



HAL
open science

BAYESIAN CALIBRATION WITH ADAPTIVE MODEL DISCREPANCY

Nicolas Leoni, Olivier Le Maitre, Maria-Giovanna Rodio, Pietro Marco
Congedo

► **To cite this version:**

Nicolas Leoni, Olivier Le Maitre, Maria-Giovanna Rodio, Pietro Marco Congedo. BAYESIAN CALIBRATION WITH ADAPTIVE MODEL DISCREPANCY. *International Journal for Uncertainty Quantification*, 2024, 14 (1), pp.19-41. 10.1615/Int.J.UncertaintyQuantification.2023046331 . hal-04323574v2

HAL Id: hal-04323574

<https://inria.hal.science/hal-04323574v2>

Submitted on 5 Dec 2023

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

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



Distributed under a Creative Commons Attribution 4.0 International License

BAYESIAN CALIBRATION WITH ADAPTIVE MODEL DISCREPANCY

Nicolas Leoni,^{1,2,} Olivier Le Maître,³ Maria-Giovanna Rodio,² & Pietro Marco Congedo¹*

¹*Inria, Centre de Mathématiques Appliquées, Ecole polytechnique, IPP, Route de Saclay, 91128 Palaiseau Cedex, France*

²*Commissariat à l’Energie Atomique et aux Energies Alternatives, ISAS, D36, 91191 Gif-sur-Yvette Cedex, France*

³*CNRS, Inria, Centre de Mathématiques Appliquées, Ecole polytechnique, IPP, Route de Saclay, 91128 Palaiseau Cedex, France*

*Address all correspondence to: Nicolas Leoni, Inria, Centre de Mathématiques Appliquées, Ecole polytechnique, IPP, Route de Saclay, 91128 Palaiseau Cedex, France, E-mail: nicolas.leoni@inria.fr

Original Manuscript Submitted: mm/dd/yyyy; Final Draft Received: mm/dd/yyyy

We investigate a computer model calibration technique inspired by the well-known Bayesian framework of Kennedy and O’Hagan. We tackle the full Bayesian formulation where model parameter and model discrepancy hyperparameters are estimated jointly and reduce the problem dimensionality by introducing a functional relationship that we call the Full Maximum a Posteriori (FMP) method. This method also eliminates the need for a true value of model parameters that caused identifiability issues in the KOH formulation. When the joint posterior is approximated as a mixture of Gaussians, the FMP calibration is proved to avoid some pitfalls of the KOH calibration, namely missing some probability regions and underestimating the posterior variance. We then illustrate two numerical examples where both model error and measurement uncertainty are estimated together. Using the solution to the full Bayesian problem as a reference, we show that the FMP results are accurate, robust and avoid the need for high-dimensional Markov Chains for sampling.

KEY WORDS: *Uncertainty Quantification, Bayesian calibration, Model error, Model discrepancy, Identifiability*

1. INTRODUCTION

Models are widely employed to understand, interpret, and predict phenomena. As the phenomenon of interest is increasingly complex, the models must be more elaborated and, eventually, have to be solved numerically using computers. The software used to solve a scientific model using numerical techniques is called a computer model or a computer code. Computer codes have become central to scientific work, thanks to the increasing computing power available and their cost, usually lower than full-scale experiments. The accuracy and reliability of computer code predictions are consequently of paramount importance today [1].

Models usually involve parameters that are not perfectly known. *Calibration* procedures aim to estimate these parameters using experimental observations. Bayesian techniques are well suited to the calibration task, as they offer a natural framework for plausible reasoning [2] and enable the incorporation of expert knowledge. This knowledge can take many forms, like favoring particular values of the parameters using an informative prior distribution or performing posterior predictive checks to ensure the statistical assumptions led to physically acceptable predictions.

All models are imperfect representations of reality; they involve *model error*, or *model discrepancy*, affecting the quality of predictions. The pioneering article of Kennedy and O’Hagan (KOH) [3] was the first to introduce model discrepancy in the calibration process. Their Bayesian framework represents the model discrepancy with Gaussian processes. It has been widely applied in various fields such as aerodynamics [4–7], fluid mechanics [8,9] and solid mechanics ([10,11]). Compared to standard calibration, inferring the model discrepancy jointly with the parameters yields more objective and less informative parameter posteriors. The work of Kennedy and O’Hagan [3] has triggered a whole body of calibration literature dealing with computer models with high-dimensional outputs [12–14] and with computer model validation [15,16]. The nature of model discrepancy and its treatment in the calibration process was also extensively discussed in [17–19]. For instance, different calibration methods with model discrepancy were recently proposed in [20], [21] and [22].

Although the calibration with model discrepancy has encountered tremendous successes, multiple questions remain unanswered. One question concerns the choice of representation of the model error. A Gaussian process is added to the model output in KOH to represent the discrepancy. Other works use constant bias, random walks [23], or deterministic functions [24]. For Gaussian process models, the user must provide the prior mean, the covariance function structure, and the prior of its hyperparameter. Setting these quantities is not an obvious task and may significantly impact the calibration results. Another approach to calibration in the presence of model error embeds representations to add uncertainties in the model parameters [25,26]. **The frequentist point of view is taken in [27,28] permitting to cure a potential L_2 -inconsistent calibration of the original KOH method.**

1 In twenty years of applications of the Bayesian calibration framework, two classes of methods have emerged:
2 fully Bayesian methods and modular methods. The fully Bayesian methods aim at solving the complete Bayesian
3 formulation of the problem accounting for all uncertainties related to the data, the surrogate for the computer model
4 if necessary, and the estimation of the model discrepancy. Although theoretically superior to others, the resolution of
5 the full Bayesian problems is seldom practical due to their complexity and computational cost [12], and numerical
6 approximation techniques are necessary to solve these problems in practice [13]. The modular methods tackle the
7 complexity of the calibration by introducing simplified formulations of the fully Bayesian problem. Typically, they
8 rely on sequential pointwise estimations of hyperparameters. These strategies significantly reduce the sample spaces'
9 dimensionality, improving the whole computational efficiency of the calibration. The authors of [29], who coined
10 the name, also claimed a theoretical advantage of modular approaches, as the separated sequential estimations avoid
11 confounding issues.

12 The KOH framework is a modular approach based on a pointwise estimation of the model discrepancy before
13 estimating the model parameters. As a result, the method underestimates the uncertainty related to the model dis-
14 crepancy. Although generally considered a "second-order uncertainty," recent studies [19] show that underestimating
15 uncertainties in the model discrepancy might impact the posterior predictions, which are of utmost importance to
16 assess the calibration quality and predict new values with objective confidence intervals. Therefore, there is a need for
17 new modular techniques that approximate the fully Bayesian calibration problem better while retaining the advantage
18 of reduced sample spaces. **In this paper, we propose a new calibration method, named Full Maximum a Posteriori
19 (FMP), which involves a parametric estimation of the model discrepancy hyperparameters. Specifically, the model
20 error is explicitly dependent on model parameter values in our calibration framework. The resulting predictive uncer-
21 tainty split into two contributions: the calibrated model error and the residual uncertainty. The FMP method leads to
22 simple approximations of the full Bayesian parameters and predictive posterior distributions, for reduced cost com-
23 pared to the full Bayesian inference. However, the FMP method corrects the defects of the KOH estimations when
24 the posterior of the hyper-parameters has several modes.**

25 The organization of the paper is as follows. In section 2 we present the Bayesian calibration problem, introducing
26 the general framework with model discrepancy in section 2.1. Sections 2.2 and 2.3 discuss the fully Bayesian and
27 modular methods, respectively. In section 3, we introduce the FMP method, discussing its underlying hypotheses and
28 providing the expressions of the resulting parameters' posterior and posterior predictive density. We show that the
29 FMP estimation of the parameter has higher plausibility than any other modular approach. In section 3.3, we contrast
30 the calibration results for the KOH and FMP for analytic examples consisting of a Gaussian or a mixture of Gaussians

1 for the full Bayesian posterior. The superiority of the FMP method, compared to the KOH approach, is demonstrated
 2 in these analytical examples. In section 4, we consider more complex examples. The first one uses a simple model
 3 having two sets of parameter values yielding a comparable explanation of the observations. The second example
 4 concerns the calibration of a boiling model on experimental measurements. The FMP method proves superior to the
 5 KOH method in approaching the full Bayesian solution in the two examples.

6 **2. BAYESIAN CALIBRATION OF COMPUTER CODES**

7 **2.1 General framework and reference solution**

8 The first requirement for parameter calibration is to specify a statistical model to explain the observations of a quantity
 9 y . We restrict ourselves to a scalar observed quantity, $y \in \mathbb{R}$, of which we have n observations collected in the
 10 observations vector $\mathbf{y}_{\text{obs}} \in \mathbb{R}^n$. The observations consist of measurements for different experimental conditions; a
 11 particular experimental condition is described by a vector $\mathbf{x} \in \mathbb{R}^d$. For instance, in an experiment with fluids flowing
 12 through a pipe, one may measure the pressure loss along the pipe for different conditions \mathbf{x} , consisting of the flow
 13 rate, the fluid density, temperature, . . . We denote $\{\mathbf{x}_i\}_{1 \leq i \leq n}$ the set of experimental conditions corresponding to the
 14 observations in \mathbf{y}_{obs} .

15 Let f be a computer model that provides predictions of y . The inputs for f includes the conditions \mathbf{x} and other
 16 model parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. The latter are usually imperfectly known, and the goal of the calibration procedure
 17 is to learn their value from the observation \mathbf{y}_{obs} . In the example, $\boldsymbol{\theta}$ might consists of the coefficients of an empirical
 18 friction model.

19 In practice, the observations involve a measurement error $\epsilon(\mathbf{x})$. The simplest observation model relates the model
 20 predictions to the observations through the additive form

$$\mathbf{y}_{\text{obs}}(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\theta}) + \epsilon(\mathbf{x}). \quad (1)$$

21 A statistical model for the measurement error is needed. For simplicity, we shall assume that the measurement error
 22 does not depend on the condition \mathbf{x} and that the errors of different observations are independent. For a Gaussian
 23 distribution of the error we have $\epsilon(\mathbf{x}_i) \sim N(0, \sigma_\epsilon^2)$, where σ_ϵ^2 is the mean-squared measurement error. These settings
 24 are standard and used in many calibration studies, such as [30].

25 As discussed below, a Bayesian procedure can be employed to learn the parameters $\boldsymbol{\theta}$ with the observation model
 26 in (1). The procedure updates the prior distribution of $\boldsymbol{\theta}$ to obtain its posterior distribution and, subsequently, posterior

1 predictions using the computer code. The Bayesian procedure was successfully applied to numerous calibration prob-
 2 lems. However, there are situations where it yields inconsistent results, with residuals (differences between posterior
 3 predictions and observations) that do not follow the assumed measurement error distribution. Such inconsistencies
 4 may have two causes: an incorrect measurement error distribution and an observation model that is too poor. This
 5 work considers situations with adequate measurement error distribution and focuses on the observations model. For
 6 instance, the residuals may be highly correlated with significant mean values when the measurement errors are known
 7 to be unbiased and independent for different x_i . **These inconsistencies are interpreted in the work of Kennedy and**
 8 **O'Hagan [3] as a model error describing the *discrepancy between the observed data and the model predictions from***
 9 ***even the best-fitting parameter values.*** The authors proposed to account for the model error by completing the obser-
 10 vation model with a model discrepancy term $z(\mathbf{x})$, leading to the following calibration equation:

$$\mathbf{y}_{\text{obs}}(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\theta}) + z(\mathbf{x}) + \epsilon(\mathbf{x}). \quad (2)$$

11 Note that the original formulation in [3] includes a multiplicative coefficient ρ on the computer model prediction f ,
 12 but the expression in (2) is generally adopted in the calibration literature.

13 More statistical assumptions are needed to proceed with the calibration of the parameters. First, measurement
 14 error and model discrepancy are usually supposed independent: $z(\mathbf{x}_i) \perp\!\!\!\perp \epsilon(\mathbf{x}_i)$. Second, the model discrepancy is
 15 taken as a Gaussian process: $z|\boldsymbol{\psi}_z \sim GP(\mu(\cdot), c(\cdot, \cdot))$, where μ and c are the prior mean and covariance of the
 16 process. The covariance of z depends on some hyperparameters $\boldsymbol{\psi}_z$. For the prior mean, we set $\mu = 0$ as advocated
 17 in [12,31] for better identifiability. However, $\mu(\mathbf{x})$ can be expressed as a linear combination of basis functions in the
 18 context of Universal Kriging [32].

19 We denote $\boldsymbol{\psi}$ the vector of hyperparameters of the entire error model, so that $\boldsymbol{\psi} = (\boldsymbol{\psi}_z, \sigma_\epsilon)$ or $\boldsymbol{\psi} = \sigma_\epsilon$ depending
 20 if model error is considered or not in the calibration. The space of hyperparameters is $\Psi \subset \mathbb{R}^h$.

21 In a Bayesian framework, the posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}_{\text{obs}})$ provides all information about parameters and
 22 hyperparameters after seeing the observations. Its derivation requires the specification of a prior distribution $p(\boldsymbol{\theta}, \boldsymbol{\psi})$
 23 reflecting our knowledge about their values before seeing the observations and of a likelihood function $p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi})$.
 24 Then, the posterior follows the Bayes' Rule:

$$p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}_{\text{obs}}) \propto p(\boldsymbol{\theta}, \boldsymbol{\psi})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi}). \quad (3)$$

1 The log-likelihood function corresponding to the calibration equation without model error (1) is the probability den-
2 sity function of a multivariate normal law with diagonal covariance matrix:

$$\log p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \sigma_{\epsilon}) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma_{\epsilon}^2 - \frac{1}{2\sigma_{\epsilon}^2} \|\mathbf{y}_{\text{obs}} - \mathbf{f}_{\boldsymbol{\theta}}\|^2, \quad (4)$$

3 where $\mathbf{f}_{\boldsymbol{\theta}} = (f(\boldsymbol{\theta}, \mathbf{x}_1), f(\boldsymbol{\theta}, \mathbf{x}_2), \dots, f(\boldsymbol{\theta}, \mathbf{x}_n))^T$ is the vector of evaluations of the computer code at the observation
4 points.

