}^n\Big)$.

3. APPLICATION

The previously introduced methods are tested on an application case well suited for calibration with unobserved outputs, as presented here.

3.1 Case Description

The test case considered here deals with the Taylor cylinder impact test [27], as illustrated in Fig. 1. More precisely, it models the impact of a cylinder with length ℓ_0, radius r_0, launched at velocity v_0 against a rigid wall. This computer code requires four physical parameters associated with the Mie-Grüneisen equation of state [28] that are unknown and need to be calibrated. These parameters are ρ_0 the intial density of the cylinder, C_0 the bulk speed of sound, Γ_0 the Grüneisen coefficient at the reference state, and S the Hugoniot slope parameter. Besides these unknown parameters, the computer code depends on two additional variables, Y_0 and Y_M, associated with Ludwik's equation, which describes the electroplastic deformation process: $\tau = Y_0 + Y_M \epsilon_p^{n_0}$, where τ is the stress, Y_0 the yield stress, Y_M the strength index, ϵ_p the plastic strain, and n_0 is the strain hardening index [29]. The two parameters Y_0 and Y_M have been calibrated using another experiment, the tensile test, this is why they are considered known here and do not initially require calibration. However, they are reintroduced in the calibration process as additional parameters in $\mathcal{L}_{\text{err}}$ to capture the model error. As previously explained, the aim is that incorporating this uncertainty will enhance the predictive capability of the model. Thus, in our study we have

- three control variables $\mathbf{x} = (\ell_0, r_0, v_0)$,
- $p = 4$ parameters to calibrate ρ_0, C_0, Γ_0, S, and
- $q - p = 2$ additional parameters Y_0 and Y_M.

This computer code, validated on results from [30], is interesting in our context because it produces $T = 3$ outputs:

- the length difference $\ell_f - \ell_0$ after impact,

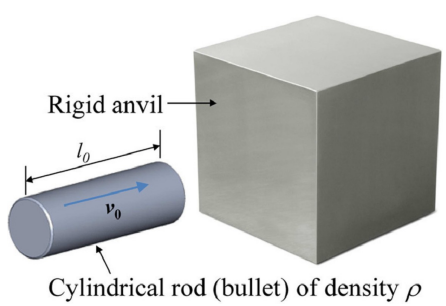


FIG. 1: Description of the Taylor cylinder impact test on a copper rod

- the final radius r_f, and
- the maximum strain over time $\epsilon_{\text{max}} = \max\ \epsilon_p$.

Therefore, we will consider three different transpositions: the first case where we have observations of $\ell_f - \ell_0$ only, the second with observations of r_f only, and finally the third with observations of ϵ_{max} only. All input variables are constrained by predefined bounds, beyond which the simulation is not feasible. In the following, to simplify, they are all considered as normalized between 0 and 1.

3.2 Model Error

Unfortunately, no real experimental data were available for this study. However, one can easily generate virtual measurements and incorporate a custom model error to work within the previously described framework. This custom model error is introduced through a perturbation of the control variables, ensuring that the synthetic data are generated using a method whose error structure does not exactly replicate that of the inference approach we aim to apply. Since our hierarchical methods rely on fluctuations of Y_0 and Y_M, the model error must not be based on these parameters. Therefore, we choose to perturb the control variables instead.

For a given triplet $\mathbf{x} = (\ell_0, r_0, v_0)$, we define $\omega(\mathbf{x}) = (\ell_0, r_0, v_0 + \Delta V)$, where ΔV is fixed to a constant in our work. For $1 \leq t \leq T$, the true value is introduced as $y_t(\mathbf{x}) = f_t(\omega(\mathbf{x}), \boldsymbol{\lambda}_0)$ for a given $\boldsymbol{\lambda}_0 \in \mathcal{L}$, leading to a model error, as there is no specific value of $\boldsymbol{\lambda}$ that verifies that $\forall \mathbf{x} \in \mathcal{X}, f_t(\mathbf{x}, \boldsymbol{\lambda}) = y_t(\mathbf{x})$. Then the protocol is the following:

- Select n points $\mathbb{X} = (\mathbf{x}_j)_{j=1}^n \in \mathcal{X}^n$ with a Maximin Latin Hypercube Sampling [31]. n is small to work with a limited number of observations.
- Compute $(y_t(\mathbf{x}_j))_{j=1}^n = (f_t(\omega(\mathbf{x}_j), \boldsymbol{\lambda}_0))_{j=1}^n$.
- Add a realization $\boldsymbol{\mathcal{E}} = (\mathcal{E}_j)_{j=1}^n$ of the vector $\boldsymbol{E} \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$ as the measurement error for the observed variable.

Finally, it provides the vector of experimental data

$$\boldsymbol{y_{1,\text{obs}}} = (y_1(\mathbf{x}_j) + \mathcal{E}_j)_{j=1}^n.$$

3.3 Performance Metrics

Beyond the measurements of the first variable $\boldsymbol{y_{1,\text{obs}}} = (y_{1,\text{obs},j})_{j=1}^n = (y_1(\mathbf{x}_j) + \mathcal{E}_j)_{j=1}^n \in \mathbb{R}^n$, here we also have access to the true values $y_t(\mathbf{x}_j)$ for $1 \leq t \leq T, 1 \leq j \leq n$. A leave-one-out (LOO) approach is used, to predict $y_t(\mathbf{x}_j)$ for $1 \leq j \leq n$ based on the noisy observations $\boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}} = (y_1(\mathbf{x}_{j'}) + E_{j'})_{\substack{j'=1 \\ j' \neq j}}^n$. This makes it possible to assess the performance of the method with a small dataset. Then a specific $\boldsymbol{\alpha}_\ell^\star$ is identified for every j, which will be denoted as $\boldsymbol{\alpha}_\ell^{j\star}$.

More precisely, for $j = 1, \ldots, n$ and $t = 1, \ldots, T$, we have a random prediction of $Y_t(\mathbf{x}_j) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}}$. Its distribution is called the predictive distribution, which will be compared to the true value $y_t(\mathbf{x}_j)$. In order to have a direct comparison of the different investigated methods, the scoring rule introduced in Eq. (25) of [32], based on the mean and the covariance of the predictive distribution, is selected here. In the general case, it is expressed as

$$s(\boldsymbol{Y}, \boldsymbol{y}) = -\log(\det(\boldsymbol{\Sigma})) - (\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}), \quad \text{with} \tag{20}$$

- $\boldsymbol{Y}$ a random prediction vector of mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and
- $\boldsymbol{y}$ the vector of true values of the phenomenon.

Here, in the LOO scheme, the predictions $\left(Y_t(\mathbf{x}_j) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}}\right)_{j=1}^{n}$ are independent. The estimated mean and variance of the prediction $Y_t(\mathbf{x}_j) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}}$ are defined as $\hat{\mu}_{t,N,M}(\mathbf{x}_j)$ and $\hat{\mathrm{v}}_{t,N,M}(\mathbf{x}_j)$, computed using the estimators $\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^{j\star}}\left(Y_t(\mathbf{x}_j)\right)$ and $\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^{j\star}}\left(Y_t(\mathbf{x}_j)^2\right) - \hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^{\star}}\left(Y_t(\mathbf{x}_j)\right)^2$ from Eq. (19).

For $j = 1, \ldots, n$, the score is introduced as

$$s_{t,N,M}(\mathbf{x}_j) = -\log(\hat{\mathrm{v}}_{t,N,M}(\mathbf{x}_j)) - \frac{(y_t(\mathbf{x}_j) - \hat{\mu}_{t,N,M}(\mathbf{x}_j))^2}{\hat{\mathrm{v}}_{t,N,M}(\mathbf{x}_j)},$$

and we finally investigate the overall prediction score,

$$s_{t,N,M}(\mathbb{X}) = \sum_{j=1}^{n} s_{t,N,M}(\mathbf{x}_j),$$

which corresponds to the score in Eq. (20) under the assumption of independent predictions, i.e., with $\boldsymbol{\Sigma}$ as a diagonal matrix. This score provides a meaningful balance between prediction accuracy and uncertainty quantification, making it particularly well suited to our application. The score is computed for all the investigated methods in our study, and the results are compared in Section 3.5, with higher scores indicating better predictive performance.

3.4 Comparison of Methods

To perform a comprehensive analysis of our method, we implement different approaches and compare the previously introduced performance metrics (see Section 3.3). All these approaches are summarized in Table 1.

The first method is to consider no model error in our Bayesian framework, and is called *No error*. We then consider $\boldsymbol{\lambda} = (\rho_0, C_0, \Gamma_0, S) \in \mathcal{L}_{\text{cal}} \subset \mathbb{R}^p$ with a noninformative prior distribution $\mathcal{U}_{[0,1]^p}$. Then we can implement the method described in Section 2.1, and work with the estimator $\hat{E}_M\left(\psi(Y_t(\mathbf{x}_0))\right)$ with $p_{\boldsymbol{\Lambda}} = \mathbb{1}_{[0,1]^p}$.

The second method is to introduce the additionnal parameters Y_0 and Y_M to represent the model error, but without a hierarchical model. Indeed, we consider $\boldsymbol{\lambda} = (\rho_0, C_0, \Gamma_0, S, Y_0, Y_M) \in \mathcal{L}_{\text{cal}} \times \mathcal{L}_{\text{err}} \subset \mathbb{R}^q$ and investigate $\hat{E}_M\left(\psi(Y_t(\mathbf{x}_0))\right)$ with $p_{\boldsymbol{\Lambda}} = \mathbb{1}_{[0,1]^q}$. It is called *Uniform error*.

