Bayesian calibration with adaptive model discrepancy

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

Bayesian calibration with adaptive model discrepancy

Nicolas Leoni[a,b], Olivier Le Maître[c], Maria-Giovanna Rodio[b], Pietro Marco Congedo[a]

[a] *Inria, Centre de Mathématiques Appliquées, Ecole polytechnique, IPP, Route de Saclay, 91128 Palaiseau Cedex, France*
[b] *Commissariat à l'Energie Atomique et aux Energies Alternatives, ISAS, D36, 91191 Gif-sur-Yvette Cedex, France*
[c] *CNRS, Inria, Centre de Mathématiques Appliquées, Ecole polytechnique, IPP, Route de Saclay, 91128 Palaiseau Cedex, France*

Abstract

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.

Keywords: 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

Email address: nicolas.leoni@inria.fr (Nicolas Leoni)

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, 5, 6, 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, 13, 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, 18, 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. A frequentist approach can avoid this issue [25, 26]. Another approach to calibration in the presence of model error embeds representations to add uncertainties in the model parameters [27, 28].

In twenty years of applications of the Bayesian calibration framework, two classes of methods have emerged: fully Bayesian methods and modular methods. The fully Bayesian methods aim at solving the complete Bayesian formulation of the problem accounting for all uncertainties related to the data, the surrogate for the computer model if necessary, and the estimation of the model discrepancy. Although theoretically superior to others, the resolution of the full Bayesian problems is seldom practical due to their complexity and computational cost [12], and numerical approximation techniques are necessary to solve these problems in practice [13]. The modular methods tackle the complexity of the calibration by introducing simplified formulations of the fully Bayesian prob-

lem. Typically, they rely on sequential pointwise estimations of hyperparameters. These strategies significantly reduce the sample spaces' dimensionality, improving the whole computational efficiency of the calibration. The authors of [29], who coined the name, also claimed a theoretical advantage of modular approaches, as the separated sequential estimations avoid confounding issues.

The KOH framework is a modular approach based on a pointwise estimation of the model discrepancy before estimating the model parameters. As a result, the method underestimates the uncertainty related to the model discrepancy. Although generally considered a "second-order uncertainty," recent studies [19] show that underestimating uncertainties in the model discrepancy might impact the posterior predictions, which are of utmost importance to assess the calibration quality and predict new values with objective confidence intervals. Therefore, there is a need for new modular techniques that approximate the fully Bayesian calibration problem better while retaining the advantage of reduced sample spaces. We propose a new calibration method, named Full Maximum a Posteriori (FMP), which involves a parametric estimation of the model discrepancy hyperparameters. The FMP method leads to simple approximations of the full Bayesian parameters and predictive posterior distributions.

The organization of the paper is as follows. In section 2 we present the Bayesian calibration problem, introducing the general framework with model discrepancy in section 2.1. Sections 2.2 and 2.3 discuss the fully Bayesian and modular methods, respectively. In section 3, we introduce the FMP method, discussing its underlying hypotheses and providing the expressions of the resulting parameters' posterior and posterior predictive density. We show that the FMP estimation of the parameter has higher plausibility than any other modular approach. In section 3.3, we contrast the calibration results for the KOH and FMP for analytic examples consisting of a Gaussian or a mixture of Gaussians for the full Bayesian posterior. The superiority of the FMP method, compared to the KOH approach, is demonstrated in these analytical examples. In section 4, we consider more complex examples. The first one uses a simple model having two sets of parameter values yielding a comparable explanation of the observations. The second example concerns the calibration of a boiling model on experimental measurements. The FMP method proves superior to the KOH method in approaching the full Bayesian solution in the two examples.

2. Bayesian calibration of computer codes

2.1. General framework and reference solution

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

We denote $\{\mathbf{x}_i\}_{1\leq i\leq n}$ the set of experimental conditions corresponding to the observations in $\mathbf{y}_{\text{obs}}$.

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

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

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

A statistical model for the measurement error is needed. For simplicity, we shall assume that the measurement error does not depend on the condition $\mathbf{x}$ and that the errors of different observations are independent. For a Gaussian 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 are standard and used in many calibration studies, such as [30].

As discussed below, a Bayesian procedure can be employed to learn the parameters $\boldsymbol{\theta}$ with the observation model in (1). The procedure updates the prior distribution of $\boldsymbol{\theta}$ to obtain its posterior distribution and, subsequently, posterior predictions using the computer code. The Bayesian procedure was successfully applied to numerous calibration problems. However, there are situations where it yields inconsistent results, with residuals (differences between posterior predictions and observations) that do not follow the assumed measurement error distribution. Such inconsistencies may have two causes: an incorrect measurement error distribution and an observation model that is too poor. This work considers situations with adequate measurement error distribution and focuses on the observations model. For instance, the residuals may be highly correlated with significant mean values when the measurement errors are known to be unbiased and independent for different x_i. These inconsistencies are interpreted in the work of Kennedy and O'Hagan [3] as a model error that induces an irreducible gap between predictions and the actual process. The authors proposed to account for the model error by completing the observation model with a model discrepancy term $z(\mathbf{x})$, leading to the following calibration equation:

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

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

More statistical assumptions are needed to proceed with the calibration of the parameters. First, measurement error and model discrepancy are usually supposed independent: $z(\mathbf{x}_i) \perp\!\!\!\perp \epsilon(\mathbf{x}_i)$. Second, the model discrepancy is 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 process. The covariance of z depends on some hyperparameters $\boldsymbol{\psi}_z$. For the prior mean, we set $\mu = 0$ as advocated in [12, 31]

for better identifiability. However, $\mu(\mathbf{x})$ can be expressed as a linear combination of basis functions in the context of Universal Kriging [32].

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 if model error is considered or not in the calibration. The space of hyperparameters is $\Psi \subset \mathbb{R}^h$.

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

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

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

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

where $\mathbf{f}_{\boldsymbol{\theta}} = (f(\boldsymbol{\theta}, \mathbf{x}_1), f(\boldsymbol{\theta}, \mathbf{x}_2), \ldots, f(\boldsymbol{\theta}, \mathbf{x}_n))^T$ is the vector of evaluations of the computer code at the observation points, and $\mathbf{I}_n$ is the identity matrix of size n.