5 The log-likelihood function corresponding to the calibration equation with model discrepancy (2) writes:

$$\log p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi}_z, \sigma_{\epsilon}) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log \det(\boldsymbol{\Sigma}_{\boldsymbol{\psi}} + \sigma_{\epsilon}^2 \mathbf{I}_n) - \frac{1}{2} (\mathbf{y}_{\text{obs}} - \mathbf{f}_{\boldsymbol{\theta}})^T (\boldsymbol{\Sigma}_{\boldsymbol{\psi}} + \sigma_{\epsilon}^2 \mathbf{I}_n)^{-1} (\mathbf{y}_{\text{obs}} - \mathbf{f}_{\boldsymbol{\theta}}), \quad (5)$$

6 where \mathbf{I}_n is the identity matrix of size n and $\boldsymbol{\Sigma}_{\boldsymbol{\psi}}$ is the prior covariance matrix of z : $(\boldsymbol{\Sigma}_{\boldsymbol{\psi}})_{i,j} = c_{\boldsymbol{\psi}_z}(\mathbf{x}_i, \mathbf{x}_j)$.

7 The choice of prior distributions in Bayesian statistics is not straightforward and affects the calibration results.
8 Model parameters and hyperparameters are generally considered independent a priori: $p(\boldsymbol{\theta}, \boldsymbol{\psi}) = p(\boldsymbol{\theta})p(\boldsymbol{\psi})$. Model
9 parameters might be physical quantities with a range of plausible values in the literature. Consequently, it is often
10 possible to formulate informative priors for them. For the hyperparameters of the model discrepancy, classical choices
11 are uniform priors and Jeffrey's prior (e.g., $p(\sigma) \propto 1/\sigma$). **Inverse-gamma priors for correlation lengths l have the**
12 **advantage of ruling-out the cases of vanishing l , inducing a clear separation in the estimation of model error and**
13 **measurement error (assumed uncorrelated).**

14 The marginal posterior distribution of $\boldsymbol{\theta}$ is obtained by integration over the hyperparameters $\boldsymbol{\psi}$:

$$p_{\text{Bayes}}(\boldsymbol{\theta}) := p(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) = \int_{\Psi} p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}_{\text{obs}}) d\boldsymbol{\psi}. \quad (6)$$

15 The posterior predictive density is used to make predictions of the true process at any experimental condition \mathbf{x}_* ,
16 possibly not observed:

$$p(y(\mathbf{x}_*)|\mathbf{y}_{\text{obs}}) = \int_{\Theta} \int_{\Psi} p(y(\mathbf{x}_*)|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y}_{\text{obs}}) p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}_{\text{obs}}) d\boldsymbol{\theta} d\boldsymbol{\psi}, \quad (7)$$

17 The distribution $p(y(\mathbf{x}_*)|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y}_{\text{obs}})$ is normal, obtained from the Gaussian Process predictive equations, which

1 writes [33]:

$$\begin{aligned}
 y(\mathbf{x}_*)|\boldsymbol{\psi}, \boldsymbol{\theta}, \mathbf{y}_{\text{obs}} &\sim \mathcal{N}(\boldsymbol{\mu}_{\text{pred}}, \boldsymbol{\sigma}_{\text{pred}}^2), \\
 \boldsymbol{\mu}_{\text{pred}} &= f(\mathbf{x}_*, \boldsymbol{\theta}) + \mathbf{k}_*^T (\boldsymbol{\Sigma}_{\boldsymbol{\psi}} + \sigma_{\epsilon}^2 \mathbf{I}_n)^{-1} (\mathbf{y}_{\text{obs}} - \mathbf{f}_{\boldsymbol{\theta}}), \\
 \boldsymbol{\sigma}_{\text{pred}}^2 &= c_{\boldsymbol{\psi}}(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (\boldsymbol{\Sigma}_{\boldsymbol{\psi}} + \sigma_{\epsilon}^2 \mathbf{I}_n)^{-1} \mathbf{k}_*.
 \end{aligned} \tag{8}$$

2 with $\mathbf{k}_* = (c_{\boldsymbol{\psi}}(\mathbf{x}_*, \mathbf{x}_1), c_{\boldsymbol{\psi}}(\mathbf{x}_*, \mathbf{x}_2), \dots, c_{\boldsymbol{\psi}}(\mathbf{x}_*, \mathbf{x}_n))^T$.

3 In the following, we refer to the general calibration framework with model error presented above as the *full*
 4 *Bayesian* framework. The distributions defined in equations (6) and (7) constitute the exact solution to the calibration
 5 problem, as was proposed for example in [12,13]. The interest in the full Bayesian calibration comes from its rigorous
 6 derivation and the characterization of the posterior uncertainties in the model parameters and model discrepancy.
 7 Further, the framework is flexible and can adapt to various situations. For instance, in the case of an expensive
 8 computer code requiring its substitution by a surrogate model, the framework can be extended introducing a surrogate
 9 error model with new hyperparameters.

10 Despite its apparent simplicity, this calibration can not be solved precisely in most applications. The reason is
 11 that it requires an accurate estimation of the posterior density, which might exhibit substantial variations and many
 12 local maxima over the joint space $\Theta \times \Psi$. Further, for calibrations involving multiple quantities and experimental con-
 13 figurations, the number of hyperparameters might become significant, making the estimation in high dimensions very
 14 complex. Monte-Carlo techniques [34] and other advanced sampling techniques have been employed to generate sam-
 15 ples from the high dimensional posterior distributions and approximate the required integrals. The following section
 16 discusses some works proposed in the literature on direct full Bayesian calibration. However, the so-called modular
 17 approaches ([29]) have gained popularity as alternatives to full Bayesian calibration. These approaches, discussed in
 18 section 2.3, reduce the dimensionality of the sample space by estimating point values for some hyperparameters.

19 **2.2 Fully Bayesian approaches**

20 An example of application of full Bayesian method is in [12], where the full joint density is sampled using the
 21 Metropolis-Hastings algorithm in dimension six (one parameter and five hyperparameters), allowing a complete rep-
 22 resentation of the posterior uncertainty. In [13], the authors deal with a high-dimensional output code and a sample
 23 space with a dimension greater than 30. Besides, more than 18,000 observations are considered making one evalua-
 24 tion of the likelihood function costly. Consequently, the authors relied on reduced basis representations, block matrix

1 inversion formulas, and adapted informative priors to make the problem tractable and ease the posterior sampling.

2 Another example of a fully Bayesian technique is presented in [15], where the authors chose a covariance function
3 for z that involves variance and correlation length hyperparameters. Before the sampling, correlation lengths are
4 estimated with **Maximum Likelihood (ML)**, as well as any eventual prior mean for z . ML estimates of variance
5 hyperparameters serve to construct a prior for them. The MCMC sampling itself is simplified by allowing only
6 variance hyperparameters to vary (as well as other hyperparameters involved in the surrogate for the computer model),
7 using Gibbs updates as a conjugate prior was chosen. The authors declare that combining ML estimates with sampling
8 achieves an efficient approximation to the Full Bayesian calibration. Note that they claim that ML estimates are a
9 sufficient approximation to Full Bayesian, but only concerning hyperparameters for the computer model surrogate;
10 they recognize that the variations of model discrepancy hyperparameters must be considered. Conjugate priors are
11 also central to the sampling with Gibbs steps proposed in [17].

12 Yuan and Ng [35] approximated the full Bayesian solution using an Expectation-Maximization algorithm, which
13 can be interpreted as a repetition of a modular method of the first class (see next section). In [36], the full posterior
14 distribution was sampled with a Gibbs sampler. In [37], the authors proposed a weighted normal approximation to the
15 posterior to reduce the overall cost of evaluation and accommodate large numbers of hyperparameters.

16 We now discuss two calibration frameworks that account for the uncertainty in model error through its depen-
17 dence on the model parameters. In [20], a "true value" of model parameters is defined as the minimizer of a loss
18 function. It leads to an orthogonality condition for the model discrepancy and the gradient of the computer model.
19 For each θ , a model discrepancy is estimated using a covariance function that verifies the orthogonality condition.
20 This calibration aims to find the "true parameters value", a questionable objective when the model predictions can
21 not satisfactorily represent the observations. Nonetheless, using an adaptive representation of model error instead of
22 a distribution is a way to include multiple model discrepancies that might be very different in structure. In [38], the
23 calibration was recast as an optimization problem, where a single value of θ is sought to maximize the likelihood
24 of the corresponding model discrepancy term. These applications show that the full Bayesian calibration remains of
25 practical interest despite being less popular than modular approaches due to its cost and the difficulties involved in
26 sampling.

27 **2.3 Modular approaches**

28 The principle underlying modular approaches is to separate into groups (modules) the treatment of parameters and
29 hyperparameters during the calibration and obtain a problem of sequential estimations. In [29], three modules are

1 considered in the calibration of computer models: the surrogate of f , the measurement error in the observations,
 2 and the model discrepancy. Each module is then separately estimated. This approach breaks the high-dimensional
 3 calibration problem into smaller ones. One expects that these separated problems are more accessible to sample than
 4 the original problem. The modular approach can also improve the identifiability of the hyperparameters belonging to
 5 different modules.

6 Since the calibration is sequential, the calibration problem for the modules must be adapted, and, in particular,
 7 assumptions regarding subsequent modules not yet calibrated are needed. Since the posterior distribution of the pa-
 8 rameters θ is arguably the goal of the calibration, the last module estimates $p(\theta | \mathbf{y}_{\text{obs}}, \boldsymbol{\psi} = \hat{\boldsymbol{\psi}})$ where the estimator $\hat{\boldsymbol{\psi}}$
 9 of $\boldsymbol{\psi}$ have been estimated in previous modules such that $p(\theta | \mathbf{y}_{\text{obs}}, \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}) \approx p_{\text{Bayes}}(\theta)$. We now distinguish between
 10 two classes of methods proposed in the calibration literature.

11 In the first class of methods, one starts by determining a value θ_0 such that $f(\mathbf{x}, \theta_0)$ is an accurate prediction of
 12 $y(\mathbf{x})$. Then, the hyperparameters are estimated, usually by ML, considering $\theta = \theta_0$. With our notations, this amounts
 13 to the following estimator:

$$\hat{\boldsymbol{\psi}}_1 = \arg \max_{\boldsymbol{\psi}} p(\mathbf{y}_{\text{obs}} | \boldsymbol{\psi}, \theta_0). \quad (9)$$

14 The methods in the first class differ by the definition of θ_0 : some works advocate for elicited values such as the prior
 15 mean [29,39]. Other common approaches estimate θ_0 from the observation, either minimizing the residuals between
 16 the model prediction and the observations [40], or solving the calibration problem without model error (1) to propose
 17 a plausible value [11,19,41,42]. These approaches inherently assess the relevance of introducing a model discrepancy
 18 term by judging the quality of the model predictions using $\theta = \theta_0$. In [43], it is demonstrated that $f(\mathbf{x}, \theta_0) + z(\mathbf{x})$ can
 19 be an accurate prediction of $y(\mathbf{x})$, even for a poor value of θ_0 , because the model discrepancy term z can compensate
 20 for a wide range of residuals.

21 In the second class of methods, one averages the posterior over the model parameters value to obtain the distri-
 22 bution $p(\mathbf{y}_{\text{obs}} | \boldsymbol{\psi})$, and then the hyperparameters are estimated using:

$$\hat{\boldsymbol{\psi}}_2 = \arg \max_{\boldsymbol{\psi}} p(\mathbf{y}_{\text{obs}} | \boldsymbol{\psi}). \quad (10)$$

23 This approach is used for instance in [14,31,44]. It was originally proposed by Kennedy and O'Hagan [3,45], with
 24 also the consideration of the hyperparameter prior, leading to the Maximum A Posteriori (MAP) estimate:

$$\hat{\boldsymbol{\psi}}_{\text{KOH}} = \arg \max_{\boldsymbol{\psi} \in \Psi} p(\boldsymbol{\psi} | \mathbf{y}_{\text{obs}}) = \arg \max_{\boldsymbol{\psi} \in \Psi} p(\boldsymbol{\psi}) \int_{\Theta} p(\theta) p(\mathbf{y}_{\text{obs}} | \theta, \boldsymbol{\psi}) d\theta, \quad (11)$$

1 and the following estimation for the parameters posterior distribution:

$$p_{\text{KOH}}(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\text{KOH}}). \quad (12)$$

2 Note also that an approximation of $p(\mathbf{y}_{\text{obs}}|\boldsymbol{\psi})$ is sometimes used to reduce the computational cost; Kennedy and
 3 O’Hagan propose to use a Gaussian approximation in [45]. Compared to the first class of methods, the computation
 4 of $p(\mathbf{y}_{\text{obs}}|\boldsymbol{\psi})$ is generally more costly than estimating a single parameter value θ_0 . Compared to the first class, the
 5 second class of methods benefits from the complete account of the parameter uncertainty when estimating the model
 6 discrepancy.

7 The two classes of modular methods are compared in [19], where the second class estimates higher model dis-
 8 crepancies, showing the benefit of including the parameter uncertainty. In general, the modular approaches provide
 9 fairly accurate parameter posteriors. Nonetheless, due to the single point estimate for $\boldsymbol{\psi}$, they do not account for the
 10 complete uncertainty about the model discrepancy term. In what follows, we propose a calibration technique where
 11 the uncertainty about the values of $\boldsymbol{\psi}$ is estimated to include all relevant uncertainties in the calibration framework.

12 3. THE FULL MAXIMUM A POSTERIORI METHOD

13 This section introduces the FMP approximation. It starts with its founding assumptions and the resulting approxi-
 14 mations to the distributions of interest, presented in section (3.1). The FMP method is shown to produce marginal
 15 likelihoods greater than any other similar estimation technique in section (3.2). Then, we derive the exact FMP, KOH
 16 and Bayes calibration solutions, in the case of a Gaussian joint posterior in section (3.3.1), and a mixture of Gaussians
 17 with well-separated modes in section (3.3.2).