The third method, called *Hierarchical MAP*, includes the additionnal parameters Y_0 and Y_M with the hierarchical model using a plug-in strategy. It investigates $\boldsymbol{\alpha}_\ell^\star$ as described in Section 2.3 and then studies $\hat{E}_M\left(\psi(Y_t(\mathbf{x}_0))\right)$ with $p_{\boldsymbol{\Lambda}}(\cdot \mid \boldsymbol{\alpha}_\ell^\star)$. In our work, we consider

TABLE 1: Summary of all the methods investigated in our study

	Formulation	Distribution
No error	$Y_t(\mathbf{x}) = f_t(\mathbf{x}, \boldsymbol{\Lambda})$ $\boldsymbol{\Lambda} \in \mathbb{R}^p$	$p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}) = \mathbb{1}_{[0,1]^p}$
Uniform error	$Y_t(\mathbf{x}) = f_t(\mathbf{x}, \boldsymbol{\Lambda})$ $\boldsymbol{\Lambda} \in \mathbb{R}^q$	$p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}) = \mathbb{1}_{[0,1]^q}$
Hierarchical MAP	$Y_t(\mathbf{x}) = f_t(\mathbf{x}, \boldsymbol{\Lambda})$ $\boldsymbol{\Lambda} \in \mathbb{R}^q$	$p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}_\ell^\star) = \mathbb{1}_{[0,1]^p}(\lambda_1, \ldots, \lambda_p)$ $\times f_{\mathcal{N}_t(\boldsymbol{\alpha}_\ell^\star, 0.45^2, 0, 1)}(\lambda_{p+1}, \ldots, \lambda_q)$
Hierarchical full Bayes	$Y_t(\mathbf{x}) = f_t(\mathbf{x}, \boldsymbol{\Lambda})$ $\boldsymbol{\Lambda} \in \mathbb{R}^q$	$p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda} \mid \mathbf{A}) = \mathbb{1}_{[0,1]^p}(\lambda_1, \ldots, \lambda_p)$ $\times f_{\mathcal{N}_t(\mathbf{A}, 0.45^2, 0, 1)}(\lambda_{p+1}, \ldots, \lambda_q)$
Embedded discrepancy	$Y_t(\mathbf{x}) = f_t(\mathbf{x}, \mathbf{A}^1 + \delta(\mathbf{A}^2, \boldsymbol{\Xi}))$ $\mathbf{A}^1 \in \mathbb{R}^q,\ \mathbf{A}^2 \in \mathbb{R}^q$	$\delta(\mathbf{A}^2, \boldsymbol{\Xi}) = \text{diag}(\mathbf{A}^2 \boldsymbol{\Xi}^T)$ $\boldsymbol{\Xi} \sim \mathcal{U}([-1, 1]^q)$ $p_{(\mathbf{A}^1, \mathbf{A}^2)}(\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2) \propto \prod_{i=1}^q \mathbb{1}_{\boldsymbol{\alpha}_i^2 > 0} \mathbb{1}_{\boldsymbol{\alpha}_1^1 - \boldsymbol{\alpha}_i^2 > 0} \mathbb{1}_{\boldsymbol{\alpha}_i^1 + \boldsymbol{\alpha}_i^2 < 1}$

$$p_{\mathbf{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}) = \mathbb{1}_{[0,1]^p}(\lambda_1, \ldots, \lambda_p) \times f_{\mathcal{N}_t(\boldsymbol{\alpha}, 0.45^2, 0, 1)}(\lambda_{p+1}, \ldots, \lambda_q), \tag{21}$$

where $f_{\mathcal{N}_t(\mu,\sigma^2,a,b)}$ is the pdf of a Gaussian distribution with mean μ and standard deviation σ truncated between a and b. This choice of standard deviation of the prior $\sigma = 0.45$ aims to balance exploratory behavior, allowing the prior distribution to capture model errors while still concentrating on relevant regions for the additional parameters, in contrast to the uniform error method. Further details on this selection are provided in Appendix B.

The fourth method is the one described in Section 2.4, considering a model error with a hierarchical description but with a full-Bayesian approach, to compute the estimator $\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^\star}\Big(\psi(Y_t(\mathbf{x}_0))\Big)$. We call it *Hierarchical full Bayes*. We consider the prior $p_{\mathbf{\Lambda}} = p_{\mathbf{\Lambda}}(\cdot \mid \boldsymbol{\alpha})$ described in Eq. (21), and $p_{\mathbf{A}}$ the pdf of the uniform distribution $\mathcal{U}_{\mathcal{I}(\boldsymbol{\alpha}_\ell^\star,\kappa)}$, with $\mathcal{I}(\boldsymbol{\alpha}_\ell^\star, \kappa) = [(\boldsymbol{\alpha}_\ell^\star)_1 - \kappa, (\boldsymbol{\alpha}_\ell^\star)_1 + \kappa] \times \cdots \times [(\boldsymbol{\alpha}_\ell^\star)_r - \kappa, (\boldsymbol{\alpha}_\ell^\star)_r + \kappa] \cap [-10, 10]^r$, and $\kappa \in \mathbb{R}$. The constant κ is introduced to prevent the investigation of zones with high fluctuations of $p_{\mathbf{\Lambda}}(\mathbf{\Lambda}_k \mid \boldsymbol{\alpha})/p_{\mathbf{\Lambda}}(\mathbf{\Lambda}_k \mid \boldsymbol{\alpha}_\ell^\star)$. We work with $\kappa = 4$ here. Note that this choice of prior $p_{\mathbf{A}}$ allows for negative values of $\boldsymbol{\alpha}$ (or values higher than 1) that can concentrate $p_{\mathbf{\Lambda}}(\cdot \mid \boldsymbol{\alpha})$ on a zone near 0 (or a zone near 1, respectively).

Finally, the last approach is the one presented in [13], called *Embedded discrepancy* here. Their work is particularly interesting in our context, as they consider a model error that is embedded in the calibration parameters and then can be set up in our transposition situation. The method is detailed in Appendix C, but it is important to highlight here the difference between the *Embedded discrepancy* and the *Hierarchical full Bayes* approach. With our hierarchical model, we consider $Y_t(\mathbf{x}) = f_t(\mathbf{x}, \mathbf{\Lambda})$, where the prior of $\mathbf{\Lambda}$ depends on random hyperparameters $\mathbf{A}$, and a Bayesian inference is conducted for both $\mathbf{\Lambda}$ and $\mathbf{A}$. For a given $\boldsymbol{\alpha}$, the posterior distribution of $\mathbf{\Lambda}$ is proportionnal to $p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\lambda}) p_{\mathbf{\Lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\alpha})$. However, in the *Embedded discrepancy* approach, the model is $Y_t(\mathbf{x}) = f_t(\mathbf{x}, \mathbf{\Lambda})$, where the distribution of $\mathbf{\Lambda}$ depends on random hyperarameters $\mathbf{A} = (\mathbf{A}^1, \mathbf{A}^2)$. Here, the Bayesian inference is conducted only on the hyperparameters $\mathbf{A}$. Particularly, for a given $\boldsymbol{\alpha}$, the distribution of $\mathbf{\Lambda}$ is directly obtained by $p(\mathbf{\Lambda} \mid \boldsymbol{\alpha})$, without any posterior consideration. The parametrization of $\mathbf{\Lambda} \mid \boldsymbol{\alpha}$ plays a crucial role here. Following the recommendations in [13], we adopt a Legendre-uniform polynomial chaos representation for the distribution of $\mathbf{\Lambda} \mid \boldsymbol{\alpha}$, as the parameters $\mathbf{\Lambda}$ are bounded here.

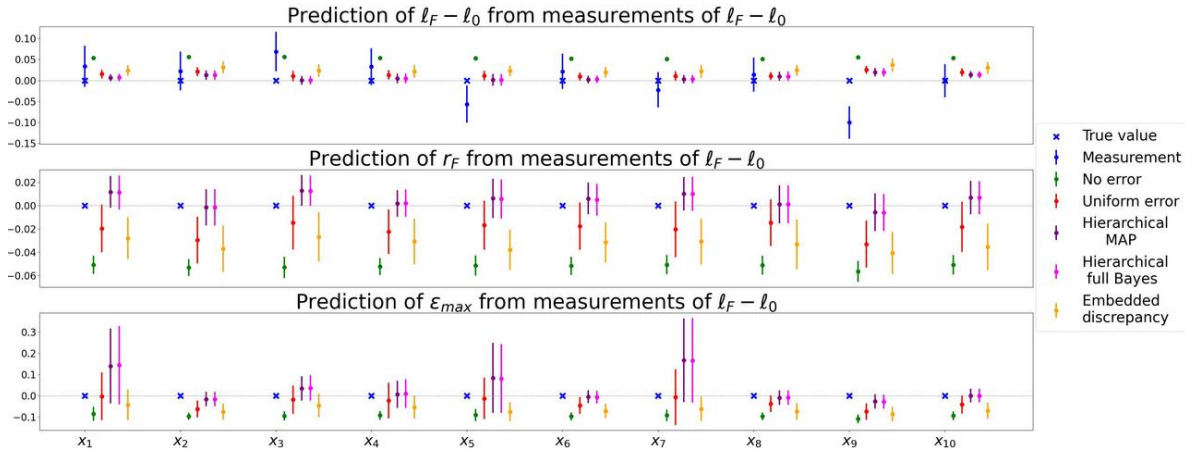
Every MCMC sampling is implemented with the package pymcmcstat [33], and the optimization of the hyperparameters for the Gaussian process regression is performed with pylibkriging [34]. The same Gaussian process is used for all the methods. The number of observations is set to $n = 10$. $L = 10^4$ samples are considered to estimate $\boldsymbol{\alpha}_{\text{MAP}}$, while $M = 3000$ posterior samples of $\mathbf{\Lambda}$ and $N = 750$ samples of $\mathbf{A}$ are generated to compute $\hat{E}_{N,M}^{\boldsymbol{\alpha}_\ell^\star}\Big(\psi(Y_t(\mathbf{x}_0))\Big)$. Note that $N < M$ as $\mathcal{A} \subset \mathbb{R}^2$ compared to $\mathcal{L} \subset \mathbb{R}^6$. $L' = 2 \times 10^4$ samples are used for the estimation of the confidence levels associated with $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) < \beta p(\boldsymbol{\alpha}_\ell^\star \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$. The value of the stopping threshold τ in Algorithm 1 results from a trade-off between choosing a small value to ensure precision and avoiding unnecessary computational cost. In our test case, $\boldsymbol{\alpha}$ corresponds to the mean of a bidimensional Gaussian distribution (since $q - p = 2$), truncated between 0 and 1. We set $\tau = 0.02$, considering that below this value, the Euclidean distance between two successive means, $\boldsymbol{\alpha}_\ell^\star$ and $\boldsymbol{\alpha}_{\ell+1}^\star$, is negligible.