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

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

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

The choice of prior distributions in Bayesian statistics is not straightforward and affects the calibration results. Model parameters and hyperparameters are generally considered independent a priori: $\mathrm{p}(\boldsymbol{\theta}, \boldsymbol{\psi}) = \mathrm{p}(\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\psi})$. Model parameters might be physical quantities with a range of plausible values in the literature. Consequently, it is often possible to formulate informative priors for them. For the hyperparameters of the model discrepancy, classical choices are uniform priors and Jeffrey's prior (e.g., $\mathrm{p}(\sigma) \propto 1/\sigma$). Inverse-gamma priors may be employed for the hyperparameters fixing the correlation with $\mathbf{x}$ of z, as they put zero probability mass on zero and infinite correlation lengths to separate model error from measurement error.

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

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

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

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

The distribution $\mathrm{p}(y(\mathbf{x}_*)|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y}_{\mathrm{obs}})$ is normal, obtained from the Gaussian Process predictive equations, which writes [33]:

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

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

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

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

2.2. Fully Bayesian approaches

An example of application of full Bayesian method is in [12], where the full joint density is sampled using the Metropolis-Hastings algorithm in dimension six (one parameter and five hyperparameters), allowing a complete representation of the posterior uncertainty. In [13], the authors deal with a high-dimensional output code and a sample space with a dimension greater than 30. Besides, more than 18,000 observations are considered making one evaluation of the likelihood function costly. Consequently, the authors relied on reduced basis representations, block matrix inversion formulas, and adapted informative priors to make the problem tractable and ease the posterior sampling.

Another example of a fully Bayesian technique is presented in [15], where the authors chose a covariance function for z that involves variance and correlation

length hyperparameters. Before the sampling, correlation lengths are estimated with ML, as well as any eventual prior mean for z. ML estimates of variance hyperparameters serve to construct a prior for them. The MCMC sampling itself is simplified by allowing only variance hyperparameters to vary (as well as other hyperparameters involved in the surrogate for the computer model), using Gibbs updates as a conjugate prior was chosen. The authors declare that combining ML estimates with sampling achieves an efficient approximation to the Full Bayesian calibration. Note that they claim that ML estimates are a sufficient approximation to Full Bayesian, but only concerning hyperparameters for the computer model surrogate; they recognize that the variations of model discrepancy hyperparameters must be considered. Conjugate priors are also central to the sampling with Gibbs steps proposed in [17].

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

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

2.3. Modular approaches

The principle underlying modular approaches is to separate into groups (modules) the treatment of parameters and hyperparameters during the calibration and obtain a problem of sequential estimations. In [29], three modules are considered in the calibration of computer models: the surrogate of f, the measurement error in the observations, and the model discrepancy. Each module is then separately estimated. This approach breaks the high-dimensional calibration problem into smaller ones. One expects that these separated problems are more accessible to sample than the original problem. The modular approach can also improve the identifiability of the hyperparameters belonging to different modules.

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

In the first class of methods, one starts by determining a value $\boldsymbol{\theta}_0$ such that $f(\mathbf{x}, \boldsymbol{\theta}_0)$ is an accurate prediction of $y(\mathbf{x})$. Then, the hyperparameters are estimated, usually by Maximum Likelihood, considering $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. With our notations, this amounts to the following estimator:

$$\widehat{\boldsymbol{\psi}}_1 = \arg\max_{\Psi} \mathrm{p}(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\psi}, \boldsymbol{\theta}_0). \tag{9}$$

The methods in the first class differ by the definition of $\boldsymbol{\theta}_0$: some works advocate for elicited values such as the prior mean [29, 39]. Other common approaches estimate $\boldsymbol{\theta}_0$ from the observation, either minimizing the residuals between the model prediction and the observations [40], or solving the calibration problem without model error (1) to propose a plausible value [41, 11, 42, 19]. These approaches inherently assess the relevance of introducing a model discrepancy term by judging the quality of the model predictions using $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. In [43], it is demonstrated that $f(\mathbf{x}, \boldsymbol{\theta}_0) + z(\mathbf{x})$ can be an accurate prediction of $y(\mathbf{x})$, even for a poor value of $\boldsymbol{\theta}_0$, because the model discrepancy term z can compensate for a wide range of residuals.

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

$$\widehat{\boldsymbol{\psi}}_2 = \arg\max_{\Psi} \mathrm{p}(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\psi}). \tag{10}$$

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

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

and the following estimation for the parameters posterior distribution:

$$\mathrm{p}_{\mathrm{KOH}}(\boldsymbol{\theta}) \propto \mathrm{p}(\boldsymbol{\theta})\mathrm{p}(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\theta}, \widehat{\boldsymbol{\psi}}_{\mathrm{KOH}}). \tag{12}$$

Note also that an approximation of $\mathrm{p}(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\psi})$ is sometimes used to reduce the computational cost; Kennedy and O'Hagan propose to use a Gaussian approximation in [45]. Compared to the first class of methods, the computation of $\mathrm{p}(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\psi})$ is generally more costly than estimating a single parameter value $\boldsymbol{\theta}_0$. Compared to the first class, the second class of methods benefits from

the complete account of the parameter uncertainty when estimating the model discrepancy.

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

3. The Full Maximum A Posteriori method

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

3.1. Calibration framework

3.1.1. Fundamental hypotheses

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

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

Therefore, the calibration equation becomes:

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

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

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

where FMP stands for "Full Maximum A Posteriori", a combination between full Bayesian analysis and MAP estimation. Two additional assumptions are necessary to ensure the approximation of the posterior:

1. For all $\boldsymbol{\theta} \in \Theta$, the distribution $\text{p}(\boldsymbol{\psi}|\boldsymbol{\theta}, \mathbf{y}_{\text{obs}})$ can be approximated as a point mass distribution at its mode $\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})$.