18 3.1 Calibration framework

19 3.1.1 Fundamental hypotheses

20 We follow the idea that the plausibility of parameter values should be compared by considering their most favorable
 21 model discrepancy distributions. This is done by first recognizing that the model discrepancy term depends on the
 22 model parameter:

$$z_{\boldsymbol{\theta}}(\mathbf{x}) = y(\mathbf{x}) - f(\mathbf{x}, \boldsymbol{\theta}). \quad (13)$$

1 Therefore, the calibration equation becomes:

$$y_{\text{obs}}(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\theta}) + z_{\boldsymbol{\theta}}(\mathbf{x}) + \epsilon(\mathbf{x}). \quad (14)$$

2 This equation appeared previously in the work of [20], but we do not modify the covariance of $z_{\boldsymbol{\theta}}$ as they do. Instead,
 3 we perform a joint estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ like in the full Bayesian framework of [12]. To make the estimation cheaper
 4 we introduce a functional relationship between the parameters and the "optimal hyperparameters" defined as follows

$$\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\psi}} p(\boldsymbol{\psi}) p(\mathbf{y}_{\text{obs}} | \boldsymbol{\theta}, \boldsymbol{\psi}), \quad (15)$$

5 where FMP stands for "Full Maximum A Posteriori", a combination between full Bayesian analysis and MAP esti-
 6 mation. **Two additional assumptions are introduced to derive the approximation of the posterior.**

7 1. **For all $\boldsymbol{\theta} \in \Theta$, the distribution $p(\boldsymbol{\psi} | \boldsymbol{\theta}, \mathbf{y}_{\text{obs}})$ can be approximated as a point mass distribution at its mode**
 8 **$\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})$. This assumption is standard in calibration contexts. For instance, the underlying assumption of**
 9 **KOH's framework is a point mass approximation of the distribution $p(\boldsymbol{\psi} | \mathbf{y}_{\text{obs}})$ at its mode. Here we rely on**
 10 **a point mass approximation of the distributions $p(\boldsymbol{\psi} | \boldsymbol{\theta}, \mathbf{y}_{\text{obs}})$. In this way, we do not consider one value of $\boldsymbol{\psi}$**
 11 **that fits the error for all model predictions (when $\boldsymbol{\theta}$ varies). Instead, the errors for each model prediction can be**
 12 **fitted well by individual values of $\boldsymbol{\psi}$.**

13 2. **We assume that, for all $\boldsymbol{\theta} \in \Theta$ with significant posterior,**

$$\int_{\Theta} p(\boldsymbol{\theta}' | \mathbf{y}_{\text{obs}}) p(\mathbf{y}_{\text{obs}} | \boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}')) d\boldsymbol{\theta}' \approx p(\mathbf{y}_{\text{obs}} | \boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})). \quad (16)$$

14 **This assumption is exploited below to simplify the expression and computation of the FMP posterior. In prac-**
 15 **tice, it holds true for multi-modal posteriors that are not well-separated in both $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. Otherwise, the approx-**
 16 **imation in 16 introduces a bias in the relative strengths of the modes, but retain the global multi-modal structure**
 17 **of the $\boldsymbol{\theta}$'s posterior, in contrast with the KOH method. This is illustrated in the example of section 3.3.2.**

1 3.1.2 Posterior approximation and predictive density

With the previous assumptions, the parameter posterior, named $p_{\text{Bayes}}(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}})$, can be expressed as

$$\begin{aligned}
 p_{\text{Bayes}}(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) &= \int_{\Psi} p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y}_{\text{obs}})p(\boldsymbol{\psi}|\mathbf{y}_{\text{obs}}) d\boldsymbol{\psi} \\
 &= \int_{\Psi} p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y}_{\text{obs}}) \left(\int_{\Theta} p(\boldsymbol{\psi}|\boldsymbol{\theta}', \mathbf{y}_{\text{obs}})p(\boldsymbol{\theta}'|\mathbf{y}_{\text{obs}}) d\boldsymbol{\theta}' \right) d\boldsymbol{\psi} \\
 &= \int_{\Theta} p(\boldsymbol{\theta}'|\mathbf{y}_{\text{obs}}) \left(\int_{\Psi} p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y}_{\text{obs}})p(\boldsymbol{\psi}|\boldsymbol{\theta}', \mathbf{y}_{\text{obs}}) d\boldsymbol{\psi} \right) d\boldsymbol{\theta}' \\
 &\approx \int_{\Theta} p(\boldsymbol{\theta}'|\mathbf{y}_{\text{obs}})p(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}')) d\boldsymbol{\theta}' && \text{(Assumption 1, Sec. 3.1.1)} \\
 &\propto \int_{\Theta} p(\boldsymbol{\theta}'|\mathbf{y}_{\text{obs}})p(\boldsymbol{\theta})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}')) d\boldsymbol{\theta}' && \text{(Bayes Theorem)} \\
 &\propto p(\boldsymbol{\theta})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})) \doteq p_{\text{FMP}}(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) && \text{(Assumption 2, Sec. 3.1.1).} \quad (17)
 \end{aligned}$$

2 We denote with p_{FMP} the last term of Eq. (17), *i.e.* the probability density proportional to $p(\boldsymbol{\theta})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}))$.

3 From the previous equation, p_{FMP} is a reasonable approximation of p_{Bayes} .

4 Note that for most covariance functions, $c_{\boldsymbol{\psi}}(x, x')$ is a continuous function of $\boldsymbol{\psi}$ for all (x, x') . By the Maximum
 5 Theorem [46,47] it can be shown that p_{FMP} is a continuous function of $\boldsymbol{\theta}$. With the additional assumption that for
 6 each value of $\boldsymbol{\theta}$ there is a unique value of optimal hyperparameters (which is not true in general), we have that $\hat{\boldsymbol{\psi}}_{\text{FMP}}$
 7 is a continuous function of $\boldsymbol{\theta}$. These continuity properties pave the way for calibration techniques based on surrogate
 8 modelling of the FMP quantities that will be explored in subsequent work.

The predictive density for the true process at an unobserved experimental condition \mathbf{x}^* is given by:

$$\begin{aligned}
 p(y(\mathbf{x}^*)|\mathbf{y}_{\text{obs}}) &= \int_{\Psi} \int_{\Theta} p(y(\mathbf{x}^*)|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y}_{\text{obs}})p(\boldsymbol{\psi}|\boldsymbol{\theta}, \mathbf{y}_{\text{obs}})p(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) d\boldsymbol{\theta} d\boldsymbol{\psi} \\
 &= \int_{\Theta} p(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) \left(\int_{\Psi} p(y(\mathbf{x}^*)|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y}_{\text{obs}})p(\boldsymbol{\psi}|\boldsymbol{\theta}, \mathbf{y}_{\text{obs}}) d\boldsymbol{\psi} \right) d\boldsymbol{\theta} \\
 &\approx \int_{\Theta} p(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}})p(y(\mathbf{x}^*)|\boldsymbol{\theta}, \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}), \mathbf{y}_{\text{obs}}) d\boldsymbol{\theta}. \quad (18)
 \end{aligned}$$

9 The mean and variance of the predictive density can be computed explicitly. It comes

$$\begin{aligned}
 \mathbb{E}[y(\mathbf{x}_*)|\mathbf{y}_{\text{obs}}] &= \mathbb{E}_{\boldsymbol{\theta}}[\mathbb{E}[y(\mathbf{x}_*)|\mathbf{y}_{\text{obs}}, \boldsymbol{\theta}]] \\
 &= \underbrace{\mathbb{E}_{\boldsymbol{\theta}}[f(\mathbf{x}_*, \boldsymbol{\theta})]}_{\text{averaged model prediction at } \mathbf{x}_*} + \underbrace{\mathbb{E}_{\boldsymbol{\theta}}[\mathbf{k}_*^T (\boldsymbol{\Sigma} + \sigma_{\epsilon}^2 \mathbf{I}_n)^{-1} (\mathbf{y}_{\text{obs}} - \mathbf{f}_{\boldsymbol{\theta}})]}_{\text{averaged model discrepancy at } \mathbf{x}_*}, \quad (19)
 \end{aligned}$$

1 and

$$\begin{aligned}
\text{Var}[y(\mathbf{x}_*)|\mathbf{y}_{\text{obs}}] &= \text{Var}_{\boldsymbol{\theta}}[\mathbb{E}[y(\mathbf{x}_*)|\mathbf{y}_{\text{obs}}, \boldsymbol{\theta}]] + \mathbb{E}_{\boldsymbol{\theta}}[\text{Var}[y(\mathbf{x}_*)|\mathbf{y}_{\text{obs}}, \boldsymbol{\theta}]] \\
&= \underbrace{\text{Var}_{\boldsymbol{\theta}}[f(\mathbf{x}_*, \boldsymbol{\theta}) + \mathbf{k}_*^T (\boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})} + \sigma_{\epsilon}^2 \mathbf{I}_n)^{-1} (\mathbf{y}_{\text{obs}} - \mathbf{f}_{\boldsymbol{\theta}})]}_{\text{uncertainty in the corrected model}} \\
&\quad + \underbrace{\mathbb{E}_{\boldsymbol{\theta}}[c_{\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})}(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (\boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})} + \sigma_{\epsilon}^2 \mathbf{I}_n)^{-1} \mathbf{k}_*]}_{\text{residual uncertainty}}
\end{aligned} \tag{20}$$

2 where $\mathbf{k}_* = (c_{\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})}(\mathbf{x}_*, \mathbf{x}_1), c_{\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})}(\mathbf{x}_*, \mathbf{x}_2), \dots, c_{\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})}(\mathbf{x}_*, \mathbf{x}_n))^T$.

3 Equation (20) provides a decomposition of the predictive error into two contributions. The first contribution,
4 named "uncertainty in the corrected model", is the variability of the corrected predictions. It is significant when the
5 corrected predictions vary with $\boldsymbol{\theta}$, indicating that the model discrepancy term cannot adequately correct the predic-
6 tions. The second term, named "residual uncertainty," is the part of the posterior uncertainty that the corrected model
7 cannot explain. It corresponds to the remaining uncertainty in calibration using a finite amount of noisy observations.

8 The FMP framework avoids ambiguity in defining the "true value" of parameters. Nevertheless, other types of
9 identifiability issues might still plague the calibration process. If two distinct values of $\boldsymbol{\theta}$ with the same prior value
10 lead to the same computer model prediction, one might deem them unidentifiable. Likewise, if the prior distribution of
11 z is not suited to the actual model discrepancy, optimal hyperparameters might be meaningless and posterior samples
12 of z would be inaccurate. These issues can be addressed by testing several prior distributions for $\boldsymbol{\theta}$, z , and $\boldsymbol{\psi}$. The
13 multivariate Gaussian example in section 3.3.2 below gives a graphical representation of these identifiability cases.

14 3.2 Marginal likelihood of the FMP approximation

15 We now examine the relevance of the FMP method in the scope of Bayesian Model Comparison (BMC). For a
16 model M and an observations set \mathbf{y}_{obs} , the *marginal likelihood* $p(\mathbf{y}_{\text{obs}}|M)$ is the probability of observing \mathbf{y}_{obs} given
17 that its generative model is M . Comparing the marginal likelihood values for different models allows for determin-
18 ing which model is the most likely to have generated the observations. In our case, consider weak prior informa-
19 tion on the hyperparameters, the maximum a posteriori estimator simplifies to maximum likelihood: $\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}) =$
20 $\arg \max_{\boldsymbol{\psi}} p(\boldsymbol{\psi})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi}) \approx \arg \max_{\boldsymbol{\psi}} p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi})$.

Assume that, in an alternative approximation method, the functional relationship $\hat{\boldsymbol{\psi}}_h(\boldsymbol{\theta}) = h(\boldsymbol{\theta})$ is proposed,
with h a generic function. The posterior of the model parameters would become $p_h(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) \propto p(\boldsymbol{\theta})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi} =$

$\boldsymbol{\psi}_h(\boldsymbol{\theta})$). This setting also encompasses single-point estimation methods, such as the KOH or cross-validation estimator of the hyperparameters. We have:

$$\begin{aligned} & p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})) \geq p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi} = \boldsymbol{\psi}_h(\boldsymbol{\theta})), \quad \forall \boldsymbol{\theta} \in \Theta \\ \Rightarrow & \int_{\Theta} p(\boldsymbol{\theta})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})) d\boldsymbol{\theta} \geq \int_{\Theta} p(\boldsymbol{\theta})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi} = \boldsymbol{\psi}_h(\boldsymbol{\theta})) d\boldsymbol{\theta} \\ \Leftrightarrow & p(\mathbf{y}_{\text{obs}}|\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_{\text{FMP}}) \geq p(\mathbf{y}_{\text{obs}}|\boldsymbol{\psi} = \boldsymbol{\psi}_h). \end{aligned}$$

1 Thus, the likelihood of the observations when making the FMP approximation will always be higher than any other
 2 functional relationship between hyperparameters and parameters. The BMC will then systematically favor the FMP
 3 method against any other approach based on such a relationship or single-point estimation.

4 3.3 Elementary examples

5 To illustrate the behavior of the FMP calibration, we consider two examples. In the first one, the (Bayesian) joint
 6 posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\psi})$ is Gaussian. In the second one, we assume it is a multimodal mixture of Gaussians.
 7 The Bayes, KOH, and FMP calibration methods feature analytical solutions in these two cases.

8 **A Gaussian posterior is unlikely in practice because priors on hyperparameters are rarely Gaussian but classically**
 9 **follow uniform or inverse Gamma distributions. However, if a *single* value for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ perfectly fits the data, then**
 10 **the posterior is unimodal and a Gaussian approximation might be reasonable.** However, the Gaussian case is very
 11 instructive, and posterior approximations by Gaussian distribution are standard approximation when the posteriors
 12 become increasingly sharp around their maximum as the number of observations increases. **When several, well-**
 13 **separated, values of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ explain the data equally well, the multimodal Gaussian assumption is more appropriate.**
 14 The Gaussian mixture can approximate more complex distributions, and this example helps us understand the case of
 15 alternative explanations of the observations are plausible.