3.5 Numerical Results

The methods are implemented on different designs $\mathbb{X} = (\mathbf{x}_j)_{j=1}^{10}$ for consistency. While only one design is presented here, the remaining results are provided in the "codes/" directory of the Supplementary Material [23], and the conclusions remain consistent across all designs.

3.5.1 Results Overview

The results are shown in Figs. 2–4, corresponding to observations of $\ell_f - \ell_0$, r_f, and ϵ_{max}, respectively. For each figure, the first plot shows the mean and the standard deviation of the normalized posterior distribution $(Y_t(\mathbf{x}_j) - y_t(\mathbf{x}_j))/y_t(\mathbf{x}_j) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}}$ for each method and $1 \leq j \leq 10$, and compare it to the true values $y_t(\mathbf{x}_j)$ shifted to 0 represented with blue crosses for $1 \leq t \leq 3$ [see Figs. 2(a), 3(a), and 4(a)]. From these posterior distributions, the score $s_{t,N,M}$ is evaluated for each output t and summarized in the associated tables [Figs. 2(b), 3(b), and 4(b)].

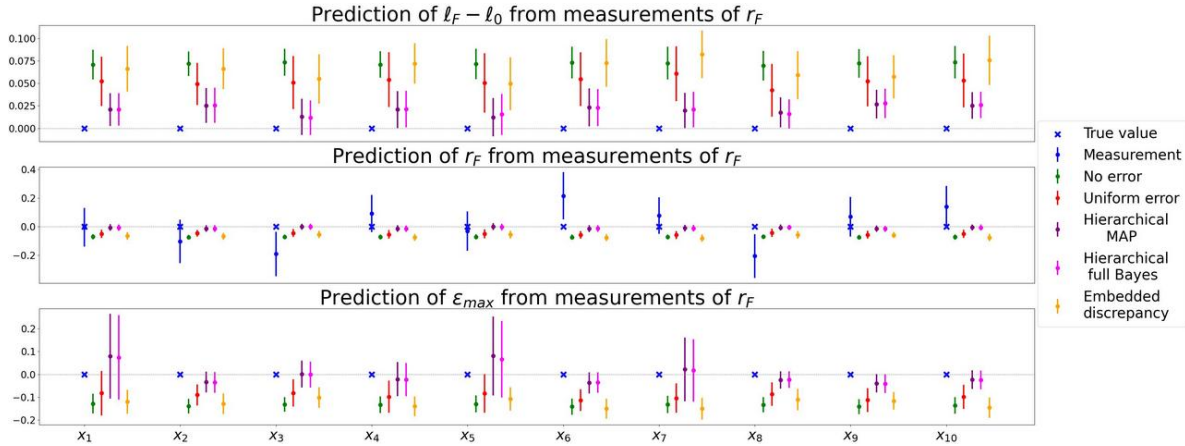


(a)

	No error	Uniform error	Hierarchical MAP	Hierarchical full Bayes	Embedded discrepancy
Prediction of $\ell_f - \ell_0$	–1642.06	6.21	21.40	21.12	–9.88
Prediction of r_f	–360.11	29.66	45.14	45.29	11.10
Prediction of ϵ_{max}	–137.73	36.52	43.16	42.89	24.41

(b)

FIG. 2: Prediction of the three outputs from observations of $\ell_f - \ell_0$. (a) Mean and standard deviation of the normalized posterior distribution $(Y_t(\mathbf{x}_j) - y_t(\mathbf{x}_j))/y_t(\mathbf{x}_j) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}}$, $1 \leq j \leq 10$, with $y_1 = \ell_f - \ell_0$, compared to the true values $y_t(\mathbf{x}_j)$ shifted to 0, shown as crosses for $1 \leq t \leq 3$. The normalized measurement of $\ell_f - \ell_0$ is indicated by the dots, with its standard deviation represented by the associated error bar. (b) Scores $s_{t,N,M}(\mathbb{X})$ from observations of $\ell_f - \ell_0$ for each method.

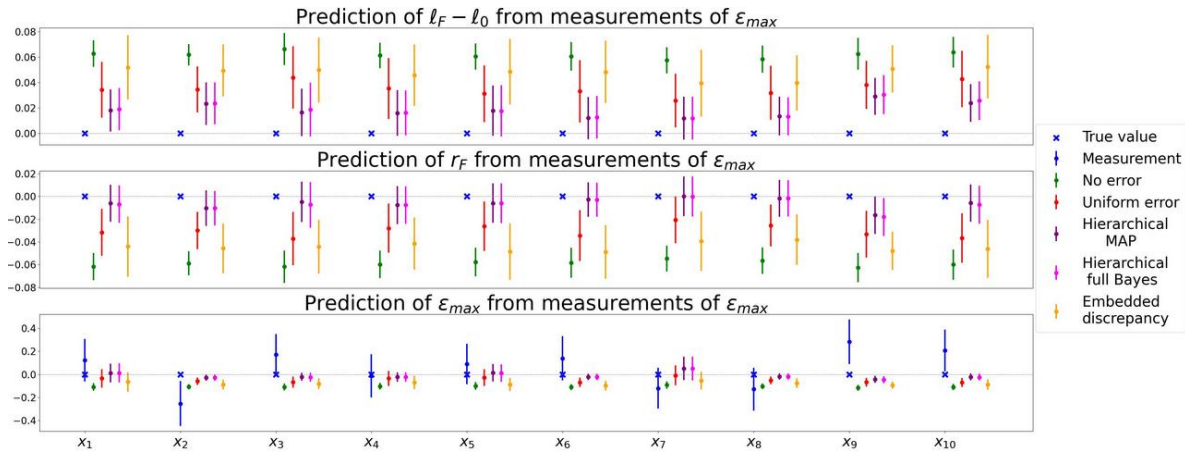


(a)

	No error	Uniform error	Hierarchical MAP	Hierarchical full Bayes	Embedded discrepancy
Prediction of $\ell_f - \ell_0$	–178.16	–22.41	5.13	4.67	–55.48
Prediction of r_f	–136.48	1.58	37.85	37.89	–39.31
Prediction of ϵ_{max}	–92.16	17.78	40.34	40.68	–29.10

(b)

FIG. 3: Prediction of the three outputs from observations of r_f. (a) Mean and standard deviation of the normalized posterior distribution $(Y_t(\mathbf{x}_j) - y_t(\mathbf{x}_j))/y_t(\mathbf{x}_j) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}}$, $1 \leq j \leq 10$, with $y_1 = r_f$, compared to the true values $y_t(\mathbf{x}_j)$ shifted to 0 shown as crosses for $1 \leq t \leq 3$. The measurement of r_f is indicated by the dots, with its standard deviation represented by the associated error bar. (b) Scores $s_{t,N,M}(\mathbb{X})$ from observations of r_f for each method.



(a)

	No error	**Uniform error**	**Hierarchical MAP**	**Hierarchical full Bayes**	**Embedded discrepancy**
Prediction of $\ell_f - \ell_0$	–305.05	–10.95	7.14	6.39	–28.00
Prediction of r_f	–182.21	18.66	43.19	42.63	–0.27
Prediction of $\epsilon_{\max}$	–92.57	31.77	48.00	47.45	15.53

(b)

FIG. 4: Prediction of the three outputs from observations of $\epsilon_{\max}$. (a) Mean and standard deviation of the normalized posterior distribution $(Y_t(\mathbf{x}_j) - y_t(\mathbf{x}_j))/y_t(\mathbf{x}_j) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-j}$, $1 \leq j \leq 10$, with $y_1 = \epsilon_{\max}$, compared to the true values $y_t(\mathbf{x}_j)$ shifted to 0, shown as crosses for $1 \leq t \leq 3$. The measurement of $\epsilon_{\max}$ is indicated by the dots, with its standard deviation represented by the associated error bar. (b) Scores $s_{t,N,M}(\mathbb{X})$ from observations of $\epsilon_{\max}$ for each method.

The first key observation is the performance improvement gained by incorporating the additional parameters Y_0 and Y_M to represent the model error. Indeed, the true values y_t are clear outliers within the prediction distribution of the *No error* approach (depicted in green on the plots), as evidenced by the significantly negative scores $s_{t,N,M}$, which are much lower than those obtained with methods that account for model error.