2. Let $g(\boldsymbol{\theta}, \boldsymbol{\theta}')$ be defined, for $(\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta^2$, as :

$$g(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathrm{p}(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}')).$$

We assume that, for all $\boldsymbol{\theta} \in \Theta$,

$$\int_{\Theta} g(\boldsymbol{\theta}, \boldsymbol{\theta}')\mathrm{p}(\boldsymbol{\theta}'|\mathbf{y}_{\mathrm{obs}})\, \mathrm{d}\boldsymbol{\theta}' \propto\!\!\!\sim g(\boldsymbol{\theta}, \boldsymbol{\theta}). \tag{16}$$

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

The second assumption supposes that $g(\boldsymbol{\theta}, \cdot)$ has a similar shape for different $\boldsymbol{\theta}$ and that the ratio between its posterior-averaged value and its maximal value (remark that $\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta}' \in \Theta} g(\boldsymbol{\theta}, \boldsymbol{\theta}')$) does not depend on $\boldsymbol{\theta}$. This is verified for example when $\hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta})$ is a slowly varying function of $\boldsymbol{\theta}$ (one value of optimal hyperparameters stands out), or when, for two different values of optimal hyperparameters $\hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}_1)$ and $\hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}_2)$, the conditional densities $\mathrm{p}(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}}, \hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}_1))$ and $\mathrm{p}(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}}, \hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}_2))$ are close.

3.1.2. Posterior approximation and predictive density

With the previous assumptions, the parameter posterior can be expressed as

$$\begin{aligned}
\mathrm{p}(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}}) &= \int_{\Psi} \mathrm{p}(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y}_{\mathrm{obs}})\mathrm{p}(\boldsymbol{\psi}|\mathbf{y}_{\mathrm{obs}})\, \mathrm{d}\boldsymbol{\psi} \\
&= \int_{\Psi} \mathrm{p}(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y}_{\mathrm{obs}}) \left(\int_{\Theta} \mathrm{p}(\boldsymbol{\psi}|\boldsymbol{\theta}', \mathbf{y}_{\mathrm{obs}})\mathrm{p}(\boldsymbol{\theta}'|\mathbf{y}_{\mathrm{obs}})\, \mathrm{d}\boldsymbol{\theta}' \right) \mathrm{d}\boldsymbol{\psi} \\
&= \int_{\Theta} \mathrm{p}(\boldsymbol{\theta}'|\mathbf{y}_{\mathrm{obs}}) \left(\int_{\Psi} \mathrm{p}(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y}_{\mathrm{obs}})\mathrm{p}(\boldsymbol{\psi}|\boldsymbol{\theta}', \mathbf{y}_{\mathrm{obs}})\, \mathrm{d}\boldsymbol{\psi} \right) \mathrm{d}\boldsymbol{\theta}' \\
&\approx \int_{\Theta} \mathrm{p}(\boldsymbol{\theta}'|\mathbf{y}_{\mathrm{obs}})\mathrm{p}(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}}, \hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}'))\, \mathrm{d}\boldsymbol{\theta}' \\
&\propto\!\!\!\sim \int_{\Theta} \mathrm{p}(\boldsymbol{\theta}'|\mathbf{y}_{\mathrm{obs}})\mathrm{p}(\boldsymbol{\theta})\mathrm{p}(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}'))\, \mathrm{d}\boldsymbol{\theta}' \\
&\propto\!\!\!\sim \mathrm{p}(\boldsymbol{\theta})\mathrm{p}(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta})).
\end{aligned} \tag{17}$$

We denote $\mathrm{p}_{\mathrm{FMP}}$ the probability density over $\boldsymbol{\theta}$ proportional to $\mathrm{p}(\boldsymbol{\theta})\mathrm{p}(\mathbf{y}_{\mathrm{obs}}|\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}))$. From the previous equation, $\mathrm{p}_{\mathrm{FMP}}$ is a reasonable approximation of $\mathrm{p}_{\mathrm{Bayes}}$.

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 Theorem [46, 47] it can be shown that $\mathrm{p}_{\mathrm{FMP}}$ is a continuous function of $\boldsymbol{\theta}$. With the additional assumption that for each value of $\boldsymbol{\theta}$ there is a unique value of optimal hyperparameters (which is

not true in general), we have that $\hat{\boldsymbol{\psi}}_{\mathrm{FMP}}$ is a continuous function of $\boldsymbol{\theta}$. These continuity properties pave the way for calibration techniques based on surrogate 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:

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

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

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

and

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

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

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

The FMP framework avoids ambiguity in defining the "true value" of parameters. Nevertheless, other types of identifiability issues might still plague the calibration process. If two distinct values of $\boldsymbol{\theta}$ with the same prior value lead to the same computer model prediction, one might deem them unidentifiable. Likewise, if the prior distribution of z is not suited to the actual model discrepancy, optimal hyperparameters might be meaningless and posterior samples of z

would be inaccurate. These issues can be addressed by testing several prior distributions for $\boldsymbol{\theta}$, z, and $\boldsymbol{\psi}$. The multivariate Gaussian example in section 3.3.2 below gives a graphical representation of these identifiability cases.

3.2. Marginal likelihood of the FMP approximation

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

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 $\mathrm{p}_h(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) \propto \mathrm{p}(\boldsymbol{\theta})\mathrm{p}(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta},\boldsymbol{\psi} = \boldsymbol{\psi}_h(\boldsymbol{\theta}))$. This setting also encompasses single-point estimation methods, such as the KOH or cross-validation estimator of the hyperparameters. We have:

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

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

3.3. Elementary examples

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

A Gaussian posterior is unlikely in practice because priors on hyperparameters are rarely Gaussian. However, the Gaussian case is very instructive, and posterior approximations by Gaussian distribution are standard when the posteriors become increasingly sharp around their maximum as the number of observations increases. Similarly, the Gaussian mixture can approximate more complex distributions, and this example helps us understand the case of a multimodal posterior distribution where alternative explanations of the observations are plausible.