16 3.3.1 Unimodal Gaussian posterior

17 Let us assume that the Bayesian procedure leads to a joint Gaussian posterior for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. We recall that $\dim(\boldsymbol{\theta}) = p$
 18 and $\dim(\boldsymbol{\psi}) = h$. Let $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\psi})$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\mu}_{\boldsymbol{\psi}})$ is the posterior mean and $\boldsymbol{\Lambda} = \begin{pmatrix} \mathbf{V}_{\boldsymbol{\theta}} & \mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}}^T \\ \mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}} & \mathbf{V}_{\boldsymbol{\psi}} \end{pmatrix}$ is the posterior

1 covariance, symmetric and positive definite. The joint posterior distribution writes:

$$p(\boldsymbol{\gamma}|\mathbf{y}_{\text{obs}}) = \frac{1}{(2\pi)^{(p+h)/2}|\boldsymbol{\Lambda}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1}(\boldsymbol{\gamma} - \boldsymbol{\mu})\right). \quad (21)$$

2 Following the properties of the multivariate Gaussian distribution, the marginal distributions are also Gaussian:

$$\boldsymbol{\theta}|\mathbf{y}_{\text{obs}} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathbf{V}_{\boldsymbol{\theta}}), \quad \text{and} \quad \boldsymbol{\psi}|\mathbf{y}_{\text{obs}} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\psi}}, \mathbf{V}_{\boldsymbol{\psi}}), \quad (22)$$

3 which directly gives:

$$p_{\text{Bayes}}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathbf{V}_{\boldsymbol{\theta}}). \quad (23)$$

4 Another property is that the conditional distributions are also Gaussian. We can thus obtain the distribution of param-
5 eters $\boldsymbol{\theta}$ conditioned to a value of hyperparameters $\boldsymbol{\psi}$:

$$\begin{aligned} \boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y}_{\text{obs}} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}|\boldsymbol{\psi}}, \mathbf{V}_{\boldsymbol{\theta}|\boldsymbol{\psi}}), \\ \boldsymbol{\mu}_{\boldsymbol{\theta}|\boldsymbol{\psi}} &= \boldsymbol{\mu}_{\boldsymbol{\theta}} + \mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}}^T \mathbf{V}_{\boldsymbol{\psi}}^{-1}(\boldsymbol{\psi} - \boldsymbol{\mu}_{\boldsymbol{\psi}}), \\ \mathbf{V}_{\boldsymbol{\theta}|\boldsymbol{\psi}} &= \mathbf{V}_{\boldsymbol{\theta}} - \mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}}^T \mathbf{V}_{\boldsymbol{\psi}}^{-1} \mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}}. \end{aligned} \quad (24)$$

6 Accordingly, the distribution of hyperparameters $\boldsymbol{\psi}$ conditioned to a value of parameters $\boldsymbol{\theta}$ has for expression

$$\begin{aligned} \boldsymbol{\psi}|\boldsymbol{\theta}, \mathbf{y}_{\text{obs}} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\psi}|\boldsymbol{\theta}}, \mathbf{V}_{\boldsymbol{\psi}|\boldsymbol{\theta}}), \\ \boldsymbol{\mu}_{\boldsymbol{\psi}|\boldsymbol{\theta}} &= \boldsymbol{\mu}_{\boldsymbol{\psi}} + \mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}} \mathbf{V}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}), \\ \mathbf{V}_{\boldsymbol{\psi}|\boldsymbol{\theta}} &= \mathbf{V}_{\boldsymbol{\psi}} - \mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}} \mathbf{V}_{\boldsymbol{\theta}}^{-1} \mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}}^T. \end{aligned} \quad (25)$$

To perform the KOH estimation, the first step is to compute the marginal of hyperparameters and retrieve its maximum. From equation (22) we obtain

$$\hat{\boldsymbol{\psi}}_{\text{KOH}} = \boldsymbol{\mu}_{\boldsymbol{\psi}}.$$

7 The approximation of the parameters' marginal is then obtained by conditioning $\boldsymbol{\theta}$ on the estimated value $\hat{\boldsymbol{\psi}}_{\text{KOH}}$.

8 Thus,

$$p_{\text{KOH}}(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) = p(\boldsymbol{\theta}|\hat{\boldsymbol{\psi}}_{\text{KOH}}, \mathbf{y}_{\text{obs}}) = p(\boldsymbol{\theta}|\boldsymbol{\mu}_{\boldsymbol{\psi}}, \mathbf{y}_{\text{obs}}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}|\boldsymbol{\mu}_{\boldsymbol{\psi}}}, \mathbf{V}_{\boldsymbol{\theta}|\boldsymbol{\mu}_{\boldsymbol{\psi}}}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathbf{V}_{\boldsymbol{\theta}|\boldsymbol{\psi}}). \quad (26)$$

1 Note the *reduction of variance* property: because the matrix $\mathbf{C}_{\theta,\psi}^T \mathbf{V}_{\psi}^{-1} \mathbf{C}_{\theta,\psi}$ is symmetric positive definite, under the
 2 Loewner order we have $\mathbf{V}_{\theta|\psi} \leq \mathbf{V}_{\theta}$. This order can be understood as the following: let $\{\lambda_i\}_{1 \leq i \leq p}$ be the eigenvalues
 3 of $\mathbf{V}_{\theta|\psi}$ in a decreasing order and $\{\lambda'_i\}_{1 \leq i \leq p}$ be the eigenvalues of \mathbf{V}_{θ} in a decreasing order. Then, for all $i \leq p$, we
 4 have $\lambda_i \leq \lambda'_i$. It is seen that p_{KOH} remains Gaussian, with correct mean $\boldsymbol{\mu}_{\theta}$, but a reduced covariance compared to
 5 p_{Bayes} .

To proceed with the FMP estimation, the first step is to estimate the optimised hyperparameters for each parameter value $\boldsymbol{\theta}$. We directly get, using equation (24),

$$\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\psi}} p(\boldsymbol{\psi}|\boldsymbol{\theta}, \mathbf{y}_{\text{obs}}) = \boldsymbol{\mu}_{\boldsymbol{\psi}|\boldsymbol{\theta}}(\boldsymbol{\theta}) = \boldsymbol{\mu}_{\boldsymbol{\psi}} + \mathbf{C}_{\theta,\psi} \mathbf{V}_{\theta}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta}).$$

Thus $\hat{\boldsymbol{\psi}}_{\text{FMP}}$ is an affine function of $\boldsymbol{\theta}$. To evaluate the approximate posterior distribution of parameters we need to express the inverse of the covariance matrix $\boldsymbol{\Lambda}$, using the Matrix Block Inversion Lemma [33],

$$\boldsymbol{\Lambda}^{-1} = \begin{pmatrix} \mathbf{V}_{\theta}^{-1} + \mathbf{V}_{\theta}^{-1} \mathbf{C}_{\theta,\psi}^T \mathbf{V}_{\psi|\theta}^{-1} \mathbf{C}_{\theta,\psi} \mathbf{V}_{\theta}^{-1} & -\mathbf{V}_{\theta}^{-1} \mathbf{C}_{\theta,\psi}^T \mathbf{V}_{\psi|\theta}^{-1} \\ -\mathbf{V}_{\psi|\theta}^{-1} \mathbf{C}_{\theta,\psi} \mathbf{V}_{\theta}^{-1} & \mathbf{V}_{\psi|\theta}^{-1} \end{pmatrix},$$

6 we obtain

$$\begin{aligned} p_{\text{FMP}}(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) &\propto p(\boldsymbol{\theta}, \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})|\mathbf{y}_{\text{obs}}) \\ &\propto \exp \left(-\frac{1}{2} \begin{pmatrix} \boldsymbol{\theta} - \boldsymbol{\mu}_{\theta} \\ \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}) - \boldsymbol{\mu}_{\boldsymbol{\psi}} \end{pmatrix}^T \boldsymbol{\Lambda}^{-1} \begin{pmatrix} \boldsymbol{\theta} - \boldsymbol{\mu}_{\theta} \\ \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}) - \boldsymbol{\mu}_{\boldsymbol{\psi}} \end{pmatrix} \right) \\ &\propto \exp \left(-\frac{1}{2} \begin{pmatrix} \boldsymbol{\theta} - \boldsymbol{\mu}_{\theta} \\ \mathbf{C}_{\theta,\psi} \mathbf{V}_{\theta}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta}) \end{pmatrix}^T \boldsymbol{\Lambda}^{-1} \begin{pmatrix} \boldsymbol{\theta} - \boldsymbol{\mu}_{\theta} \\ \mathbf{C}_{\theta,\psi} \mathbf{V}_{\theta}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta}) \end{pmatrix} \right) \\ &\propto \exp \left(-\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta})^T \begin{pmatrix} \mathbf{1}_{p \times p} \\ \mathbf{C}_{\theta,\psi} \mathbf{V}_{\theta}^{-1} \end{pmatrix} \boldsymbol{\Lambda}^{-1} \begin{pmatrix} \mathbf{1}_{p \times p} \\ \mathbf{C}_{\theta,\psi} \mathbf{V}_{\theta}^{-1} \end{pmatrix} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta}) \right) \\ &\propto \exp \left(-\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta})^T \mathbf{V}_{\theta}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta}) \right). \end{aligned} \tag{27}$$

7 The density p_{FMP} is proportional to the exponential of a quadratic form in $\boldsymbol{\theta}$, so it is Gaussian, and its mean and
 8 covariance matrix correspond to the true marginal distribution. In this case, the FMP approximation is exact.

9 The behaviour of the methods is illustrated in figure 1 for $\dim(\boldsymbol{\theta}) = \dim(\boldsymbol{\psi}) = 1$. In this example, the FMP

1 method matches exactly the Bayesian solution. The KOH solution has a reduced variance due to the conditioning on
 2 a fixed value of the hyperparameters, which might lead to the "false certitude" effect. This higher certitude (lower
 3 uncertainty) is simply an artifact of the KOH estimation method.

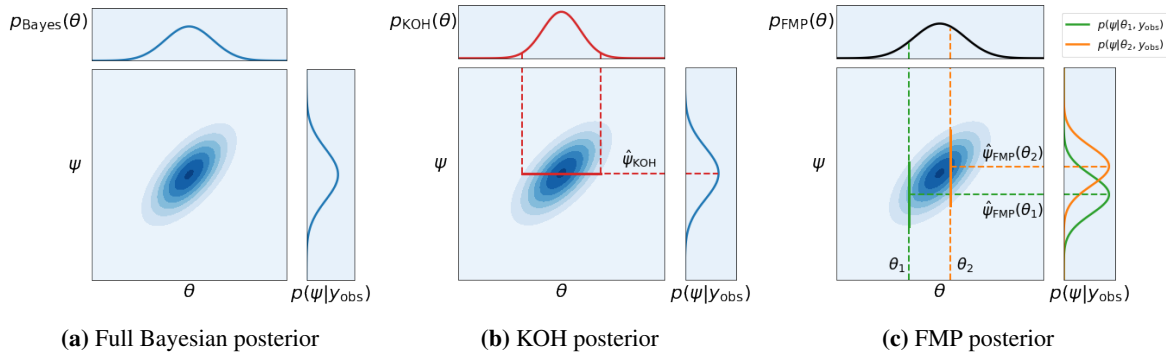


Figure 1: Joint posteriors of parameters and hyperparameters: Gaussian case.

4 3.3.2 Mixture of Gaussians

5 We now consider the case where the joint posterior $p(\boldsymbol{\gamma}|\mathbf{y}_{\text{obs}})$ is a mixture of Gaussians, with modes that are well-
 6 separated, with similar orders of magnitude. The joint density is a mixture of m Gaussians, with weights $(w_i)_{i \leq m}$,
 7 such that $\sum_{i=1}^m w_i = 1$. Their respective means are $\boldsymbol{\mu}_i = \begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{\theta}, i} \\ \boldsymbol{\mu}_{\boldsymbol{\psi}, i} \end{pmatrix}$ and their respective covariance matrices: $\boldsymbol{\Lambda}_i =$
 8 $\begin{pmatrix} \mathbf{V}_{i, \boldsymbol{\theta}} & \mathbf{C}_{i, \boldsymbol{\theta}, \boldsymbol{\psi}}^T \\ \mathbf{C}_{i, \boldsymbol{\theta}, \boldsymbol{\psi}} & \mathbf{V}_{i, \boldsymbol{\psi}} \end{pmatrix}$. The joint density is

$$\begin{aligned}
 p(\boldsymbol{\gamma}|\mathbf{y}_{\text{obs}}) &= \sum_{i=1}^m w_i p(\boldsymbol{\gamma}|\mathbf{y}_{\text{obs}}, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) \\
 &= \frac{1}{(2\pi)^{(p+h)/2}} \sum_i \frac{w_i}{\sqrt{|\boldsymbol{\Lambda}_i|}} \exp\left(-\frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\mu}_i)^T \boldsymbol{\Lambda}_i^{-1} (\boldsymbol{\gamma} - \boldsymbol{\mu}_i)\right).
 \end{aligned} \tag{28}$$

The well-separated hypothesis concerns the projections of the modes on the spaces Θ and Ψ . Geometrically, the well-separated hypothesis on Θ means that, when projecting the Gaussians on the first p coordinates, the 95% confidence ellipses of each mode do not intersect each other. It can be written as:

”The intervals $\Theta_j = \{\boldsymbol{\theta} \text{ s. t. } (\boldsymbol{\theta} - \boldsymbol{\mu}_j)^T \mathbf{V}_{\boldsymbol{\theta}, j}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_j) \leq t_{95}(p)\}$ are disjoint for $1 \leq j \leq m$,”