Regarding the model error, the two hierarchical approaches (*Hierarchical MAP* and *Hierarchical full Bayes*) yield nearly identical results and significantly outperform the uniform error approach. These methods effectively shift the prediction mean closer to the true values while maintaining a sufficiently accurate variance estimation. This is reflected in the prediction intervals $[\hat{\mu}_{t,N,M}(\mathbf{x}_j) - 2\sqrt{\hat{\mathrm{v}}_{t,N,M}(\mathbf{x}_j)}, \hat{\mu}_{t,N,M}(\mathbf{x}_j) + 2\sqrt{\hat{\mathrm{v}}_{t,N,M}(\mathbf{x}_j)}]$, which contain the true values $y_t(\mathbf{x}_j)$ for all predictions. This conclusion is further supported by the scores associated with these hierarchical approaches, which are consistently higher than those of the other methods.

The embedded discrepancy approach shows less favorable results compared to the hierarchical approach, with true values $y_{\mathbf{x}_j}$ that are often located in the tails of the prediction distributions, as highlighted by the low score (e.g., clear negative scores for the predictions from observations of r_f).

To interpret these results, it is necessary to examine $p(\boldsymbol{\lambda} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$, the pdf of the posterior distribution of $\boldsymbol{\Lambda}$ for the different methods. Since a LOO scheme is applied, a distinct posterior sample $(\boldsymbol{\lambda}_k)_{k=1}^M$ (or $\left(\boldsymbol{\alpha}_k^1 + \boldsymbol{\alpha}_k^2 \boldsymbol{\xi}_r\right)_{\substack{r=1,\dots,R \\ k=1,\dots,M}}$ for the *Embedded discrepancy*) is obtained for each $\mathbf{x}_j$ and each method, except for the *Hierarchical MAP* and *Hierarchical full Bayes* methods, where the samples are identical, as discussed in Section 2.4. Figure 5 presents the marginal distributions of posterior samples obtained for predicting $\mathbf{x}_{10}$ based on measurements of $\ell_f - \ell_0$, but similar results are observed when targeting another $\mathbf{x}_j$ or with a different calibration variable.

3.5.2 Analysis for the Hierarchical Model

In Fig. 5, it clearly appears that the hierarchical model completely drives the additional parameters Y_0 and Y_M to values near 0, which leads to better results. The influence of this approach is amplified by the fact that the observation

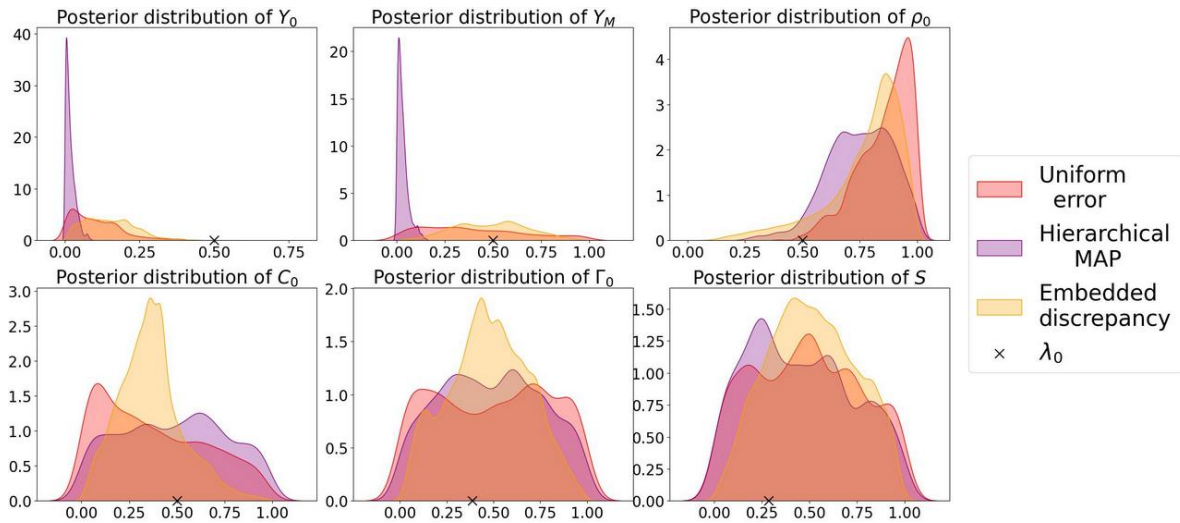


FIG. 5: Distribution of each margin of the posterior samples $(\boldsymbol{\lambda}_k)_{k=1}^M$ obtained with the *Uniform error* and the *Hierarchical MAP* strategies, and the samples $\left(\boldsymbol{\alpha}_k^1 + \boldsymbol{\alpha}_k^2 \boldsymbol{\xi}_r\right)_{\substack{r=1,\dots,R \\ k=1,\dots,M}}$ for the *Embedded discrepancy*, for the prediction at $\mathbf{x}_{10}$ from observations of $\ell_f - \ell_0$. The black cross is associated with $\boldsymbol{\lambda}_0$, the parameters used for the measurement acquisition.

noise is set high, with standard deviations of 0.9 for $\ell_f - \ell_0$ and r_f, and 0.3 for ϵ_{max} [see Figs. 2(a), 3(a), and 4(a)]. Indeed, the larger the noise variance, the less concentrated the likelihood of $\boldsymbol{\lambda}$, thereby increasing the influence of the prior. Appendix E shows the transposition results obtained with smaller observation noises.

The almost identical results between *Hierarchical MAP* and *Hierarchical full Bayes* are explained by the posterior distribution $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ that overall focuses on negative values of $\boldsymbol{\alpha}$ and then drives the normalized $\boldsymbol{\lambda}$ very close to 0, with very small difference compared to the identified $\boldsymbol{\alpha}_\ell^\star$. Appendix D illustrates this posterior distribution for a specific $\mathbf{x}_j$ and $y_1 = \epsilon_{\text{max}}$ [see Fig. D1(a)]. It concentrates the prior $p_{\boldsymbol{\Lambda}}(\cdot \mid \boldsymbol{\alpha})$ on zones near 0 for $\boldsymbol{\alpha} = \boldsymbol{\alpha}_\ell^\star$ and for all the values $\boldsymbol{\alpha}$ that are sampled with $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$. Figure D1(b) also shows the promising asymptotic confidence levels $\gamma(\boldsymbol{\alpha})$ associated with $p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha}) < 1.05 p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star) p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)$, as $\forall \boldsymbol{\alpha}, \gamma(\boldsymbol{\alpha}) > 0.99$.

3.5.3 Analysis for the Embedded Discrepancy

Regarding the performance of the *embedded discrepancy* approach, it is evident that the prediction error consistently occurs in the same direction: $\ell_f - \ell_0$ is overestimated, while r_f and ϵ_{max} are underestimated. This can be attributed to the fact that, as is often the case in the presence of model error, the highest likelihoods of $\boldsymbol{\lambda}$ are located at the boundaries, particularly for Y_0 and ρ_0 (see Fig. 5). The influence of these variables on the outputs can be seen as monotonic, as highlighted by linear regression: increasing Y_0 leads to an increase in $\ell_f - \ell_0$ but a decrease in r_f and ϵ_{max}, and the opposite is true for ρ_0. However, as expected, the *embedded discrepancy* approach is more exploratory in its sampling, considering a wider range of values for Y_0 and ρ_0 away from the boundary. This increases the predicted mean of $\ell_f - \ell_0$ and decreases it for r_f and ϵ_{max}, since no sampling is possible beyond the boundary to balance this exploration.

4. SUMMARY AND PERSPECTIVES

This article addresses the transposition problem, which occurs when observed experimental data do not correspond directly to the data of interest. It examines a scenario with multiple outputs, where only one output is experimentally observed through a limited set of observations, while the goal is to predict all outputs. The main idea is to propose a representation of the model error that is adapted to this situation and thus needs to be embedded in the parameters $\boldsymbol{\Lambda}$ to establish a relationship between the observations and the unobserved outputs. Specifically, we explore the

inclusion in $\mathbf{\Lambda}$ of additional parameters, typically numerical or known physical ones, that were not originally intended for calibration. These parameters are incorporated to account for model error through a hierarchical representation, directing this error into specific zones by introducing hyperparameters $\boldsymbol{\alpha}$ for the prior distribution. The estimation of these hyperparameters is then conducted using Bayesian inference, either by identifying the maximum *a posteriori* $\boldsymbol{\alpha}_{\text{MAP}}$ or by sampling from the posterior distribution, with an importance sampling scheme to significantly reduce the computational cost. This study is conducted in the context of expensive simulations, with Gaussian process regression as a surrogate model.

The application test case of the article models the Taylor cylinder impact test, which involves three outputs considered as observed variables one at a time, leading to three different sets of results. The performance of our method is compared to the approach proposed by [13], which introduces an *embedded discrepancy* method relevant to our scenario. It is important to keep in mind that the results of these Bayesian approaches are highly dependent on the specific application case.

The results obtained here highlight the significant impact of the hierarchical model, as evidenced by the scores computed to assess the predictive distribution in relation to the true values. This impact is somewhat reduced when a small noise measurement variance is considered, although the hierarchical approach remains well suited. In contrast, the *embedded discrepancy* method from [13] performs less effectively in our context, partly due to the fact that the most likely values of $\mathbf{\Lambda}$ are located near the boundaries. Additionally, the Legendre-uniform polynomial chaos (Legendre-uniform PC) representation used in this study could be further modified to potentially enhance these results.