3.3.1. Unimodal Gaussian posterior

Let us assume that the Bayesian procedure leads to a joint Gaussian posterior for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. We recall that $\dim(\boldsymbol{\theta}) = p$ 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 covariance, symmetric and positive definite. The joint posterior distribution writes:

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

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

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

which directly gives:

$$\mathrm{p}_{\mathrm{Bayes}}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathbf{V}_{\boldsymbol{\theta}}). \tag{23}$$

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

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

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

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

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

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

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

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

Note the *reduction of variance* property: because the matrix $\mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}}^T\mathbf{V}_{\boldsymbol{\psi}}^{-1}\mathbf{C}_{\boldsymbol{\theta},\boldsymbol{\psi}}$ is symmetric positive definite, under the Loewner order we have $\mathbf{V}_{\boldsymbol{\theta}|\boldsymbol{\psi}} \leq \mathbf{V}_{\boldsymbol{\theta}}$.

This order can be understood as the following: let $\{\lambda_i\}_{1\leq i\leq p}$ be the eigenvalues of $\mathbf{V}_{\boldsymbol{\theta}|\boldsymbol{\psi}}$ in a decreasing order and $\{\lambda_i^{'}\}_{1\leq i\leq p}$ be the eigenvalues of $\mathbf{V}_{\boldsymbol{\theta}}$ in a decreasing order. Then, for all $i \leq p$, we have $\lambda_i \leq \lambda_i^{'}$. It is seen that $\mathrm{p}_{\mathrm{KOH}}$ remains Gaussian, with correct mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$, but a reduced covariance compared to $\mathrm{p}_{\mathrm{Bayes}}$.

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),

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

Thus $\hat{\boldsymbol{\psi}}_{\mathrm{FMP}}$ 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],

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

we obtain

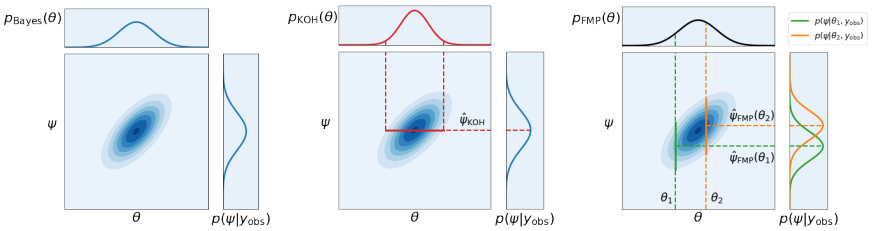
$$\begin{aligned}
\mathrm{p}_{\mathrm{FMP}}(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}}) &\propto \mathrm{p}(\boldsymbol{\theta}, \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta})|\mathbf{y}_{\mathrm{obs}}) \\
&\propto \exp\left(-\frac{1}{2}\begin{pmatrix} \boldsymbol{\theta}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

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

The behaviour of the methods is illustrated in figure 1 for $\dim(\boldsymbol{\theta}) = \dim(\boldsymbol{\psi}) = 1$. In this example, the FMP method matches exactly the Bayesian solution. The KOH solution has a reduced variance due to the conditioning on a fixed value of the hyperparameters, which might lead to the "false certitude" effect. This higher certitude (lower uncertainty) is simply an artifact of the KOH estimation method.

3.3.2. Mixture of Gaussians

We now consider the case where the joint posterior $\mathrm{p}(\boldsymbol{\gamma}|\mathbf{y}_{\mathrm{obs}})$ is a mixture of Gaussians, with modes that are well-separated, with similar orders of magnitude. The joint density is a mixture of m Gaussians, with weights $(\pi_i)_{i\leq m}$, such



(a) Full Bayesian posterior (b) KOH posterior (c) FMP posterior

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

that $\sum_{i=1}^m \pi_i = 1$. Their respective means are $\boldsymbol{\mu}_i = \begin{pmatrix} \mu_{\boldsymbol{\theta},i} \\ \mu_{\boldsymbol{\psi},i} \end{pmatrix}$ and their respective covariance matrices: $\boldsymbol{\Lambda}_i = \begin{pmatrix} \mathbf{V}_{i,\boldsymbol{\theta}} & \mathbf{C}_{i,\boldsymbol{\theta},\boldsymbol{\psi}}^T \\ \mathbf{C}_{i,\boldsymbol{\theta},\boldsymbol{\psi}} & \mathbf{V}_{i,\boldsymbol{\psi}} \end{pmatrix}$. The joint density is

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

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 $1 \leq j \leq m$,"

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

$$\begin{aligned} \mathrm{p}_{\text{Bayes}}(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) &= \int_{\boldsymbol{\psi}} \mathrm{p}(\boldsymbol{\gamma}|\mathbf{y}_{\text{obs}}) \, \mathrm{d}\boldsymbol{\psi} = \sum_{i=1}^m \pi_i \int_{\boldsymbol{\psi}} \mathrm{p}(\boldsymbol{\gamma}|\mathbf{y}_{\text{obs}}, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) \, \mathrm{d}\boldsymbol{\psi} \\ &= \frac{1}{(2\pi)^{p/2}} \sum_{i=1}^m \frac{\pi_i}{\sqrt{|\mathbf{V}_{i,\boldsymbol{\theta}}|}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta},i})^T \mathbf{V}_{i,\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \mu_{\boldsymbol{\theta},i})\right). \end{aligned} \tag{29}$$

The KOH estimation first computes the hyperparameter marginal distribution:

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

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

$$\begin{aligned}\hat{\boldsymbol{\psi}}_{\mathrm{KOH}} &= \arg\max_{\boldsymbol{\psi}} \mathrm{p}(\boldsymbol{\psi}|\mathbf{y}_{\mathrm{obs}}) = \boldsymbol{\mu}_{i_{\mathrm{KOH}},\boldsymbol{\psi}},\\ i_{\mathrm{KOH}} &= \arg\max_{i\leq m} \frac{\pi_j}{\sqrt{\mathbf{V}_{i,\boldsymbol{\psi}}}}.\end{aligned} \tag{30}$$