1 where we have noted $t_{95}(p)$ the 95% quantile of the χ_2 law with p degrees of freedom. The equivalent condition for
 2 the projections over Ψ is also supposed to be true. Another assumption is that the weights $\{w_i\}_{1 \leq i \leq m}$ have the same
 3 order of magnitude, so that, for $1 \leq i \leq m$, if (θ, ψ) is close to μ_i , then $p(\theta, \psi | \mathbf{y}_{\text{obs}}) \approx w_i p(\theta, \psi | \mathbf{y}_{\text{obs}}, \mu_i, \Lambda_i)$. By
 4 linearity, the true marginal density of the parameters is the linear combination of marginals:

$$\begin{aligned} p_{\text{Bayes}}(\theta | \mathbf{y}_{\text{obs}}) &= \int_{\Psi} p(\gamma | \mathbf{y}_{\text{obs}}) d\psi = \sum_{i=1}^m w_i \int_{\Psi} p(\gamma | \mathbf{y}_{\text{obs}}, \mu_i, \Lambda_i) d\psi \\ &= \frac{1}{(2\pi)^{p/2}} \sum_{i=1}^m \frac{w_i}{\sqrt{|\mathbf{V}_{i,\theta}|}} \exp\left(-\frac{1}{2}(\theta - \mu_{\theta,i})^T \mathbf{V}_{i,\theta}^{-1}(\theta - \mu_{\theta,i})\right). \end{aligned} \quad (29)$$

The KOH estimation first computes the hyperparameter marginal distribution:

$$p(\psi | \mathbf{y}_{\text{obs}}) = \frac{1}{(2\pi)^{h/2}} \sum_{i=1}^m \frac{w_i}{\sqrt{|\mathbf{V}_{i,\psi}|}} \exp\left(-\frac{1}{2}(\psi - \mu_{\psi,i})^T \mathbf{V}_{i,\psi}^{-1}(\psi - \mu_{\psi,i})\right).$$

5 According to the separation hypothesis over Ψ , the maximum of $p(\psi | \mathbf{y}_{\text{obs}})$ is:

$$\begin{aligned} \hat{\psi}_{\text{KOH}} &= \arg \max_{\psi} p(\psi | \mathbf{y}_{\text{obs}}) = \mu_{i_{\text{KOH}}, \psi}, \\ i_{\text{KOH}} &= \arg \max_{i \leq m} \frac{w_j}{\sqrt{|\mathbf{V}_{i,\psi}|}}. \end{aligned} \quad (30)$$

6 So, the KOH posterior is:

$$p_{\text{KOH}}(\theta | \mathbf{y}_{\text{obs}}) = p(\theta | \mathbf{y}_{\text{obs}}, \psi = \mu_{i_{\text{KOH}}, \psi}) = \mathcal{N}(\mu_{\theta, i_{\text{KOH}}}, \mathbf{V}_{i_{\text{KOH}}, \theta | \psi}), \quad (31)$$

with

$$\mathbf{V}_{i_{\text{KOH}}, \theta | \psi} = \mathbf{V}_{i_{\text{KOH}}, \theta} - \mathbf{C}_{i_{\text{KOH}}, \theta, \psi}^T \mathbf{V}_{i_{\text{KOH}}, \psi} \mathbf{C}_{i_{\text{KOH}}, \theta, \psi}.$$

7 The solution of the KOH estimation is Gaussian with reduced variance matrix. Besides, the selection of the mode
 8 is driven by the criteria defining i_{KOH} , which does not necessarily correspond to the true maximum of the Bayesian
 9 solution.

Optimal hyperparameters in the FMP method are given by the solution to the optimization problem:

$$\hat{\theta}_{\text{FMP}} = \arg \max_{\theta} p(\gamma | \mathbf{y}_{\text{obs}}) = \arg \max_{\theta} \sum_{i=1}^m \frac{w_i}{\sqrt{|\Lambda_i|}} \exp\left(-\frac{1}{2}(\gamma - \mu_i)^T \Lambda_i^{-1}(\gamma - \mu_i)\right)$$

According to the separation hypothesis, the confidence intervals $\{\Theta_i\}_{i \leq m}$ are disjoint so that

$$\text{for } \boldsymbol{\theta} \in \Theta_i, \quad p(\boldsymbol{\gamma}|\mathbf{y}_{\text{obs}}) \approx w_i p(\boldsymbol{\gamma}|\mathbf{y}_{\text{obs}}, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i),$$

$$\text{and } \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}) = \hat{\boldsymbol{\psi}}_{i,\text{FMP}}(\boldsymbol{\theta}) := \boldsymbol{\mu}_{i,\boldsymbol{\psi}} + \mathbf{C}_{i,\boldsymbol{\theta},\boldsymbol{\psi}} \mathbf{V}_{i,\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}}).$$

1 We note $p_{\text{FMP}}^*(\boldsymbol{\theta}) = p(\boldsymbol{\theta}, \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})|\mathbf{y}_{\text{obs}}) = \sum_{j=1}^m w_j p(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})|\mathbf{y}_{\text{obs}}, \boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j)$ the unnormalized FMP
2 approximation of the parameter posterior. The following two properties are true:

- 3 • $\forall \boldsymbol{\theta} \in \Theta_i, \quad w_i p(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{i,\text{FMP}}(\boldsymbol{\theta})|\mathbf{y}_{\text{obs}}, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) \gg \sum_{j \neq i} w_j p(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{i,\text{FMP}}(\boldsymbol{\theta})|\mathbf{y}_{\text{obs}}, \boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j)$. This is the application
4 of the separation property in Ψ and the low discrepancy in the weights $\{w_i\}_{1 \leq i \leq m}$.
- 5 • $w_i p(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{i,\text{FMP}}(\boldsymbol{\theta})|\mathbf{y}_{\text{obs}}, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) = \frac{w_i}{(2\pi)^{(p+h)/2} \sqrt{|\boldsymbol{\Lambda}_i|}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}})^T \mathbf{V}_{i,\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}})\right)$. This result was ob-
6 tained in the unimodal case (see equation (27)).

Thus, a natural approximation for p_{FMP}^* , for $\boldsymbol{\theta} \in \Theta$, is

$$p_{\text{FMP}}^*(\boldsymbol{\theta}) \approx \sum_{j=1}^m w_j p(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{j,\text{FMP}}(\boldsymbol{\theta})|\mathbf{y}_{\text{obs}}, \boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j),$$

and if $\boldsymbol{\theta} \in \Theta_i$ it simplifies to

$$p_{\text{FMP}}^*(\boldsymbol{\theta}) \approx \frac{w_i}{(2\pi)^{(p+h)/2} \sqrt{|\boldsymbol{\Lambda}_i|}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}})^T \mathbf{V}_{i,\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}})\right).$$

7 Since most of the probability mass of $p_{\text{FMP}}^*(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}})$ is contained within the intervals Θ_i , we have

$$p_{\text{FMP}}(\boldsymbol{\theta}) = \frac{1}{K} \sum_{i=1}^m \frac{w_i}{\sqrt{|\boldsymbol{\Lambda}_i|}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}})^T \mathbf{V}_{i,\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}})\right) \quad \text{for } \boldsymbol{\theta} \in \Theta, \quad (32)$$

with the normalizing constant

$$\begin{aligned} K &= \int_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^m \frac{w_i}{\sqrt{|\boldsymbol{\Lambda}_i|}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}})^T \mathbf{V}_{i,\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}})\right) d\boldsymbol{\theta} \\ &\approx (2\pi)^{(p/2)} \sum_{i=1}^m w_i \frac{\sqrt{|\mathbf{V}_{i,\boldsymbol{\theta}}|}}{\sqrt{|\boldsymbol{\Lambda}_i|}}. \end{aligned}$$

8 The distribution p_{FMP} is a linear combination of the m marginals associated with the original Gaussian mixture. The
9 variances of the different peaks are correctly estimated, but the estimation introduces a bias on the weights of the

1 Gaussian mixture: the FMP estimates the original weights w_i to be $w'_i = \frac{1}{k} w_i \frac{\sqrt{|\mathbf{V}_{i, \mathbf{e}}|}}{\sqrt{|\mathbf{\Lambda}_i|}}$ where k is a normalizing
 2 constant such that $\sum_{i=1}^m w'_i = 1$. Specifically, the FMP method emphasizes the importance of the Gaussians with low
 3 hyperparameter variance in the mixture.

4 An illustration of the methods is given in figure 2. In this example, the KOH estimation finds only one peak that
 5 does not correspond to the true maximum of the posterior, so the conclusions about θ are potentially misleading.
 6 In the FMP estimation, all the peaks are found, yet their relative weights might be wrongly estimated, leading to an
 7 inversion of the importance of the peaks so that the maximum a posteriori estimator is incorrect. **This feature will be**
 8 **found in any method that uses estimated hyperparameters instead of marginalization (see assumption 2 in equation 16)**
 9 **because volume effects cannot be seen.** We argue that it is still favorable to perform the FMP estimation in this case
 10 so that no possible explanation of the observations is missed.

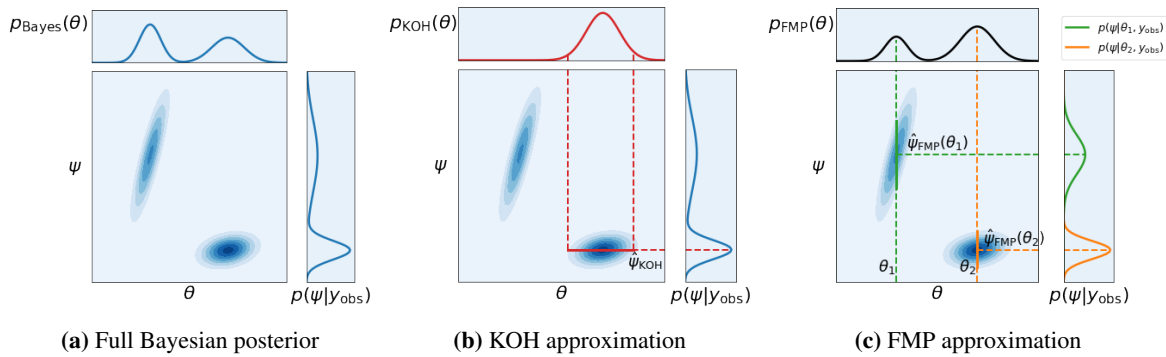
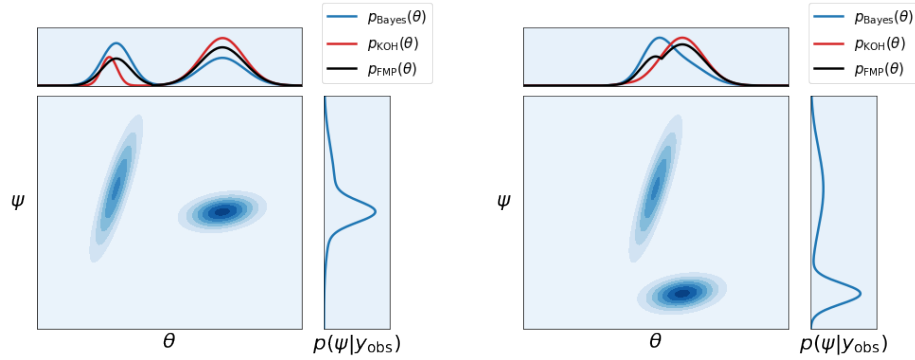


Figure 2: Joint posterior of parameters and hyperparameters: well separated Gaussian mixture case.

11 We now discuss the case where the two Gaussians of the mixture are not well-separated. In figure 3a, the Ψ -
 12 projections of the two modes overlaps. The corresponding estimations of the KOH and FMP methods are plotted in
 13 the figure. In this example, the KOH criterion selects the right mode; contrary to the well-separated case, the other
 14 (left) mode contributes to the parameter posterior because part of its probability mass is captured when conditioning
 15 on $\hat{\psi}_{\text{KOH}}$. However, KOH badly estimates the second mode, contrary to the FMP approximation, which retrieves
 16 the two modes correctly. This situation occurs when a single model discrepancy adequately applies to the whole
 17 parameters domain.

18 Figure 3b shows the case where the Θ -projections of the Gaussians overlap. This situation is challenging for
 19 the two approximation methods since they rely on a point-mass approximation of the marginals of ψ , which is not
 20 verified here. In the example shown, the FMP approximation performs slightly better with a posterior variance of θ
 21 underestimated but closer to the reference than for KOH. This situation occurs when the distribution of the model
 22 discrepancy is at odds with the observations, and no hyperparameters value stands out; it calls for a change in the

1 prior of z_θ (e.g., selecting another covariance structure).



(a) Posterior with intersecting Ψ -projection.

(b) Posterior with intersecting Θ -projection.

Figure 3: Comparison of KOH and FMP estimations: case of Gaussian mixtures without modes separation.

2 4. APPLICATIONS

3 In this section, we present two examples. The first one in Section 4.1 is an analytical model in which predictions
 4 have different shapes depending on the parameter value, yielding two explanations of the observations. The posterior
 5 is then bimodal, challenging the KOH estimation. The elementary nature of this example also allows studying the
 6 normality of the posterior by fitting a Gaussian Mixture Model.

7 The second application in Section 4.2 deals with the calibration using actual experimental measurements of a
 8 boiling model. Physical insights are required to formulate the statistical assumptions. The model error and the mea-
 9 surement uncertainty are both significant in this problem, and the FMP calibration correctly attributes the uncertainty
 10 of each source, whereas the KOH calibration sees only one.

11 4.1 Calibration of an inadequate model

12 We apply the FMP calibration to an elementary problem where model inadequacy is present, and the structure of
 13 the model predictions is sensitive to parameter variations. We also perform the KOH and full Bayesian calibrations
 14 on this example. The Bayesian calibration shows that the joint posterior can be approximated by a mixture of two
 15 Gaussian distributions, illustrating the previous theoretical results.

1 4.1.1 Problem formulation

2 In this example, the true function is $y(x) = x$ and the computer model is given by

$$f(x, \theta) = x \sin(2\theta x) + (x + 0.15)(1 - \theta). \quad (33)$$

3 The input variable x and the model parameter θ lie within the restricted range $(x, \theta) \in [0, 1] \times [-0.5, 1.5]$. We use
 4 observations of the true function at 8 points uniformly spaced in the interval $[0, 1]$. The observations are noisy with a
 5 centered Gaussian noise with standard deviation $\sigma_\epsilon = 0.1$. The observation noise is not known a priori but is learned
 6 along with the other hyperparameters of the model error.

7 The prior of the model discrepancy is a Gaussian Process with zero mean and a Squared Exponential covariance

$$c_\psi(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right). \quad (34)$$

8 Thus, a total of three hyperparameters $\psi = (\sigma, l, \sigma_\epsilon)$ are involved in the calibration. The prior for θ is uniform over
 9 $[-0.5, 1.5]$. The priors for σ and σ_ϵ are uniform on the interval $[1e-5, 1]$. The prior on l is also uniform, truncated
 10 on $[1e-3, 5]$.