Although the study presented here is comprehensive, further developments should be considered. As explained in Section 3.5, the influence of the prior on $\mathbf{\Lambda}$ is reduced when the noise measurement variance is small, although the results of the hierarchical model remain promising in this situation, as shown in Appendix E. An in-depth study of the relationship between the impact of the hierarchical representation and the noise measurement could be valuable, potentially identifying a threshold on the noise variance below which the likelihood on $\mathbf{\Lambda}$ becomes too concentrated, resulting in a noninfluential prior.

Another important aspect of the work is the surrogate model implemented with Gaussian process regression. Here, as explained in Section 2.5, a different Gaussian process is considered for every $\mathbf{x} \in \mathcal{X}$ and $1 \leq t \leq T$, and they are all considered independent, which is possible with a small number of observations. Other implementations are possible. For instance, one could consider a Gaussian process $Z_t(\mathbf{x}, \boldsymbol{\lambda})$ in the joint space $\mathcal{X} \times \mathcal{L}$, or even multioutput Gaussian processes [35–37] to account for the dependencies between the outputs of the computer code $(f_t)_{t=1}^T$.

Even though the framework presented in this article is general, it has only been tested in the context of a transposition situation involving unobserved outputs. Other transposition cases should be explored, such as situations involving a change of scale where predictions are made at a point $\mathbf{x}_0$ far from the observation points $(\mathbf{x}_j)_{j=1}^n$. Note that the transposition situations can be combined as well, with a change of scale for $\mathbf{x}$ and the prediction of an unobserved output y_t. The method with the hierarchical model and the different performance metrics remains valid in these contexts.

Finally, as previously mentioned, the results of the methods are highly influenced by the application test case. In particular, a computer code without constraints on the calibration variables would be valuable, as it would allow for the consideration of untruncated prior distributions, for instance.

ACKNOWLEDGMENTS

This research was supported by the Consortium in Applied Mathematics CIROQUO, gathering partners in technological and academia in the development of advanced methods for computer experiments (see https://doi.org/10.5281/zenodo.6581217). The numerical model used for the Taylor cylinder impact test was developped by Rémi Chauvin.

SUPPLEMENTARY MATERIAL

Supplementary material is available from Ref. [23]. The “proof/” directory contains the proof of convergence for the estimators discussed in Section 2.4. The “codes/” directory includes the datasets and associated code used to

produce the results presented in the article. As noted in Section 3.5, several experimental designs $\mathbb{X} = (\mathbf{x}_j)_{j=1}^{10}$ were investigated, with results organized into subdirectories named "design_." A readme.md file provides additional details about the structure of the repository.

REFERENCES

1. Garrett, A., New, J., and Chandler, T., Evolutionary Tuning of Building Models to Monthly Electrical Consumption, *ASHRAE Trans.*, **119**:89–100, 2013.
2. Agami Reddy, T.I.M. and Panjapornpon, C., Calibrating Detailed Building Energy Simulation Programs with Measured Data—Part II: Application to Three Case Study Office Buildings (RP-1051), *HVAC&R Res.*, **13**(2):243–265, 2007.
3. Muehleisen, R.T. and Bergerson, J., Bayesian Calibration—What, Why and How, *Proc. of the Int. High Performance Buildings Conf.*, Purdue University, West Lafayette, IN, USA, 2016.
4. Kennedy, M.C. and O'Hagan, A., Bayesian Calibration of Computer Models, *J. R. Stat. Soc.: B*, **63**(3):425–464, 2001.
5. Higdon, D., Kennedy, M., Cavendish, J.C., Cafeo, J.A., and Ryne, R.D., Combining Field Data and Computer Simulations for Calibration and Prediction, *SIAM J. Sci. Comput.*, **26**(2):448–466, 2004.
6. Bayarri, M., Berger, J., Paulo, R., Sacks, J., Kottas, A., and Tu, J., A Framework for Validation of Computer Models, *Technometrics*, **49**:138–154, 2007.
7. Boukouvalas, A., Sykes, P., Cornford, D., and Maruri-Aguilar, H., Bayesian Precalibration of a Large Stochastic Microsimulation Model, *IEEE Trans. Intell. Transp. Syst.*, **15**(3):1337–1347, 2014.
8. Campbell, K., Statistical Calibration of Computer Simulations, *Reliab. Eng. Syst. Saf.*, **91**(10):1358–1363, 2006.
9. O'Hagan, A., Bayesian Inference with Misspecified Models: Inference about What?, *J. Stat. Plann. Inf.*, **143**(10):1643–1648, 2013.
10. Ling, Y., Mullins, J., and Mahadevan, S., Selection of Model Discrepancy Priors in Bayesian Calibration, *J. Comput. Phys.*, **276**:665–680, 2014.
11. Maupin, K.A. and Swiler, L.P., Model Discrepancy Calibration across Experimental Settings, *Reliab. Eng. Syst. Saf.*, **200**:106818, 2020.
12. Oliver, T.A., Terejanu, G., Simmons, C.S., and Moser, R.D., Validating Predictions of Unobserved Quantities, *Comput. Methods Appl. Mech. Eng.*, **283**:1310–1335, 2015.
13. Sargsyan, K., Huan, X., and Najm, H.N., Embedded Model Error Representation for Bayesian Model Calibration, *Int. J. Uncertainty Quantif.*, **9**(4):365–394, 2019.
14. Damblin, G. and Gaillard, P., Bayesian Inference and Non-Linear Extensions of the CIRCE Method for Quantifying the Uncertainty of Closure Relationships Integrated into Thermal-Hydraulic System Codes, *Nucl. Eng. Des.*, **359**:110391, 2020.
15. van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M.G., Vannucci, M., Gelman, A., Veen, D., and Willemsen, J., Bayesian Statistics and Modelling, *Nat. Rev. Methods Primers*, **1**(1):1, 2021.
16. Berkson, J., Bayes' Theorem, *Ann. Math. Stat.*, **1**(1):42–56, 1930.
17. Haario, H., Laine, M., Mira, A., and Saksman, E., DRAM: Efficient Adaptive MCMC, *Stat. Comput.*, **16**:339–354, 2006.
18. Duane, S., Kennedy, A., Pendleton, B.J., and Roweth, D., Hybrid MonteCarlo, *Phys. Lett. B*, **195**(2):216–222, 1987.
19. Kemp, C., Perfors, A., and Tenenbaum, J.B., Learning Overhypotheses with Hierarchical Bayesian Models, *Dev. Sci.*, **10**(3):307–321, 2007.
20. Allenby, G.M. and Rossi, P.E., Hierarchical Bayes Models, in *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*, Thousand Oaks, CA: SAGE Publications, pp. 418–440, 2006.
21. Kloek, T. and van Dijk, H.K., Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo, *Econometrica*, **46**(1):1–19, 1978.
22. Liu, D.C. and Nocedal, J., On the Limited Memory BFGS Method for Large Scale Optimization, *Math. Program.*, **45**(1):503–528, 1989.
23. Sire, C., Supplementary Code Transposition, from https://zenodo.org/records/15582747, 2024.
24. Williams, C.K. and Rasmussen, C.E., *Gaussian Processes for Machine Learning*, Vol. 2, Cambridge, MA: MIT Press, 2006.
25. Genton, M.G., Classes of Kernels for Machine Learning: A Statistics Perspective, *J. Mach. Learn. Res.*, **2**:299–312, 2002.

26. Weiss, N., Holmes, P., and Hardy, M., *A Course in Probability*, Boston: Pearson Addison Wesley, 2005.
27. Maudlin, P., Bingert, J., House, J., and Chen, S., On the Modeling of the Taylor Cylinder Impact Test for Orthotropic Textured Materials: Experiments and Simulations, *Int. J. Plast.*, **15**(2):139–166, 1999.
28. Burshtein, A.I., *Introduction to Thermodynamics and Kinetic Theory of Matter*, John Wiley & Sons, 2008.
29. Mai, J., Peng, L., Lin, Z., and Lai, X., Experimental Study of Electrical Resistivity and Flow Stress of Stainless Steel 316L in Electroplastic Deformation, *Mater. Sci. Eng. A*, **528**(10-11):3539–3544, 2011.
30. Sambasivan, S., Kapahi, A., and Udaykumar, H., Simulation of High Speed Impact, Penetration and Fragmentation Problems on Locally Refined Cartesian Grids, *J. Comput. Phys.*, **235**:334–370, 2013.
31. Kenny, Q.Y., Li, W., and Sudjianto, A., Algorithmic Construction of Optimal Symmetric Latin Hypercube Designs, *J. Stat. Planning Inf.*, **90**(1):145–159, 2000.
32. Gneiting, T. and Raftery, A.E., Strictly Proper Scoring Rules, Prediction, and Estimation, *J. Am. Stat. Assoc.*, **102**(477):359–378, 2007.
33. Miles, P.R., Pymcmcstat: A Python Package for Bayesian Inference Using Delayed Rejection Adaptive Metropolis, *J. Open Source Software*, **4**(38):1417, 2019.
34. Richet, Y. and Have, P., libKriging, from https://github.com/libKriging/libKriging.git, 2023.
35. Le Gratiet, L. and Garnier, J., Recursive Co-Kriging Model for Design of Computer Experiments with Multiple Levels of Fidelity, *Int. J. Uncertainty Quantif.*, **4**(5), pp. 365–386, 2014.
36. Alvarez, M.A., Rosasco, L., and Lawrence, N.D., Kernels for Vector-Valued Functions: A Review, *Found. Trends Mach. Learn.*, **4**(3):195–266, 2012.
37. Gu, M. and Berger, J.O., Parallel Partial Gaussian Process Emulation for Computer Models with Massive Output, *Ann. Appl. Stat.*, **10**(3):1317–1347, 2016.
38. Korolev, V. and Shevtsova, I., An Improvement of the Berry–Esseen Inequality with Applications to Poisson and Mixed Poisson Random Sums, *Scand. Actuarial J.*, **2012**(2):81–105, 2012.
39. Terrell, G.R. and Scott, D.W., Variable Kernel Density Estimation, *Ann. Stat.*, **20**(3):1236–1265, 1992.
40. Wiener, N., The Homogeneous Chaos, *Am. J. Math.*, **60**(4):897–936, 1938.