So, the KOH posterior is:

$$\mathrm{p}_{\mathrm{KOH}}(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}}) = \mathrm{p}(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}}, \boldsymbol{\psi} = \boldsymbol{\mu}_{i_{\mathrm{KOH}},\boldsymbol{\psi}}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta},i_{\mathrm{KOH}}}, \mathbf{V}_{i_{\mathrm{KOH}},\boldsymbol{\theta}|\boldsymbol{\psi}}), \tag{31}$$

with

$$\mathbf{V}_{i_{\mathrm{KOH}},\boldsymbol{\theta}|\boldsymbol{\psi}} = \mathbf{V}_{i_{\mathrm{KOH}},\boldsymbol{\theta}} - \mathbf{C}^T_{i_{\mathrm{KOH}},\boldsymbol{\theta},\boldsymbol{\psi}} \mathbf{V}_{i_{\mathrm{KOH}},\boldsymbol{\psi}} \mathbf{C}_{i_{\mathrm{KOH}},\boldsymbol{\theta},\boldsymbol{\psi}}.$$

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

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

$$\hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\psi}} \mathrm{p}(\boldsymbol{\gamma}|\mathbf{y}_{\mathrm{obs}}) = \arg\max_{\boldsymbol{\psi}} \sum_{i=1}^{m} \frac{\pi_i}{\sqrt{|\boldsymbol{\Lambda}_i|}} \exp\left(-\frac{1}{2}(\boldsymbol{\gamma}-\boldsymbol{\mu}_i)^T \boldsymbol{\Lambda}_i^{-1}(\boldsymbol{\gamma}-\boldsymbol{\mu}_i)\right)$$

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

$$\begin{aligned}\text{for} \quad \boldsymbol{\theta} \in \Theta_i, \quad & \mathrm{p}(\boldsymbol{\gamma}|\mathbf{y}_{\mathrm{obs}}) \approx \pi_i \mathrm{p}(\boldsymbol{\gamma}|\mathbf{y}_{\mathrm{obs}}, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i),\\ \text{and} \quad & \hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta}) = \hat{\boldsymbol{\psi}}_{i,\mathrm{FMP}}(\boldsymbol{\theta}) := \boldsymbol{\mu}_{i,\boldsymbol{\psi}} + \mathbf{C}_{i,\boldsymbol{\theta},\boldsymbol{\psi}} \mathbf{V}^{-1}_{i,\boldsymbol{\theta}} (\boldsymbol{\theta} - \boldsymbol{\mu}_{i,\boldsymbol{\theta}}).\end{aligned}$$

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

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

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

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

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

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

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

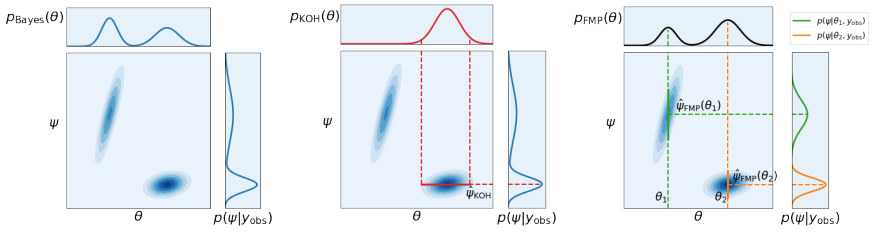
$$\mathrm{p}_{\mathrm{FMP}}(\boldsymbol{\theta}) = \frac{1}{K}\sum_{i=1}^{m} \frac{\pi_i}{\sqrt{|\boldsymbol{\Lambda}_i|}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu}_{i,\boldsymbol{\theta}})^T \mathbf{V}_{i,\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_{i,\boldsymbol{\theta}})\right)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

with the normalizing constant

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

The distribution $\mathrm{p}_{\mathrm{FMP}}$ is a linear combination of the m marginals associated with the original Gaussian mixture. The variances of the different peaks are correctly estimated, but the estimation introduces a bias on the weights of the Gaussian mixture: the FMP estimates the original weights π_i to be $\pi_i' = \frac{1}{k}\pi_i \frac{\sqrt{|\mathbf{V}_{i,\boldsymbol{\theta}}|}}{\sqrt{|\boldsymbol{\Lambda}_i|}}$ where k is a normalizing constant such that $\sum_{i=1}^{m}\pi_i' = 1$. Specifically, the FMP method emphasizes the importance of the Gaussians with low hyperparameter variance in the mixture.

An illustration of the methods is given in figure 2. In this example, the KOH estimation finds only one peak that does not correspond to the true maximum of the posterior, so the conclusions about $\boldsymbol{\theta}$ are potentially misleading. In the FMP estimation, all the peaks are found, yet their relative weights might be wrongly estimated, leading to an inversion of the importance of the peaks so that the maximum a posteriori estimator is incorrect. This feature will be found in any method that uses estimated hyperparameters instead of marginalization because volume effects cannot be seen. We argue that it is still favorable to perform the FMP estimation in this case so that no possible explanation of the observations is missed.



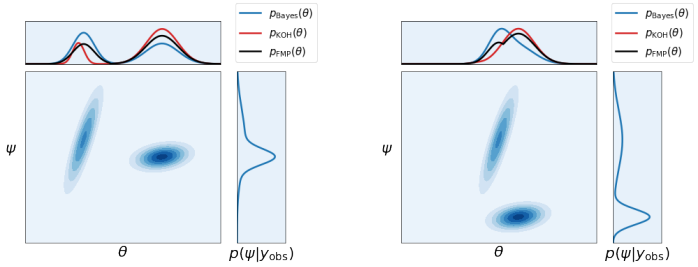
(a) Full Bayesian posterior (b) KOH approximation (c) FMP approximation

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

We now discuss the case where the two Gaussians of the mixture are not well-separated. In figure 3a, the Ψ-projections of the two modes overlaps. The

corresponding estimations of the KOH and FMP methods are plotted in the figure. In this example, the KOH criterion selects the right mode; contrary to the well-separated case, the other (left) mode contributes to the parameter posterior because part of its probability mass is captured when conditioning on $\hat{\boldsymbol{\psi}}_{\mathrm{KOH}}$. However, KOH badly estimates the second mode, contrary to the FMP approximation, which retrieves the two modes correctly. This situation occurs when a single model discrepancy adequately applies to the whole parameters domain.