11 Samples of the posterior distributions are obtained using Monte-Carlo Markov Chain methods with the Metropolis-
 12 Hastings algorithm [48]. For the KOH calibration, we first estimate the hyperparameters using (11). The MCMC
 13 method is then run for the target density $p_{\text{KOH}}(\theta)$ (see (12)). In the FMP calibration, the target density $p_{\text{FMP}}(\theta)$ is
 14 also of dimension 1, but each step of the MCMC requires the determination of the optimal hyperparameters. The cost
 15 of these optimizations is not prohibitive, thanks to the available gradients [33, Chapter 5], and because the optimal
 16 hyperparameters of the previous step are generally a good starting point. For the Bayes calibration, the target density
 17 $p(\theta, \psi | \mathbf{y}_{\text{obs}})$ is in dimension four and requires significantly more steps. The FMP and KOH chains have $5e5$ steps,
 18 and the Bayes chain has $2.5e6$ steps, all with a burn-in of 10%. From each chain, we extract a posterior sample of size
 19 5000 by taking regularly-spaced visited states. This subsampling rate is large enough compared to the self-correlation
 20 length of the chains to consider the resulting samples independent. For more details about this procedure, see [49].

21 4.1.2 Calibration Results

22 We apply a Kernel Density Estimation method to the MCMC samples to estimate the posterior marginal densities
 23 of parameters and hyperparameters. These marginals are shown in figure 4. The parameter's posterior marginals are

1 bimodal with a first mode located at $\theta = -0.025$ and the second at $\theta = 1.02$. For the reference Bayes solutions, the
 2 relative importance of the two modes is in the ratio 4.3:1, computed as the relative frequency of $\theta < 0.5$. The ratio
 3 is 2.75:1 for the FMP method and 69.4:1 for the KOH method. The FMP method overestimates the second mode's
 4 importance and puts more probability mass between the two modes. The KOH estimation almost misses the second
 5 mode focusing on a single interpretation of the observations.

6 Hyperparameters' posterior marginals reveal that the FMP method improves the calibration results. All three
 7 hyperparameters feature a variance a posteriori that the KOH estimation cannot capture. The KOH hyperparameters,
 8 by definition, are at the Bayes posterior maxima; they match the maxima of the marginals of σ_ϵ and σ , although with
 9 a slight offset for l . The FMP marginals have modes located at these maxima and have smaller variances than the
 10 Bayes solution due to the reduction in their optimal values. Note that the posterior marginal of l accumulates mass
 11 around the maximal value a priori. We have repeated the calibration with high maximal values up to $l = 20$ without
 12 finding a notable difference, as the Bayes posterior remains flat. Note that for the inferred σ_ϵ , the Bayes solution is
 13 the only one to correctly estimate the true value of 0.1 when other methods infer smaller values $\sigma_\epsilon \approx 0.06$.

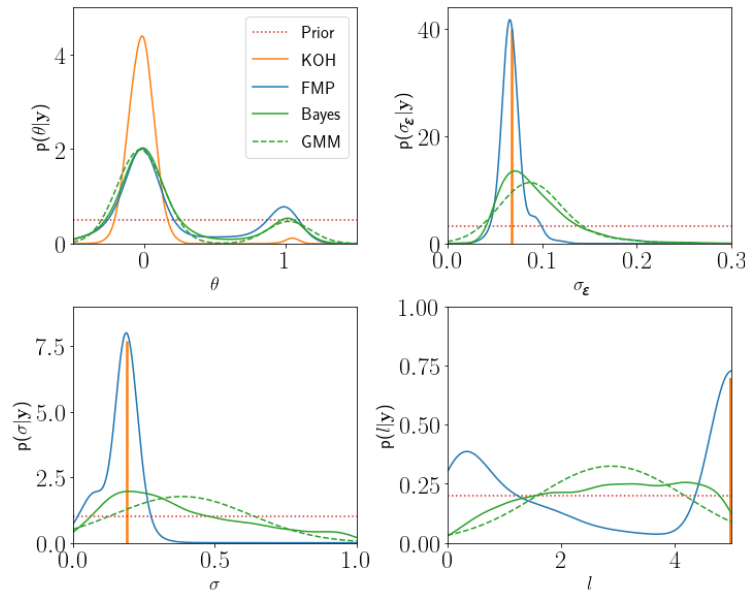


Figure 4: Prior and posterior distributions for the model parameter (top left) and hyperparameters. In the KOH calibration, hyperparameters are estimated with point mass distributions, represented with vertical lines. The Gaussian Mixture Model, fitted on the Bayes posterior, is represented with the dashed curves. In the bottom row, some probability mass lies outside the support of the densities due to the Kernel Density Estimation. All densities are normalized over their support for proper comparison.

14 The bimodality of the parameter posterior reflects in the posterior model predictions shown in figure 5. The
 15 left mode, the only one found by KOH, corresponds to quasi-linear predictions with a constant offset from the true

1 process $y(x)$. The second modes correspond to predictions close to the true process but with an oscillation in x .
 2 FMP method accounts for the two types of predictions present in the Bayes solution. We also see a clear separation
 3 between the two types of predictions in the Bayes solutions, while the FMP method produces some predictions with
 4 intermediate structures.

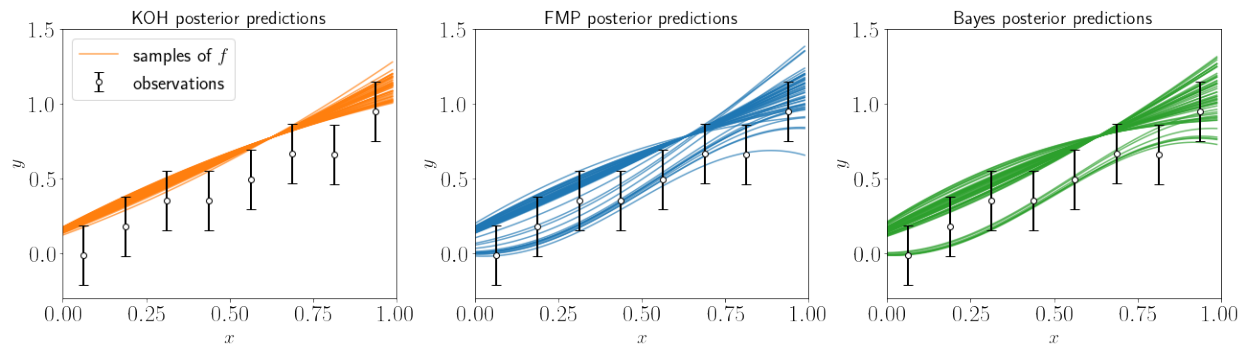


Figure 5: Posterior samples of the model predictions for all three calibration methods. 50 samples are represented on each figure, and observations are shown with $2\sigma_\epsilon$ confidence intervals. Model predictions that are quasi-linear correspond to $\theta \approx 0$, and the ones with slow oscillations correspond to $\theta \approx 1$.

5 4.1.3 Gaussian Mixture Model fit on the Bayes solution

6 As a further step, we fit a mixture of Gaussians model on the MCMC samples of the Bayes solution. To do so, we
 7 employed a hard clustering algorithm based on Expectation-Maximization *. The selected form of the mixture is a
 8 weighted sum of two Gaussians

$$p(\theta, \psi | \mathbf{y}_{\text{obs}}) = w_1 \mathcal{N}(\mu_1, \mathbf{C}_1) + w_2 \mathcal{N}(\mu_2, \mathbf{C}_2). \quad (35)$$

9 The estimated coefficients of the GMM are presented in Table 1.

10 The two estimated modes reveal that, even though they are distinct in θ , they are not in ψ . We are thus in a
 11 situation without clear separation featuring the intersection of projections in ψ space. The marginals of the GMM are
 12 shown in Figure 4. A good fit of the Bayesian and GMM marginals is reported for θ . The fit is much less satisfactory
 13 for the hyperparameters: the normality assumption is inappropriate due to the truncated supports. Nevertheless, the
 14 GMM illustrates some results of section 3.3.2 as discussed below:

- 15 • The KOH estimation selects a mode according to the criterion in equation (30). We check that $\frac{w_1}{\sqrt{|\mathbf{C}_1^{[3,3]}|}} = 70.8$
 16 is greater than $\frac{w_2}{\sqrt{|\mathbf{C}_2^{[3,3]}|}} = 10.0$, which is the reason why it selects the first mode.

*<https://perso.telecom-paristech.fr/bonald/documents/gmm.pdf>

| Coefficient | Estimation |
|----------------------|--|
| w_1 | 0.81 |
| w_2 | 0.19 |
| $\boldsymbol{\mu}_1$ | $(-0.03, 0.09, 0.40, 2.93)^T$ |
| $\boldsymbol{\mu}_2$ | $(1.02, 0.12, 0.29, 2.56)^T$ |
| \mathbf{C}_1 | $\begin{pmatrix} 2.9\text{e-}2 & -1.9\text{e-}4 & 1.3\text{e-}3 & -1.6\text{e-}3 \\ -1.9\text{e-}4 & 1.5\text{e-}3 & -8.6\text{e-}5 & -1.4\text{e-}3 \\ 1.3\text{e-}3 & -8.6\text{e-}5 & 5.7\text{e-}2 & 3.0\text{e-}2 \\ -1.6\text{e-}3 & -1.4\text{e-}3 & 3.0\text{e-}2 & 1.6 \end{pmatrix}$ |
| \mathbf{C}_2 | $\begin{pmatrix} 2.1\text{e-}2 & -3.0\text{e-}4 & -8.0\text{e-}4 & 5.1\text{e-}2 \\ -3.0\text{e-}4 & 2.9\text{e-}3 & 5.7\text{e-}4 & 7.8\text{e-}3 \\ -8.0\text{e-}3 & 5.7\text{e-}4 & 6.2\text{e-}2 & 4.9\text{e-}2 \\ 5.1\text{e-}2 & 7.8\text{e-}3 & 4.9\text{e-}2 & 2.1 \end{pmatrix}$ |

Table 1: Estimated coefficients of the Gaussian Mixture Model for the posterior distribution of $\boldsymbol{\mu} = (\theta, \sigma_{\text{mes}}, \Delta T, \sigma, l)$.

- 1 • The posterior correlation between θ and $\boldsymbol{\psi}$ is low, so the variance at the first mode is roughly unchanged
2 after projection: $\mathbf{V}_\theta = \mathbf{C}_1(0, 0) = 2.90\text{e-}2$, and $\mathbf{V}_{\theta|\boldsymbol{\psi}} = \mathbf{C}_1(0, 0) - \mathbf{C}_{1,\theta,\boldsymbol{\psi}}^T (\mathbf{C}_1^{[3,3]})^{-1} \mathbf{C}_{1,\theta,\boldsymbol{\psi}} = 2.89\text{e-}2$,
3 with $\mathbf{C}_{1,\theta,\boldsymbol{\psi}} = (-1.9\text{e-}4, 1.3\text{e}3, -1.6\text{e-}3)^T$. The KOH method underestimates the posterior variance due to
4 selecting only one mode.
- 5 • The FMP weights can be computed using (32). We have $w'_i \propto w_i \frac{\sqrt{|\mathbf{C}_i|}}{\sqrt{\mathbf{C}_i(0,0)}}$, which gives $w'_1 = 0.73$ and
6 $w'_2 = 0.27$, in a proportion 2.71:1 which corresponds to the ratio found in the FMP results.

7 Thus, even if the assumption of normal joint posterior is questionable, this example shows the key role of the
8 posterior covariance in understanding the behaviour of the calibration methods. However, this covariance can not be
9 accessed before the calibration.

To further contrast the quality of the FMP and KOH distributions, we measure their respective similarities with
the Bayes distribution using the Jensen-Shannon divergence (D_{JS}). The JS-Divergence between two distributions
with density p_* and the Bayes' density p_{Bayes} is given by

$$D_{\text{JS}}(p_* || p_{\text{Bayes}}) = \frac{1}{2} D_{\text{KL}}(p_*, p_{1/2}) + \frac{1}{2} D_{\text{KL}}(p_{\text{Bayes}}, p_{1/2}),$$

10 where $p_{1/2} = (p_* + p_{\text{Bayes}})/2$ and $D_{\text{KL}}(p_1 || p_2)$ is the standard Kullback-Leibler divergence. The JS-Divergence
11 is symmetric and is zero when the two distributions compared are equal. A larger value of D_{JS} means a higher
12 dissimilarity. The present work estimates the divergences using the MCMC sample sets. In the first stage, we use a
13 KDE method to estimate the densities. In the second stage, the densities are evaluated at the sample set's points to
14 estimate the averages in the expression of the KL divergences. Compared to the KL divergence, the JS divergence is

1 much less susceptible to sampling noise and more stable when the densities are dissimilar. In the present analysis,
 2 resampling with bootstrapping has shown that the estimates are sufficiently converged to characterize the differences
 3 between the two methods.

4 The divergence for the posterior distributions of the parameter θ and the model's prediction $f(\cdot, \theta)$ at some x are
 5 reported in Table 2. These results show that the FMP distribution has a much lower JS-Divergence with the Bayesian
 6 distribution than the KOH distribution. Specifically, in this example, the divergence from the Bayesian distribution of
 7 the FMP method is less than 7.5% than that of the FMP method.

| Quantity | JS Divergences | | Relative (%) |
|-----------------------|----------------|-----------|--------------|
| | Bayes-FMP | Bayes-KOH | |
| θ | 0.0072 | 0.1416 | 5.1 |
| $f(x = 0.56, \theta)$ | 0.0052 | 0.1222 | 4.2 |
| $f(x = 0.69, \theta)$ | 0.0074 | 0.1334 | 5.5 |
| $f(x = 0.94, \theta)$ | 0.0097 | 0.1295 | 7.5 |

Table 2: JS-Divergences between the Bayes and FMP and the Bayes and KOH distributions. The divergences are computed for the posterior distributions of the parameter (θ) and model's predictions ($f(\cdot, \theta)$) at different values of x as indicated. The last column provides the ratio of the Bayes-FMP and Bayes-KOH divergences.