APPENDIX A. ESTIMATION OF THE CONFIDENCE LEVELS $\gamma(\alpha)$

This appendix complements Section 2.3.2 by examining the confidence level $\gamma(\boldsymbol{\alpha})$ associated with the condition $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}) < \beta p(\boldsymbol{\alpha}_\ell^\star \mid \boldsymbol{y}_{\mathbf{1},\text{obs}})$ for a given $\boldsymbol{\alpha}$, where β represents a user-defined acceptable margin of error for this estimation. The objective is to prove the asymptotic formulation for $\gamma(\boldsymbol{\alpha})$ introduced in Section 2.3.2, using the law of large numbers and a statistical test.

$\gamma(\boldsymbol{\alpha})$ is equivalent to a confidence level for $p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) < \beta p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)$. We introduce

$$\begin{cases} \mu(\boldsymbol{\alpha}) = p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star), \\ v_\beta(\boldsymbol{\alpha}) = \mathbb{V}_{\boldsymbol{\alpha}_\ell^\star}\left(p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\Lambda})\dfrac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda} \mid \boldsymbol{\alpha}_\ell^\star)}\right), \end{cases}$$

where $\boldsymbol{\Lambda}$ has pdf $p_{\boldsymbol{\Lambda}}(. \mid \boldsymbol{\alpha}_\ell^\star)$, and

$$S_{L'}(\boldsymbol{\alpha})^2 = \frac{1}{L'-1}\sum_{k=1}^{L'}\left(p(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\Lambda}_k')\frac{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k' \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k' \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)}{p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}_k' \mid \boldsymbol{\alpha}_\ell^\star)}\right.$$
$$\left. - \left(\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y}_{\mathbf{1},\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)\right)\right)^2,$$

where $(\mathbf{\Lambda}'_k)_{k=1}^{L'}$ are i.i.d. with pdf $p_{\mathbf{\Lambda}}(. \mid \boldsymbol{\alpha}_\ell^\star)$. From the law of large numbers, $S_{L'}(\boldsymbol{\alpha})^2 \xrightarrow[L'\to\infty]{p} v_\beta(\boldsymbol{\alpha})$ and then $S_{L'}(\boldsymbol{\alpha})/\sqrt{v_\beta(\boldsymbol{\alpha})} \xrightarrow[L'\to\infty]{p} 1$. The central limit theorem [38] provides that, given $\boldsymbol{\alpha}_\ell^\star$, with $\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ defined by Eq. (7),

$$\frac{\sqrt{L'}\left(\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star) - \mu(\boldsymbol{\alpha})\right)}{\sqrt{v_\beta(\boldsymbol{\alpha})}} \xrightarrow[L'\to\infty]{d} \mathcal{N}(0,1).$$

Slutsky's theorem then provides

$$\frac{\sqrt{L'}\left(\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star) - \mu(\boldsymbol{\alpha})\right)}{S_{L'}(\boldsymbol{\alpha})} \xrightarrow[L'\to\infty]{d} \mathcal{N}(0,1). \tag{A.1}$$

We introduce the hypothesis of a statistical test $H_0 \ : \ \mu(\boldsymbol{\alpha}) \geq 0$. We consider $\gamma(\boldsymbol{\alpha})$ the complementary of the associated p-value, defined by

$$\begin{aligned}\gamma(\boldsymbol{\alpha}) = 1 - \mathbb{P}\Bigg[&\frac{\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)}{S_{L'}(\boldsymbol{\alpha})} \\ &\leq \frac{\hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta\hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)}{s_{L'}(\boldsymbol{\alpha})}\Bigg],\end{aligned}$$

under $\mu(\boldsymbol{\alpha}) = 0$, with $s_{L'}(\boldsymbol{\alpha})$, $\hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)$ and $\hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ the obtained realizations of the associated estimators. Then, $\gamma(\boldsymbol{\alpha})$ is indeed a confidence level associated with $\mu(\boldsymbol{\alpha}) < 0$, the alternative hypothesis. Under $\mu(\boldsymbol{\alpha}) = 0$, with Eq. (A.1), we have for $z \in \mathbb{R}$,

$$\mathbb{P}\left[\frac{\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta\hat{P}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)}{S_{L'}(\boldsymbol{\alpha})} \leq \frac{z}{\sqrt{L'}}\right] \xrightarrow[L'\to\infty]{} \Phi(z),$$

where Φ is the cdf of a Gaussian distribution with mean 0 and variance 1.

Finally, with

$$z = \frac{\sqrt{L'}\Big(\hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha}) - \beta\hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star)p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star)\Big)}{s_{L'}(\boldsymbol{\alpha})},$$

we have the following asymptotic approximation for the confidence level:

$$\gamma(\boldsymbol{\alpha}) = \Phi\left(\frac{\sqrt{L'}\Big(\beta\hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}_\ell^\star))p_{\mathbf{A}}(\boldsymbol{\alpha}_\ell^\star) - \hat{p}_{L'}^{\boldsymbol{\alpha}_\ell^\star}(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})p_{\mathbf{A}}(\boldsymbol{\alpha})\Big)}{s_{L'}(\boldsymbol{\alpha})}\right). \tag{A.2}$$

APPENDIX B. PRIOR FLUCTUATIONS OF THE OUTPUTS

As mentioned in Section 1, the fluctuations of the additional parameters must introduce sufficient variations in the outputs to adequately capture the model error. The objective here is to investigate, for our application test case, the amplitude of the variations in the three outputs when the calibration parameters $\mathbf{\Lambda}$ vary according to their prior distribution. This analysis aims to confirm the applicability of our strategy in this context and to provide insights into the appropriate standard deviation for the prior distribution.

In Fig. B1, these variations are analyzed for prior distributions with different standard deviations. The idea is to visualize the distribution of $f_t(\mathbf{x}, \mathbf{\Lambda})$ where $\mathbf{\Lambda}$ has density $p_{\mathbf{\Lambda}}^\sigma(\boldsymbol{\lambda}) = \mathbb{1}_{[0,1]^p}(\lambda_1, \ldots, \lambda_p) \times f_{\mathcal{N}_t(0.5,\sigma^2,0,1)}(\lambda_{p+1}, \ldots, \lambda_q)$

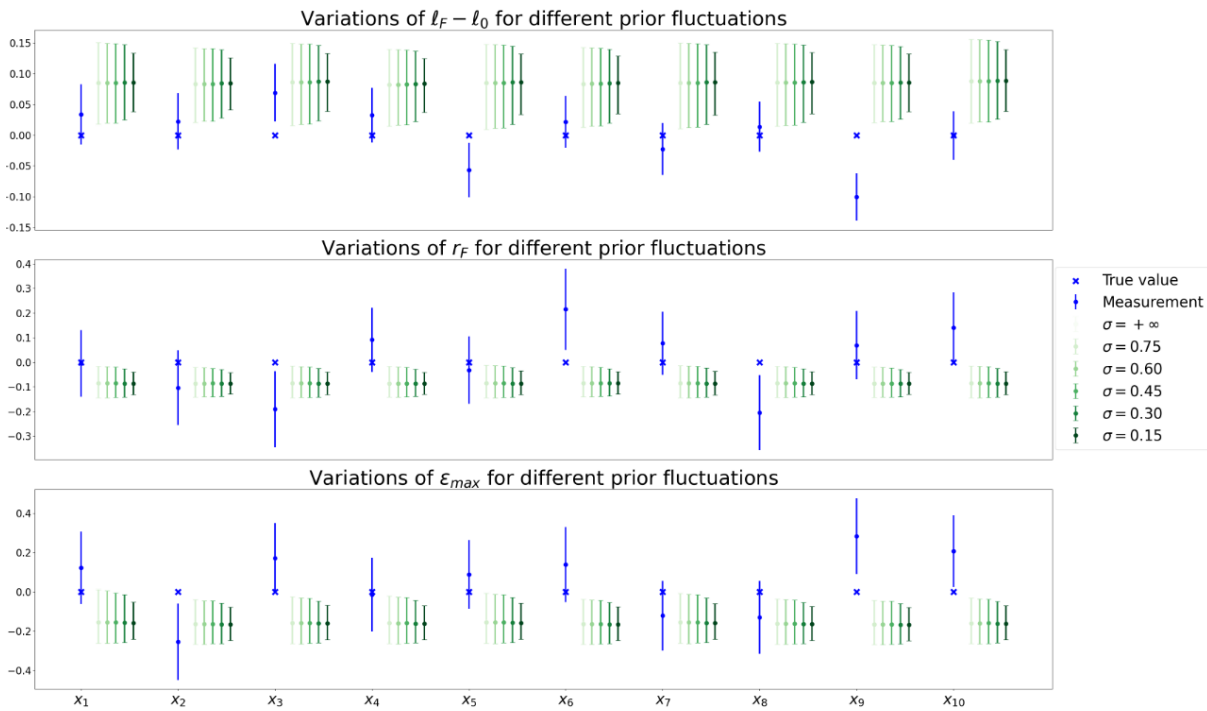


FIG. B1: Variations in the output variables due to the prior fluctuations of the parameters $\mathbf{\Lambda}$

for different values of σ, where $\mathcal{N}_t(0.5, \sigma^2, 0, 1)$ is a Gaussian distribution with mean 0.5 and standard deviation σ, truncated between 0 and 1. More precisely, i.i.d. samples $(\mathbf{\Lambda}_k)_{k=1}^{10^4}$ are generated with density $p^{\sigma}_{\mathbf{\Lambda}}(\mathbf{\lambda})$, and the interval $[y^{0.025}_{t,\sigma}(\mathbf{x}), y^{0.975}_{t,\sigma}(\mathbf{x})]$ is displayed to study the fluctuations of the outputs, where $y^{0.025}_{t,\sigma}(\mathbf{x})$ and $y^{0.975}_{t,\sigma}(\mathbf{x})$ are the empirical quantiles of levels 0.025 and 0.975 of the sample $(f_t(\mathbf{x}, \mathbf{\Lambda}_k))_{k=1}^{10^4}$.