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

4. Applications

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

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

4.1. Calibration of an inadequate model

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

4.1.1. Problem formulation

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

$$f(x, \theta) = x \sin(2\theta x) + (x + 0.15)(1 - \boldsymbol{\theta}). \tag{33}$$

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 observations of the true function at 8 points uniformly spaced in the interval $[0, 1]$. The observations are noisy with a centered Gaussian noise with standard deviation $\sigma_\epsilon = 0.1$. The observation noise is not known a priori but is learned along with the other hyperparameters of the model error.

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

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

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

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

4.1.2. Calibration Results

We apply a Kernel Density Estimation method to the MCMC samples to estimate the posterior marginal densities of parameters and hyperparameters. These marginals are shown in figure 4. The parameter's posterior marginals are bimodal with a first mode located at $\theta = -0.025$ and the second at $\boldsymbol{\theta} = 1.02$. For the reference Bayes solutions, the relative importance of the two modes is in the ratio 4.3:1, computed as the relative frequency of $\theta < 0.5$. The ratio is 2.75:1 for the FMP method and 69.4:1 for the KOH method. The FMP method overestimates the second mode's importance and puts more probability mass between the two modes. The KOH estimation almost misses the second mode focusing on a single interpretation of the observations.

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

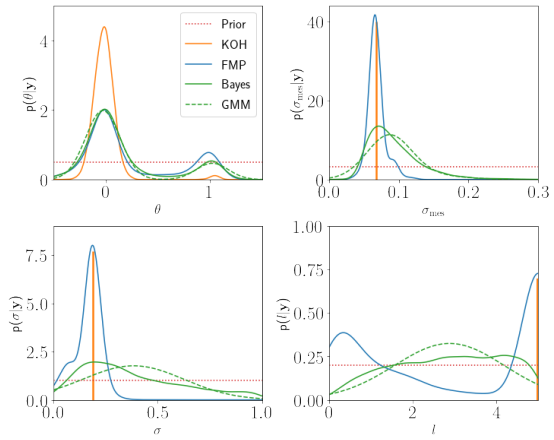


Figure 4: Prior and posterior distributions for the model parameter (top left) and hyperparameters. In the KOH calibration, hyperparameters are estimated with point mass distributions, represented with vertical lines. The Gaussian Mixture Model, fitted on the Bayes posterior, is represented with the 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.

The bimodality of the parameter posterior reflects in the posterior model

predictions shown in figure 5. The left mode, the only one found by KOH, corresponds to quasi-linear predictions with a constant offset from the true process $y(x)$. The second modes correspond to predictions close to the true process but with an oscillation in x. The FMP method accounts for the two types of predictions present in the Bayes solution. We also see a clear separation between the two types of predictions in the Bayes solutions, while the FMP method produces some predictions with intermediate structures.

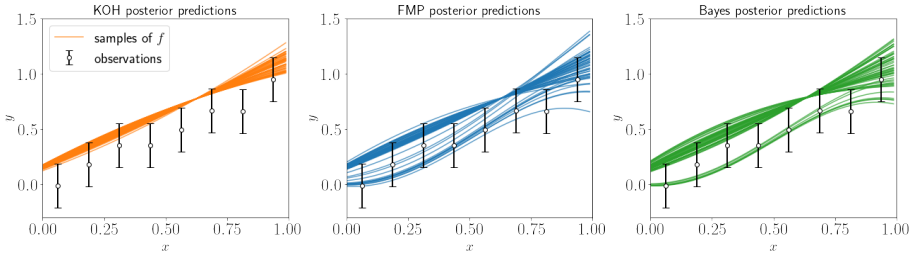


Figure 5: Posterior samples of the model predictions for all three calibration methods. 50 samples are represented on each figure, and observations are shown with $2\sigma_\epsilon$ confidence intervals. Model predictions that are quasi-linear correspond to $\theta \approx 0$, and the ones with slow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

4.1.3. Gaussian Mixture Model fit on the Bayes solution

As a further step, we fit a mixture of Gaussians model on the MCMC samples of the Bayes solution. To do so, we employed a hard clustering algorithm based on Expectation-Maximization [1]. The selected form of the mixture is a weighted sum of two Gaussians

$$\mathrm{p}(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{y}_{\mathrm{obs}}) = \pi_1 \, \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{C}_1) + \pi_2 \, \mathcal{N}(\boldsymbol{\mu}_2, \mathbf{C}_2). \tag{35}$$

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

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

- The KOH estimation selects a mode according to the criterion in equation (30). We check that $\frac{\pi_1}{\sqrt{|\mathbf{C}_1^{[3,3]}|}} = 70.8$ is greater than $\frac{\pi_2}{\sqrt{|\mathbf{C}_2^{[3,3]}|}} = 10.0$, which is the reason why it selects the first mode.

[1] `https://perso.telecom-paristech.fr/bonald/documents/gmm.pdf`

Coefficient	Estimation
π_1	0.81
π_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 $\boldsymbol{\mu} = (\theta, \sigma_{\text{mes},\Delta T}, \sigma, l)$.

- The posterior correlation between θ and $\boldsymbol{\psi}$ is low, so the variance at the first mode is roughly unchanged after projection: $\mathbf{V}_{\boldsymbol{\theta}} = \mathbf{C}_1(0,0) =$ 2.90e−2, and $\mathbf{V}_{\boldsymbol{\theta}|\boldsymbol{\psi}} = \mathbf{C}_1(0,0) - \mathbf{C}_{1,\theta,\boldsymbol{\psi}}^T(\mathbf{C}_1^{[3,3]})^{-1}\mathbf{C}_{1,\theta,\boldsymbol{\psi}} =$ 2.89e−2, 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 selecting only one mode.

- The FMP weights can be computed using (32). We have $\pi_i^{'} \propto \pi_i \frac{\sqrt{|\mathbf{C}_i|}}{\sqrt{\mathbf{C}_i(0,0)}}$, which gives $\pi_1^{'} = 0.73$ and $\pi_2^{'} = 0.27$, in a proportion 2.71:1 which corresponds to the ratio found in the FMP results.