8 4.2 Calibration of a boiling model

9 We now tackle the calibration of a thermal model using experimental observations. Three parameters are selected for
 10 the calibration for their strongly non-linear impact on the model predictions. The model discrepancy distribution is
 11 formulated using physical knowledge.

12 4.2.1 Problem formulation

13 The MIT Boiling model [50] simulates the boiling of fluids in contact with a heated wall. The inputs of the model
 14 are experimental variables (fluid pressure, velocity, temperature, and configuration geometry) and the wall superheat
 15 ΔT_{sup} , defined as the difference between the wall temperature and the saturation temperature of the fluid. The model
 16 output is the heat flux ϕ from the wall to the liquid. The model is semi-empirical and based on experimental correla-
 17 tions, with low evaluation cost. The experimental observations used in this study come from the boiling experiment of
 18 Kennel [51], precisely case number 6 (one of the observation sets used by Kommajosyula to validate the MIT Boiling
 19 model). It consists of 8 joint measurements of heat flux and wall temperature in different boiling regimes, ranging
 20 from no-boiling to fully developed nucleate boiling.

21 Three calibration parameters ($\theta_1, \theta_2, \theta_3$) are considered; the motivations for this choice and the significance of

1 each parameter is discussed in [52]. The measurement uncertainty lies primarily on ΔT_{sup} , which is an input model
 2 quantity, not an output. This specificity requires an adaptation of the framework. Following [53], the sensitivity
 3 (derivatives) of \mathbf{y}_{obs} at the uncertain inputs is estimated from a polynomial fit over the observations and used to
 4 include this uncertainty in the framework (see below).

5 We base the formulation of the model discrepancy on two considerations. First, because ϕ is a smooth increasing
 6 function of ΔT_{sup} , the model discrepancy should be smooth too. The second consideration concerns the magnitude of
 7 the discrepancy. At low values of ΔT_{sup} , there is no boiling, and the model predictions are accurate (linear tendency).
 8 At high values of ΔT_{sup} , the emergence of nucleate boiling induces an exponential increase of the heat flux that might
 9 be incorrectly represented by the model, requiring a higher variability of the discrepancy term. Consequently, the
 10 selected model discrepancy is a Gaussian Process with mean zero and covariance function given by:

$$c_{\psi}(\Delta T_{\text{sup}}, \Delta T'_{\text{sup}}) = \sigma^2 \Delta T_{\text{sup}} \Delta T'_{\text{sup}} \exp\left(-\frac{|\Delta T_{\text{sup}} - \Delta T'_{\text{sup}}|^2}{2l^2}\right). \quad (36)$$

11 The vector of observations has a normal distribution, with covariance matrix $\Sigma_{\psi} + \sigma_{\text{mes}, \Delta T}^2 \mathbf{D}$, where \mathbf{D} is the diagonal
 12 matrix with diagonal coefficients equal to the derivatives of \mathbf{y}_{obs} at the observations.

13 Table 3 reports the prior distributions considered for the calibration. The priors are truncated normals and uniform
 14 distributions for the parameters and hyperparameters, respectively. The priors' ranges are selected to ensure physically
 15 meaningful value (parameters) and using a priori considerations on the magnitude of the model and measurement
 16 errors (hyperparameters).

17 As in the first example, we sample the posteriors with the Metropolis-Hastings algorithm. The KOH and FMP
 18 chains have an acceptance rate of $\approx 11\%$ and the Bayes chain $\approx 3\%$. Figure 6 shows the self-correlation of the
 19 MCMC chains. The mixing lengths of the chains are $\tau = 28$ for KOH, $\tau = 30$ for FMP, and $\tau = 470$ for Bayes.
 20 To ensure sample sets of similar quality, we generate chains with respective lengths $5e5$, $5e5$, and $1e7$ for the KOH,
 21 FMP, and Bayes methods and extract 1,000 samples from the chains by uniform subsampling.

22 A polynomial fit of the model is used to estimate derivatives. The FMP optimizations use the C++ library NLOpt[†],
 23 with the global algorithm MLSL, using the local optimizer SBPLX. A maximal time of $10e-4$ s is allocated to
 24 individual optimization. For the KOH estimation of hyperparameters, the integral over Θ is computed by quadrature,
 25 using a QMC grid of size 300 over the parameters domain.

[†]<https://nlopt.readthedocs.io/en/latest/>

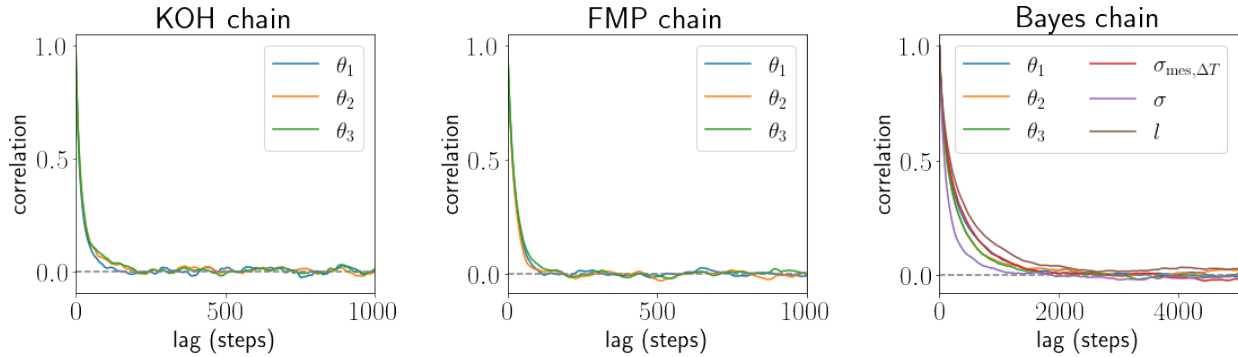


Figure 6: Self-correlation of the chains for the three methods applied to the MIT Boiling model calibration.

1 4.2.2 Results

2 Table 3 summarizes the results of the inference problem for the three methods. The most noticeable result is that
 3 the KOH method estimates a zero model error σ and attributes all the discrepancies to measurement uncertainty.
 4 Consistently, the KOH method yields an erroneous estimation of θ_1 with an underestimated variance. On the contrary,
 5 the FMP method identifies a model error comparable to the full Bayesian calibration and, despite underestimating the
 6 hyperparameters' variances, provides a globally correct inference of the parameters and hyperparameters.

| Calibration | θ_1 | θ_2 | θ_3 | σ | l | $\sigma_{\text{mes}, \Delta T}$ |
|--|------------------------------------|------------------------------------|------------------------------------|-------------------------|------------------------|---------------------------------|
| Prior distributions | | | | | | |
| - | $\mathcal{N}_{t, I_1}(0.5, 0.3^2)$ | $\mathcal{N}_{t, I_2}(0.5, 0.3^2)$ | $\mathcal{N}_{t, I_2}(0.5, 0.3^2)$ | $\mathcal{U}([0, 1e6])$ | $\mathcal{U}([1, 30])$ | $\mathcal{U}([0.1, 0.8])$ |
| Posterior summaries (mean \pm std) | | | | | | |
| KOH | 0.45 ± 0.22 | 0.50 ± 0.28 | 0.37 ± 0.12 | 0 ± 0 | 15.0 ± 0 | 0.48 ± 0 |
| FMP | 0.28 ± 0.29 | 0.53 ± 0.28 | 0.41 ± 0.14 | $5.7e3 \pm 8.6e3$ | 13.2 ± 5.5 | 0.32 ± 0.14 |
| Bayes | 0.27 ± 0.29 | 0.48 ± 0.28 | 0.43 ± 0.18 | $5.6e4 \pm 1.0e5$ | 17.2 ± 7.8 | 0.37 ± 0.17 |

Table 3: Prior distributions and posterior summaries obtained for each calibration technique. Notations are \mathcal{U} for uniform distributions, and $\mathcal{N}_{t, I_k}(\mu, \sigma^2)$ refers to a truncated normal distribution over the interval I_k , with $I_1 = [-0.5, 1.75]$ and $I_2 = [0, 1.5]$.

7 Figure 7 shows the posterior of the corrected model prediction ($f + z_\theta$) of the true process y . The credible
 8 intervals of the corrected prediction in FMP and Bayes are generally larger than for KOH, especially at low values of
 9 ΔT_{sup} where no observations are available. Note that the uncertainty is zero at ΔT_{sup} because of the structure of the
 10 model discrepancy covariance. The predictive variance is split into the model and residual contributions, as described
 11 in Section 3.1.2. For KOH, the variance a posteriori of z_θ is uniformly zero, as σ is inferred to be zero. This result is
 12 unsatisfactory, especially at low values of ΔT_{sup} where few observations are available. All approaches agree on the
 13 dominance of the model predictions variance for the high range of ΔT_{sup} values.

14 More insight can be gained by plotting the projection of the posterior samples in the $\sigma_{\text{mes}, \Delta T}$ and l plane; see

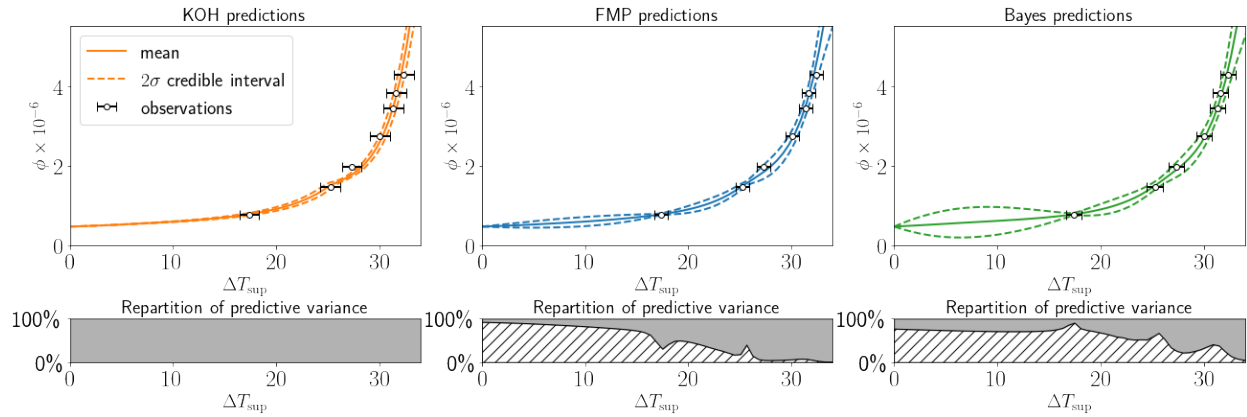


Figure 7: Posterior predictions of the corrected model ($f + z_\theta$) with uncertainty. Observations are shown with inferred error bars of length $2\mathbb{E}_\theta[\sigma_{\text{mes},\Delta T}]$. The bottom plots show the decomposition of the predictive variance in residual uncertainty (stripped region) and uncertainty in the corrected model (grey filled region) as explained in Section 3.1.2.

1 the left plot of Fig. 8. The Bayes samples cover the full support of the prior; the FMP samples fall primarily in two
 2 regions corresponding to two distinct interpretations of the observations: high measurement error with zero model
 3 error or a combination of both. The KOH estimator falls into the former. Since KOH estimates $\sigma = 0$, all lengths l are
 4 equally likely, and the reported value $l = 15.0$ comes from the initialization of the optimization problem. The second
 5 and third plots of Fig. 8 report the FMP model predictions corresponding to these two regions. In the interpretation
 6 with measurement error only, plausible model predictions pass through all observations and have tight dispersions.
 7 In the interpretation with non-zero model error, some model predictions come close to most observations, but others
 8 are further away, following the observations' trend. From a practical perspective, acknowledging the possibility of
 9 low measurement and non-zero model errors is crucial to properly assess the uncertainty in the model predictions
 10 of non-observed quantities that can not be corrected. The FMP and Bayes methods achieve this goal of considering
 11 alternative interpretations of the observations.

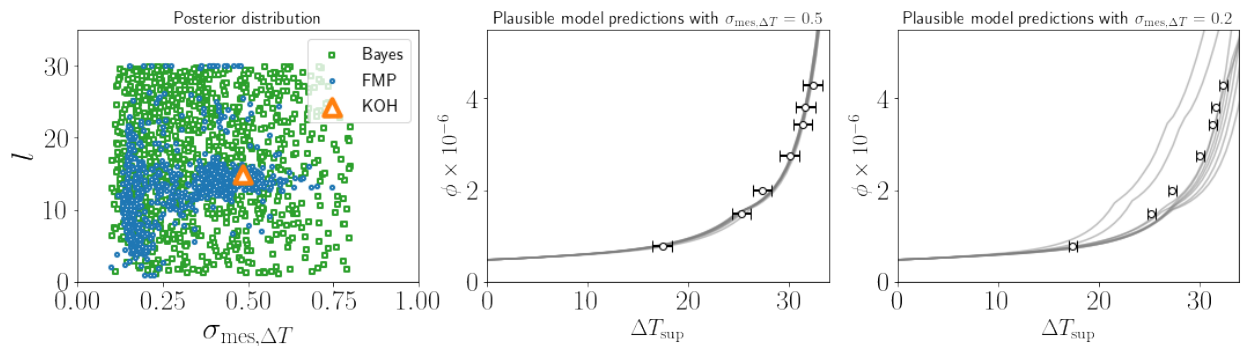


Figure 8: Samples in the σ_ϵ and l plane for the three methods (left plot). Samples of the FMP model predictions conditioned on a zero (middle) and non-zero model error (right).

1 5. CONCLUSIONS AND DISCUSSION

2 We have proposed an approach to estimate both model parameters and model discrepancy. For the first time in
3 literature, to our knowledge, the model error is explicitly dependent on model parameter values in the calibration
4 framework. We showed that it solves the identifiability problem between model error and parameter uncertainty. The
5 resulting predictive uncertainty splits into two natural contributions: the error in the calibrated model and the residual
6 uncertainty. The FMP approximation has a reduced cost compared to the full Bayesian inference while correcting the
7 defects of the KOH estimations when the posterior of the hyper-parameters has several modes.