The objective is to identify standard deviations that are too small to capture the model discrepancy when using centered prior distributions. In Fig. B1, the case $\sigma = +\infty$ corresponds to uniform sampling in $[0, 1]^q$, which is naturally the most exploratory scenario. This confirms that our strategy is applicable here, as the variations in the outputs are large enough to account for the model discrepancy.

Moreover, the exploration achieved with uniform sampling appears nearly identical to that obtained with prior distributions using $\sigma = 0.75$, $\sigma = 0.60$, and $\sigma = 0.45$. In contrast, smaller values such as $\sigma = 0.30$ and $\sigma = 0.15$ lead to reduced output variations, which do not allow for generating simulations sufficiently close to the true values of the phenomenon.

To further investigate the choice of σ, we introduce the function

$$g^j_t(\sigma) = \inf_{y \in [y^{0.025}_{t,\sigma}(\mathbf{x}_j), y^{0.975}_{t,\sigma}(\mathbf{x}_j)]} |y_t(\mathbf{x}_j) - y|, \quad \text{for } 1 \leq j \leq n.$$

This function measures how the variations in the output $f_t(\mathbf{x}_j, \mathbf{\Lambda})$ allow for getting closer to the true value $y_t(\mathbf{x}_j)$ when $\mathbf{\Lambda}$ is sampled according to the density $p^{\sigma}_{\mathbf{\Lambda}}$. To obtain a global metric, we define the aggregated function:

$$g_t(\sigma) = \frac{1}{n} \sum_{j=1}^{n} g^j_t(\sigma).$$

Analyzing the evolution of g_t as σ increases in Fig. B2, we observe that small values of σ do not provide sufficient exploration to approximate the true values of the phenomenon. This results in an initial regime where g_t decreases sharply until $\sigma \approx 0.45$. Beyond this threshold, the gain in exploration becomes less significant, leading to a regime

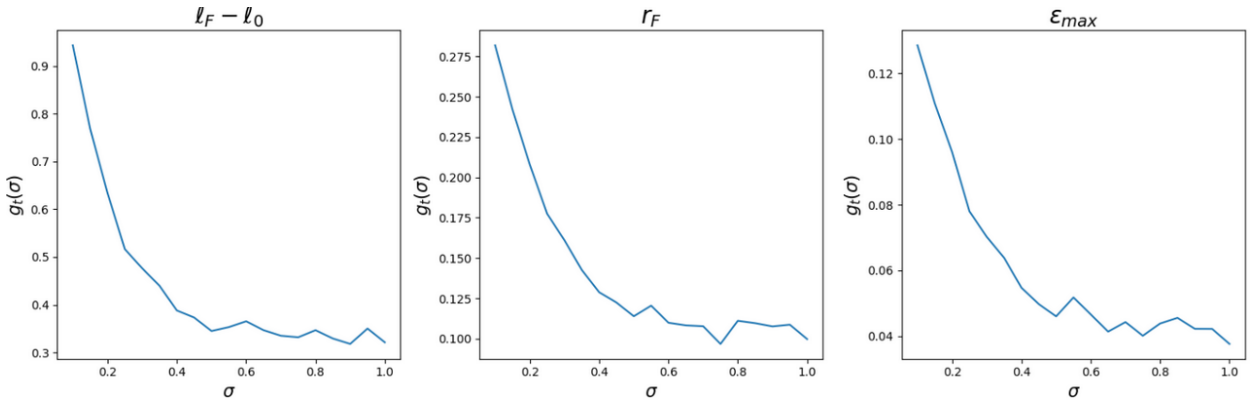


FIG. B2: Evolution of $g_t(\sigma)$ for the three outputs

close to a plateau. This observation supports the choice of a standard deviation around $\sigma = 0.45$. Note that the function g_t is not monotonic here due to sampling errors.

APPENDIX C. EMBEDDED DISCREPANCY

This section provides a brief summary of the work of Ref. [13], which proposes another type of embedded error applicable in our transposition context. Since this method has been implemented in our work for comparison with our numerical results, these details are essential for understanding the fundamentals of their approach.

Their idea is to consider an embedded random discrepancy δ, parameterized by $\boldsymbol{\alpha}^2$. In the following, this discrepancy is denoted $\delta(\boldsymbol{\alpha}^2, \boldsymbol{\Xi})$, where $\boldsymbol{\Xi}$ highlights explicitly the stochastic dimension.

This formulation leads to considering the random output $f_t(\mathbf{x}, \boldsymbol{\alpha}^1 + \delta(\boldsymbol{\alpha}^2, \boldsymbol{\Xi}))$, where $\boldsymbol{\Lambda} = \boldsymbol{\alpha}^1 + \delta(\boldsymbol{\alpha}^2, \boldsymbol{\Xi})$ is a random variable parametrized by $\boldsymbol{\alpha}^1$ and $\boldsymbol{\alpha}^2$. The problem is treated as a Bayesian estimation of the parameters $\boldsymbol{\alpha}^1$ and $\boldsymbol{\alpha}^2$, that are parameters of a pdf. In the following, we denote $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2)$ the set of augmented parameters. Then, we consider the GP formulation

$$Y_t(\mathbf{x}) = \hat{f}_t(\mathbf{x}, \mathbf{A}^1 + \delta(\mathbf{A}^2, \boldsymbol{\Xi})) + \sqrt{v_1(\mathbf{x}, \mathbf{A}^1 + \delta(\mathbf{A}^2, \boldsymbol{\Xi}))} E^f,$$

with $\mathbf{A}^1 \in \mathbb{R}^q,\ \mathbf{A}^2 \in \mathbb{R}^q$.

APPENDIX C.1 Likelihood Estimation

We need to investigate $p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha})$ the likelihood of $\boldsymbol{\alpha}$. We first define the random vectors

$$\begin{cases} \hat{\mathbf{F}}(\boldsymbol{\alpha}) = \left(\hat{f}_1(\mathbf{x}_j, \boldsymbol{\alpha}^1 + \delta(\boldsymbol{\alpha}^2, \boldsymbol{\Xi}))\right)_{j=1}^n, \\ \mathbf{H}(\boldsymbol{\alpha}) = \left(\hat{f}_1(\mathbf{x}_j, \boldsymbol{\alpha}^1 + \delta(\boldsymbol{\alpha}^2, \boldsymbol{\Xi})) + \sqrt{v_1(\mathbf{x}, \boldsymbol{\alpha}^1 + \delta(\boldsymbol{\alpha}^2, \boldsymbol{\Xi}))} E^f + E_j\right)_{j=1}^n. \end{cases}$$

Then we have

$$p(\boldsymbol{y_1}_{,\text{obs}} \mid \boldsymbol{\alpha}) = \pi_{\mathbf{H}(\boldsymbol{\alpha})}(\boldsymbol{y_1}_{,\text{obs}}),$$

with $\pi_{\mathbf{H}(\boldsymbol{\alpha})}$ the pdf of $\mathbf{H}(\boldsymbol{\alpha})$. One challenge is to evaluate $\pi_{\mathbf{H}(\boldsymbol{\alpha})}$ for every $\boldsymbol{\alpha}$. A high-dimensional kernel density estimation (KDE, [39]) would provide a precise computation but is too costly in our context. Several likelihood approximations are proposed in [13], and we have decided to work with an independent normal approximation here, for simplicity reasons. The idea is to get R samples $(\boldsymbol{\xi}_r)_{r=1}^R$ that are i.i.d. realizations of the random variable $\boldsymbol{\Xi}$ and then compute for $1 \leq j \leq n$,

$$\begin{cases} \hat{\mu}_j = \dfrac{1}{R}\displaystyle\sum_{r=1}^{R} \hat{f}_1(\mathbf{x}_j, \boldsymbol{\alpha}^1 + \delta(\boldsymbol{\alpha}^2, \boldsymbol{\xi}_r)), \\ \hat{\sigma}_j^2 = \dfrac{1}{R-1}\displaystyle\sum_{r=1}^{R} (\hat{f}_1(\mathbf{x}_j, \boldsymbol{\alpha}^1 + \delta(\boldsymbol{\alpha}^2, \boldsymbol{\xi}_r)) - \hat{\mu}_j)^2 + \dfrac{1}{R}\displaystyle\sum_{r=1}^{R} \nu_1(\mathbf{x}_j, \boldsymbol{\alpha}^1 + \delta(\boldsymbol{\alpha}^2, \boldsymbol{\xi}_r)) + \sigma_\varepsilon^2. \end{cases}$$