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

4.2. Calibration of a boiling model

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

4.2.1. Problem formulation

The MIT Boiling model [50] simulates the boiling of fluids in contact with a heated wall. The inputs of the model are experimental variables (fluid pressure, velocity, temperature, and configuration geometry) and the wall superheat ΔT_{sup}, defined as the difference between the wall temperature and the saturation temperature of the fluid. The model output is the heat flux ϕ from the

wall to the liquid. The model is semi-empirical and based on experimental correlations, with low evaluation cost. The experimental observations used in this study come from the boiling experiment of Kennel [51], precisely case number 6 (one of the observation sets used by Kommajosyula to validate the MIT Boiling model). It consists of 8 joint measurements of heat flux and wall temperature in different boiling regimes, ranging from no-boiling to fully developed nucleate boiling.

Three calibration parameters $(\theta_1, \theta_2, \theta_3)$ are considered; the motivations for this choice and the significance of each parameter is discussed in [52]. The measurement uncertainty lies primarily on ΔT_{sup}, which is an input model quantity, not an output. This specificity requires an adaptation of the framework. Following [53], the sensitivity (derivatives) of $\mathbf{y}_{\text{obs}}$ at the uncertain inputs is estimated from a polynomial fit over the observations and used to include this uncertainty in the framework (see below).

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

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

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

Table 2 reports the prior distributions considered for the calibration. The priors are truncated normals and uniform distributions for the parameters and hyperparameters, respectively. The priors' ranges are selected to ensure physically meaningful value (parameters) and using a priori considerations on the magnitude of the model and measurement errors (hyperparameters).

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

A polynomial fit of the model is used to estimate derivatives. The FMP

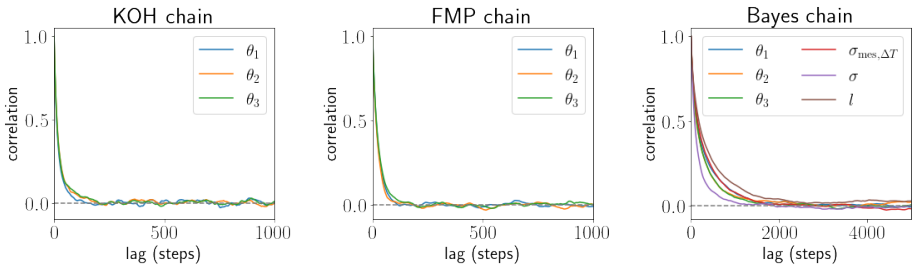


Figure 6: Self-correlation of the chains for the three methods applied to the MIT Boiling model calibration.

optimizations use the C++ library NLopt[2], with the global algorithm MLSL, using the local optimizer SBPLX. A maximal time of 10e−4s is allocated to individual optimization. For the KOH estimation of hyperparameters, the integral over Θ is computed by quadrature, using a QMC grid of size 300 over the parameters domain.

4.2.2. Results

Table 2 summarizes the results of the inference problem for the three methods. The most noticeable result is that the KOH method estimates a zero model error σ and attributes all the discrepancies to measurement uncertainty. Consistently, the KOH method yields an erroneous estimation of θ_1 with an underestimated variance. On the contrary, the FMP method identifies a model error comparable to the full Bayesian calibration and, despite underestimating the 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 ± 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 2: 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 $I_2 = [0, 1.5]$.

Figure 7 shows the posterior of the corrected model prediction $(f + z_{\boldsymbol{\theta}})$ of the true process y. The credible intervals of the corrected prediction in FMP and Bayes are generally larger than for KOH, especially at low values of ΔT_{sup} where

[2] `https://nlopt.readthedocs.io/en/latest/`

no observations are available. Note that the uncertainty is zero at ΔT_{sup} because of the structure of the model discrepancy covariance. The predictive variance is split into the model and residual contributions, as described in Section 3.1.2. For KOH, the variance a posteriori of $z_{\boldsymbol{\theta}}$ is uniformly zero, as σ is inferred to be zero. This result is unsatisfactory, especially at low values of ΔT_{sup} where few observations are available. All approaches agree on the dominance of the model predictions variance for the high range of ΔT_{sup} values.

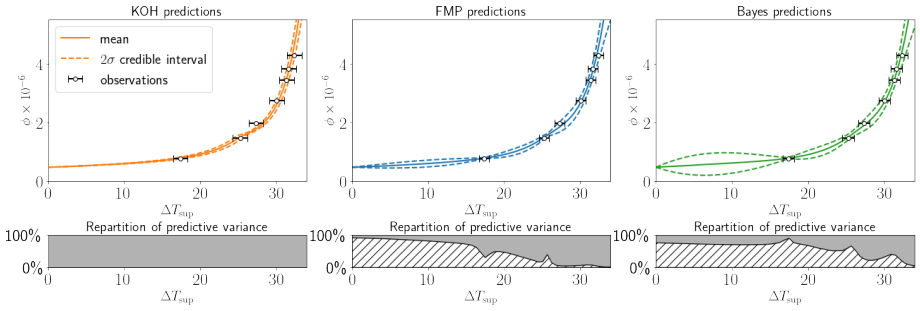


Figure 7: Posterior predictions of the corrected model ($f + z_{\boldsymbol{\theta}}$) with uncertainty. Observations are shown with inferred error bars of length $2\mathbb{E}_{\boldsymbol{\theta}}[\sigma_{\text{mes},\Delta T}]$. The bottom plots show the decomposition of the predictive variance in residual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

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

5. Conclusions

We have proposed an approach to estimate both model parameters and model discrepancy. For the first time in literature, to our knowledge, the model