8 In two applications, we found that the FMP calibration performs better than the KOH calibration in terms of un-
9 certainty estimation, values of parameters, and posterior predictions. Allowing some variance in the hyperparameters
10 of the model discrepancy helps avoid pitfalls such as the false certainty effect or missing entire probability regions.
11 The FMP approach proves to be an accurate approximation of the full Bayesian calibration and significantly reduces
12 the dimension of the MCMC sample space. Thus, the FMP method is well suited for situations requiring complex
13 model error terms with many hyperparameters or when performing calibration on a collection of experiments with
14 independent model discrepancy terms.

15 The reduction of the sample space in the FMP method is achieved by using an optimized value for the hyper-
16 parameters, rather than sampling its conditional distribution. The FMP method then solves an optimization problem
17 for the pointwise estimation of the hyperparameters, at each step of the MCMC, which induces a significant com-
18 putational overhead compared to using a global pointwise estimates. Specifically, we solve the optimization problem
19 in equation (15), where each evaluation of the likelihood requires inverting the covariance matrix of the experimen-
20 tal data. This is the most costly step of the FMP procedure since this operation is repeated for each iteration of the
21 MCMC algorithm, especially when the number of observations is high or there are many model error hyperparame-
22 ters. Because of the relatively low dimensionality of the two application problems presented in this paper, we have
23 not observed significant computational benefits of using the FMP method compared to the full Bayesian approach
24 in our numerical experiments. We are currently developing adaptive methods that construct surrogate models of the
25 FMP hyperparameters $\hat{\psi}_{\text{FMP}}(\theta)$ in an offline stage. These surrogates bypass the optimization steps with considerable
26 computational savings compared to the full Bayesian method, without significantly altering the quality of posterior
27 approximation. With the hyperparameters surrogates, the cost of sampling the FMP posterior is comparable to the
28 cost of the KOH method. These results will be presented in an article in preparation.

29 Our calibration equation offers new challenges to propose appropriate priors. Consistently with their definition
30 of model error, previous works used model discrepancy hyperparameters a priori independent of model parameters.

1 We also adopt this hypothesis in the present work, but one could imagine more complex prior. We advocate for
2 using model discrepancy distributions that are more complex and more physics-informed to provide more accurate
3 predictions.

4 **SUPPLEMENTARY MATERIAL**

5 The implementation of the MIT Boiling model, as well as the Kennel measurements data set used in section 4.2, are
6 available online at <https://github.com/nleoni95/MITB>.

7 **ACKNOWLEDGMENTS**

8 This work was partly funded by the French Innovation Agency for Defense (Agence de l’Innovation de Défense,
9 AID). N. Leoni would also like to thank Ravikishore Kommajosyula for providing the boiling model, and Kathryn
10 Maupin for helpful discussions.

11 This project was partly funded from the Clean Sky 2 Joint Undertaking (JU) under grant agreement No 101008257.

12 **References**

- 13 1. Oberkampf, W.L. and Roy, C.J., *Verification and Validation in Scientific Computing*, Cambridge University Press, Cambridge,
14 2010.
- 15 2. Jaynes, E.T. and Bretthorst, G.L., *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK;
16 New York, NY, 2003.
- 17 3. Kennedy, M.C. and O’Hagan, A., Bayesian calibration of computer models, *Journal of the Royal Statistical Society: Series B*
18 (*Statistical Methodology*), 63(3):425–464, 2001.
- 19 4. Edeling, W., Cinnella, P., Dwight, R., and Bijl, H., Bayesian estimates of parameter variability in the k- ϵ turbulence model,
20 *Journal of Computational Physics*, 258:73–94, February 2014.
- 21 5. Xiao, H. and Cinnella, P., Quantification of model uncertainty in RANS simulations: A review, *Progress in Aerospace Sci-*
22 *ences*, 108:1–31, July 2019.
- 23 6. Nadiga, B., Jiang, C., and Livescu, D., Leveraging Bayesian analysis to improve accuracy of approximate models, *Journal of*
24 *Computational Physics*, 394:280–297, October 2019.
- 25 7. Doronina, O.A., Murman, S.M., and Hamlington, P.E., Parameter Estimation for RANS Models Using Approximate Bayesian
26 Computation, *arXiv:2011.01231 [physics]*, November 2020.

- 1 8. Damblin, G., Keller, M., Barbillon, P., Pasanisi, A., and Parent, É., Bayesian Model Selection for the Validation of Computer
2 Codes, *Quality and Reliability Engineering International*, 32(6):2043–2054, 2016.
- 3 9. del Val, A., Le Maître, O.P., Magin, T.E., Chazot, O., and Congedo, P.M., A surrogate-based optimal likelihood function for the
4 bayesian calibration of catalytic recombination in atmospheric entry protection materials, *Applied Mathematical Modelling*,
5 101:791–810, 2022.
- 6 10. Gally, T., Groche, P., Hoppe, F., Kuttich, A., Matei, A., Pfetsch, M.E., Rakowitsch, M., and Ulbrich, S., Identification of Model
7 Uncertainty via Optimal Design of Experiments applied to a Mechanical Press, *arXiv:1910.08408 [cs, math, stat]*, February
8 2020.
- 9 11. Maupin, K.A. and Swiler, L.P., Model Discrepancy Calibration Across Experimental Settings, *Reliability Engineering &*
10 *System Safety*, p. 106818, January 2020.
- 11 12. Higdon, D., Kennedy, M., Cavendish, J.C., Cafoe, J.A., and Ryne, R.D., Combining Field Data and Computer Simulations for
12 Calibration and Prediction, *SIAM Journal on Scientific Computing*, 26(2):448–466, January 2004.
- 13 13. Higdon, D., Gattiker, J., Williams, B., and Rightley, M., Computer Model Calibration Using High-Dimensional Output, *Jour-*
14 *nal of the American Statistical Association*, 103(482):570–583, 2008.
- 15 14. Arendt, P.D., Apley, D.W., Chen, W., Lamb, D., and Gorsich, D., Improving Identifiability in Model Calibration Using Mul-
16 tiple Responses, *Journal of Mechanical Design*, 134(10), October 2012.
- 17 15. Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafoe, J.A., Cavendish, J., Lin, C.H., and Tu, J., A Framework for Validation
18 of Computer Models, *Technometrics*, 49(2):138–154, May 2007.
- 19 16. Bayarri, M.J., Berger, J.O., Cafoe, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R.J., Paulo, R., Sacks, J., and
20 Walsh, D., Computer model validation with functional output, *The Annals of Statistics*, 35(5):1874–1906, October 2007.
- 21 17. Brynjarsdóttir, J. and O’Hagan, A., Learning about physical parameters: The importance of model discrepancy, *Inverse Prob-*
22 *lems*, 30(11):114007, October 2014.
- 23 18. Ling, Y., Mullins, J., and Mahadevan, S., Selection of model discrepancy priors in Bayesian calibration, *Journal of Computa-*
24 *tional Physics*, 276:665–680, November 2014.
- 25 19. Gardner, P., Rogers, T.J., Lord, C., and Barthorpe, R.J., Learning model discrepancy: A Gaussian process and sampling-based
26 approach, *Mechanical Systems and Signal Processing*, 152:107381, May 2021.
- 27 20. Plumlee, M., Bayesian Calibration of Inexact Computer Models, *Journal of the American Statistical Association*,
28 112(519):1274–1285, July 2017.
- 29 21. Gu, M. and Wang, L., Scaled Gaussian Stochastic Process for Computer Model Calibration and Prediction, *SIAM/ASA Journal*
30 *on Uncertainty Quantification*, 6(4):1555–1583, January 2018.
- 31 22. Xie, F. and Xu, Y., Bayesian Projected Calibration of Computer Models, *arXiv:1803.01231 [math, stat]*, February 2019.

- 1 23. Arhonditsis, G.B., Papantou, D., Zhang, W., Perhar, G., Massos, E., and Shi, M., Bayesian calibration of mechanistic aquatic
2 biogeochemical models and benefits for environmental management, *Journal of Marine Systems*, 73(1-2):8–30, September
3 2008.
- 4 24. DeCarlo, E.C., Mahadevan, S., and Smarslok, B.P., Bayesian Calibration of Aerothermal Models for Hypersonic Air Vehicles,
5 In *54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Boston, Massachusetts,
6 April 2013. American Institute of Aeronautics and Astronautics.
- 7 25. He, Y. and Xiu, D., Numerical strategy for model correction using physical constraints, *Journal of Computational Physics*,
8 313:617–634, May 2016.
- 9 26. Sargsyan, K., Huan, X., and Najm, H.N., Embedded Model Error Representation for Bayesian Model Calibration, *Internation-*
10 *al Journal for Uncertainty Quantification*, 9(4), 2019.
- 11 27. Tuo, R. and Wu, C.F.J., A theoretical framework for calibration in computer models: Parametrization, estimation and conver-
12 gence properties, *arXiv:1508.07155 [stat]*, August 2015.
- 13 28. Tuo, R. and Wu, C.F.J., Prediction based on the Kennedy-O’Hagan calibration model: Asymptotic consistency and other
14 properties, *arXiv:1703.01326 [math, stat]*, March 2017.
- 15 29. Liu, F., Bayarri, M.J., and Berger, J.O., Modularization in Bayesian analysis, with emphasis on analysis of computer models,
16 *Bayesian Analysis*, 4(1):119–150, March 2009.
- 17 30. Cox, D.D., Park, J.S., and Singer, C.E., A statistical method for tuning a computer code to a data base, *Computational Statistics*
18 *& Data Analysis*, 37(1):77–92, July 2001.
- 19 31. Carmassi, M., Barbillon, P., Keller, M., Parent, E., and Chiodetti, M., Bayesian calibration of a numerical code for prediction,
20 *arXiv:1801.01810 [stat]*, March 2019.
- 21 32. Stein, M.L., *Interpolation of Spatial Data*, Springer Series in Statistics, Springer New York, New York, NY, 1999.
- 22 33. Rasmussen, C.E. and Williams, C.K.I., *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learn-
23 ing, MIT Press, Cambridge, Mass, 2006.
- 24 34. Robert, C.P. and Casella, G., *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer, New York, NY, 2. ed.,
25 softcover reprint of the hardcover 2. ed. 2004 edition, 2010.
- 26 35. Yuan, J. and Ng, S.H., A sequential approach for stochastic computer model calibration and prediction, *Reliability Engineering*
27 *& System Safety*, 111:273–286, March 2013.
- 28 36. Yuan, J., Nian, V., Su, B., and Meng, Q., A simultaneous calibration and parameter ranking method for building energy
29 models, *Applied Energy*, 206:657–666, November 2017.
- 30 37. Yuan, J. and Ng, S.H., An Integrated Method for Simultaneous Calibration and Parameter Selection in Computer Models,
31 *ACM Transactions on Modeling and Computer Simulation*, 30(1):7:1–7:23, February 2020.

- 1 38. Gramacy, R.B., Bingham, D., Holloway, J.P., Grosskopf, M.J., Kuranz, C.C., Rutter, E., Trantham, M., and Drake, R.P., Cali-
2 brating a large computer experiment simulating radiative shock hydrodynamics, *The Annals of Applied Statistics*, 9(3):1141–
3 1168, September 2015.
- 4 39. Wu, X., Kozłowski, T., Meidani, H., and Shirvan, K., Inverse uncertainty quantification using the modular Bayesian approach
5 based on Gaussian Process, Part 2: Application to TRACE, *Nuclear Engineering and Design*, 335:417–431, August 2018.
- 6 40. Joseph, V.R. and Melkote, S.N., Statistical Adjustments to Engineering Models, *Journal of Quality Technology*, 41(4):362–
7 375, October 2009.
- 8 41. Wong, R.K.W., Storlie, C.B., and Lee, T.C.M., A frequentist approach to computer model calibration, *Journal of the Royal*
9 *Statistical Society: Series B (Statistical Methodology)*, 79(2):635–648, 2017.
- 10 42. Rumsey, K., Huerta, G., Brown, J., and Hund, L., Dealing with Measurement Uncertainties as Nuisance Parameters in
11 Bayesian Model Calibration, *SIAM/ASA Journal on Uncertainty Quantification*, pp. 1287–1309, January 2020.
- 12 43. Loeppky, J.L., Bingham, D., and Welch, W.J., Computer Model Calibration or Tuning in Practice, Tech. Rep., 2006.
- 13 44. Arendt, P.D., Apley, D.W., and Chen, W., Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifi-
14 ability, *Journal of Mechanical Design*, 134(10):100908, October 2012.
- 15 45. Kennedy, M.C. and O’Hagan, A. Supplementary details on Bayesian Calibration of Computer Models. 2001.
- 16 46. Berge, C., *Topological Spaces: Including a Treatment of Multi-Valued Functions, Vector Spaces, and Convexity*, Macmillan
17 Co., New York, 1963.
- 18 47. Tian, G. and Zhou, J., The maximum Theorem and the existence of Nash equilibrium of (generalized) games without lower
19 semicontinuities, *Journal of Mathematical Analysis and Applications*, 166(2):351–364, May 1992.
- 20 48. Andrieu, C. and Thoms, J., A tutorial on adaptive MCMC, *Statistics and Computing*, 18(4):343–373, December 2008.
- 21 49. Gamerman, D. and Lopes, H.F., *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, number 68 in
22 Texts in Statistical Science Series, Taylor & Francis, Boca Raton, 2nd ed edition, 2006.
- 23 50. Kommajosyula, R., Development and assessment of a physics-based model for subcooled flow boiling with application to
24 CFD, PhD thesis, Massachusetts Institute of Technology, 2020.
- 25 51. Kennel, W.E., Local boiling of water and superheating of high pressure steam in annuli, Thesis, Massachusetts Institute of
26 Technology, 1949.
- 27 52. Leoni, N., Bayesian inference of model error for the calibration of two-phase CFD codes, PhD thesis, Institut Polytechnique
28 de Paris, 2022.
- 29 53. Mchutchon, A. and Rasmussen, C., Gaussian Process Training with Input Noise, In *Advances in Neural Information Process-*
30 *ing Systems*, Vol. 24. Curran Associates, Inc., 2011.