Finally, with the independent normal approximation, the pdf $\pi_{\mathbf{H}(\boldsymbol{\alpha})}$ is defined as

$$\pi_{\mathbf{H}(\boldsymbol{\alpha})}(\mathbf{y}) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\hat{\sigma}_j} \exp\left(-\frac{(y_j - \hat{\mu}_j)^2}{2\hat{\sigma}_j^2}\right).$$

APPENDIX C.2 Polynomial Chaos Representation

The random discrepancy $\delta(\boldsymbol{\alpha}^2, \boldsymbol{\Xi})$ needs to be parametrized. As presented in [13], we have opted for a polynomial chaos (PC, [40]) representation, and more precisely, Legendre-uniform PC that can respect the support of the calibration parameters. It gives

$$\boldsymbol{\alpha}^1 + \delta(\boldsymbol{\alpha}^2, \boldsymbol{\Xi}) = \boldsymbol{\alpha}^1 + \mathrm{diag}(\boldsymbol{\alpha}^2 \boldsymbol{\Xi}^T),$$

with $\boldsymbol{\Xi}_i \sim \mathcal{U}(-1, 1),\ 1 \leq i \leq q$ and $\boldsymbol{\Xi}_i \perp\!\!\!\perp \boldsymbol{\Xi}_j$ if $i \neq j$.

APPENDIX C.3 Prior Construction

The choice of the prior is here driven by the respect of the support of the parameters once again. We have opted for the uniform prior that respects the constraints $\boldsymbol{\alpha}_i^1 - \boldsymbol{\alpha}_i^2 \geq 0$ and $\boldsymbol{\alpha}_i^1 + \boldsymbol{\alpha}_i^2 \leq 1$, and $\boldsymbol{\alpha}_i^2 \geq 0$ as well, to avoid bimodal distributions. We then have

$$p_{(\mathbf{A}^1, \mathbf{A}^2)}(\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2) \propto \prod_{i=1}^{q} \mathbb{1}_{\boldsymbol{\alpha}_i^2 > 0} \mathbb{1}_{\boldsymbol{\alpha}_i^1 - \boldsymbol{\alpha}_i^2 > 0} \mathbb{1}_{\boldsymbol{\alpha}_i^1 + \boldsymbol{\alpha}_i^2 < 1}.$$

APPENDIX C.4 Computation Cost

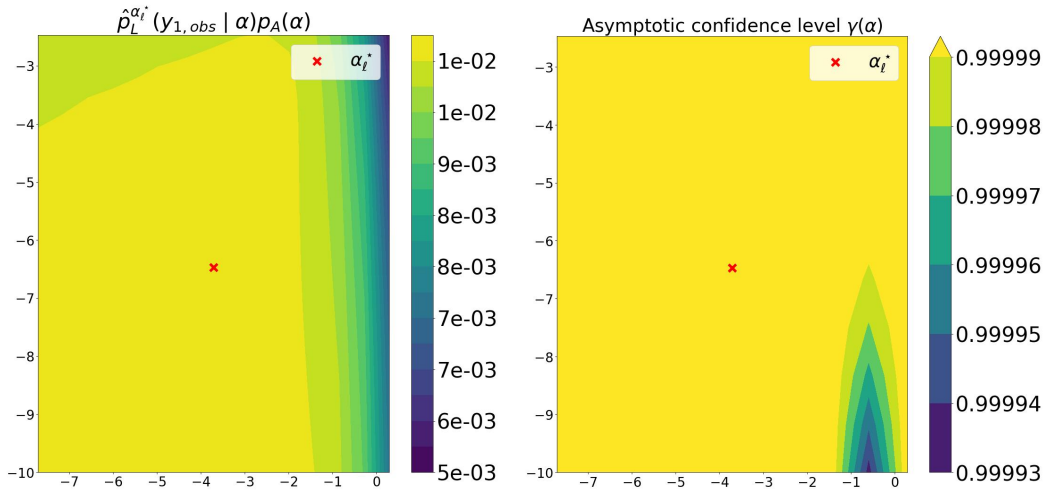
This strategy is relevant in our context but involves a high computation cost compared to the method presented in this article. Indeed, computing the likelihood for a single $\boldsymbol{\alpha}$ requires $R \times n$ calls to the simulator (or the surrogate model), compared to n calls in our method. This is a major difference, as R is associated with sampling in dimension q and thus must be large.

APPENDIX D. POSTERIOR DISTRIBUTION AND MAP FOR $\boldsymbol{\alpha}$

This appendix provides details about the posterior distribution of $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}})$ in the application test case. Figure D1(a) presents the estimated unnormalized posterior distribution $\hat{p}_L^{\boldsymbol{\alpha}_\ell^{j\star}}(\boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}} \mid \boldsymbol{\alpha}) p_{\mathbf{A}}(\boldsymbol{\alpha})$ for $j = 6$, derived from observations of ϵ_{max}. It highlights that this distribution is concentrated on negative values of $\boldsymbol{\alpha}$. Figure D1(b) provides the asymptotic confidence levels $\gamma(\boldsymbol{\alpha})$ associated to $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}}) < 1.05 p(\boldsymbol{\alpha}_\ell^{j\star} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}})$, estimated to validate that the posterior pdf at $\boldsymbol{\alpha}_\ell^{j\star}$ is close to the maximum posterior density across all $\boldsymbol{\alpha} \in \mathcal{A}$. Here, the confidence levels are promising with $\forall \boldsymbol{\alpha}, \gamma(\boldsymbol{\alpha}) > 0.99$.

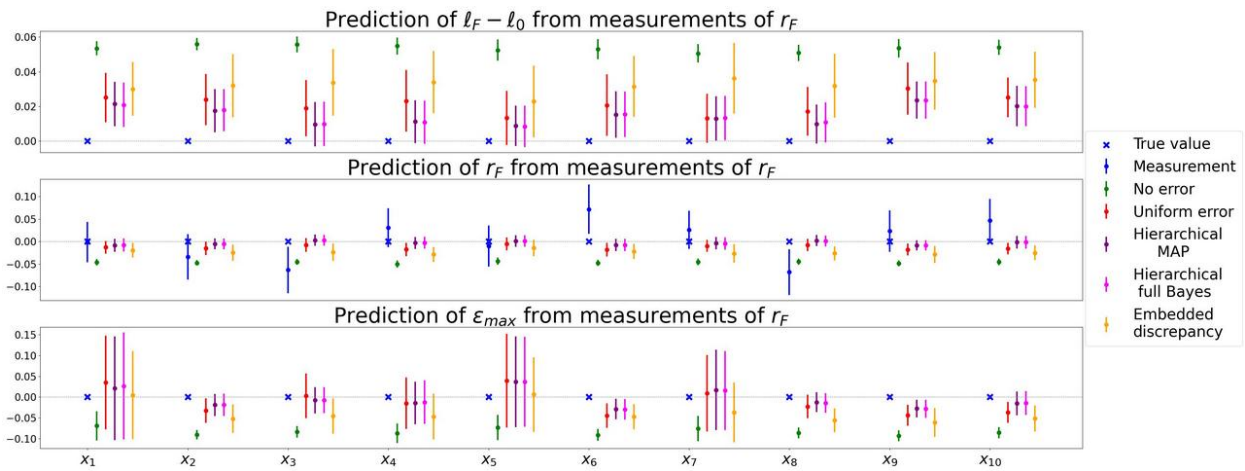
APPENDIX E. EXAMPLE OF RESULTS WITH SMALLER NOISE

This appendix examines a scenario where the observation noise variance is reduced compared to the results shown in Section 3.5. Figure E1 shows the results obtained from measurements of r_f with a noise standard deviation of 0.3, instead of 0.9 in the core of the article. The difference between the *Uniform error* approach and the two hierarchical model–based methods is less pronounced, as the prior has a reduced influence.



(a) (b)

FIG. D1: Illustration of the estimation of the posterior distribution $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}})$ (a) and the confidence level $\gamma(\boldsymbol{\alpha})$ associated to $p(\boldsymbol{\alpha} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}}) < 1.05p(\boldsymbol{\alpha}_\ell^{j\star} \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}})$ (b) with $y_1 = \epsilon_{\text{max}}$ for the predication at $\mathbf{x}_6$



(a)

	No error	Uniform error	Hierarchical MAP	Hierarchical full Bayes	Embedded discrepancy
Prediction of $\ell_f - \ell_0$	–1298.05	1.15	9.72	9.47	–13.62
Prediction of r_f	–385.63	38.98	48.55	48.61	24.42
Prediction of ϵ_{max}	–235.49	41.72	49.02	48.85	34.97

(b)

FIG. E1: Prediction of the three outputs from observations of r_f, with a standard deviation of 0.3 for the measurement noise. (a) Mean and standard deviation of the normalized posterior distribution $(Y_t(\mathbf{x}_j) - y_t(\mathbf{x}_j))/y_t(\mathbf{x}_j) \mid \boldsymbol{y}_{\mathbf{1},\text{obs}}^{-\boldsymbol{j}}$, $1 \leq j \leq 10$, with $y_1 = r_f$, compared to the true values $y_t(\mathbf{x}_j)$ shifted to 0 shown as crosses for $1 \leq t \leq 3$. The measurement of r_f is indicated by the dots, with its standard deviation represented by the associated error bar. (b) Scores $s_{t,N,M}(\mathbb{X})$ from observations of r_f for each method.