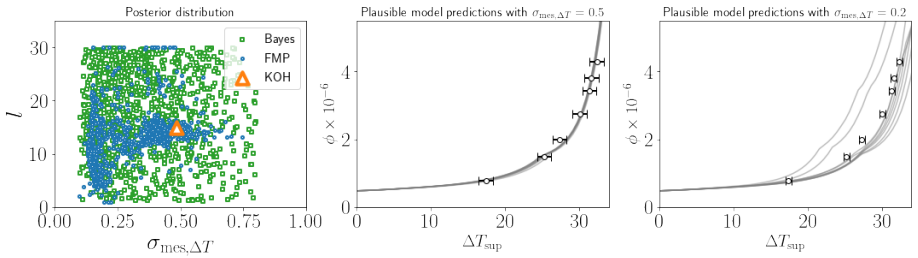


Figure 8: Samples in the σ_ϵ and l plane for the three methods (left plot). Samples of the FMP model predictions conditioned on a zero (middle) and non-zero model error (right).

error is explicitly dependent on model parameter values in the calibration framework. We showed that it solves the identifiability problem between model error and parameter uncertainty. The resulting predictive uncertainty splits into two natural contributions: the error in the calibrated model and the residual uncertainty. The FMP approximation has a reduced cost compared to the full Bayesian inference while correcting the defects of the KOH estimations when the posterior of the hyper-parameters has several modes.

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

Reducing the sample space dimension comes at the cost of solving an optimization problem at each step of the MCMC. From there comes the main computational cost of the FMP technique. Due to the relatively low dimensionality of the two application problems presented in this paper, we have not observed significant computational benefits of using the FMP method compared to the full Bayesian approach in our numerical experiments. In [52], we build a surrogate model of the optimal hyperparameters to bypass the optimization step and accelerate the FMP method with considerable computational savings compared to the full Bayesian method.

Our calibration equation offers new challenges to propose appropriate priors. Consistently with their definition of model error, previous works used model discrepancy hyperparameters a priori independent of model parameters. We also adopt this hypothesis in the present work, but one could imagine more complex prior. We advocate for using model discrepancy distributions that are more complex and more physics-informed to provide more accurate predictions.

Finally, the FMP method solves an optimization problem for the pointwise

estimation of the hyperparameters at every parameter value (see equation (15)). Solving these optimization problems may induce a significant computational overhead compared to other modular methods relying on global pointwise estimates. We are currently developing methods that construct surrogate models of the FMP hyperparameters $\hat{\boldsymbol{\psi}}_{\mathrm{FMP}}(\boldsymbol{\theta})$ in an offline stage. The sampling of the FMP posterior can then proceed with the surrogate at a cost comparable to the KOH method. These numerical developments will be published elsewhere.

SUPPLEMENTARY MATERIAL

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

Acknowledgements

This work was partly funded by the French Innovation Agency for Defense (Agence de l'Innovation de Défense, AID). N. Leoni would also like to thank Ravikishore Kommajosyula for providing the boiling model, and Kathryn Maupin for helpful discussions.
This project was partly funded from the Clean Sky 2 Joint Undertaking (JU) under grant agreement No 101008257.

References

[1] W. L. Oberkampf, C. J. Roy, Verification and Validation in Scientific Computing, Cambridge University Press, Cambridge, 2010. doi:10.1017/CBO9780511760396.

[2] E. T. Jaynes, G. L. Bretthorst, Probability Theory: The Logic of Science, Cambridge University Press, Cambridge, UK; New York, NY, 2003.

[3] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464. doi:10.1111/1467-9868.00294.

[4] W. Edeling, P. Cinnella, R. Dwight, H. Bijl, Bayesian estimates of parameter variability in the k–ϵ turbulence model, Journal of Computational Physics 258 (2014) 73–94. doi:10.1016/j.jcp.2013.10.027.

[5] H. Xiao, P. Cinnella, Quantification of model uncertainty in RANS simulations: A review, Progress in Aerospace Sciences 108 (2019) 1–31. doi:10.1016/j.paerosci.2018.10.001.

[6] B. Nadiga, C. Jiang, D. Livescu, Leveraging Bayesian analysis to improve accuracy of approximate models, Journal of Computational Physics 394 (2019) 280–297. doi:10.1016/j.jcp.2019.05.015.

[7] O. A. Doronina, S. M. Murman, P. E. Hamlington, Parameter Estimation for RANS Models Using Approximate Bayesian Computation, arXiv:2011.01231 [physics] (Nov. 2020). arXiv:2011.01231.

[8] G. Damblin, M. Keller, P. Barbillon, A. Pasanisi, É. Parent, Bayesian Model Selection for the Validation of Computer Codes, Quality and Reliability Engineering International 32 (6) (2016) 2043–2054. doi:10.1002/qre.2036.

[9] A. del Val, O. P. Le Maître, T. E. Magin, O. Chazot, P. M. Congedo, A surrogate-based optimal likelihood function for the bayesian calibration of catalytic recombination in atmospheric entry protection materials, Applied Mathematical Modelling 101 (2022) 791–810. doi:https://doi.org/10.1016/j.apm.2021.07.019.

[10] T. Gally, P. Groche, F. Hoppe, A. Kuttich, A. Matei, M. E. Pfetsch, M. Rakowitsch, S. Ulbrich, Identification of Model Uncertainty via Optimal Design of Experiments applied to a Mechanical Press, arXiv:1910.08408 [cs, math, stat] (Feb. 2020). arXiv:1910.08408.

[11] K. A. Maupin, L. P. Swiler, Model Discrepancy Calibration Across Experimental Settings, Reliability Engineering & System Safety (2020) 106818doi:10.1016/j.ress.2020.106818.

[12] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, R. D. Ryne, Combining Field Data and Computer Simulations for Calibration and Prediction, SIAM Journal on Scientific Computing 26 (2) (2004) 448–466. doi:10.1137/S1064827503426693.

[13] D. Higdon, J. Gattiker, B. Williams, M. Rightley, Computer Model Calibration Using High-Dimensional Output, Journal of the American Statistical Association 103 (482) (2008) 570–583.

[14] P. D. Arendt, D. W. Apley, W. Chen, D. Lamb, D. Gorsich, Improving Identifiability in Model Calibration Using Multiple Responses, Journal of Mechanical Design 134 (10) (Oct. 2012). doi:10.1115/1.4007573.

[15] M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, J. Tu, A Framework for Validation of Computer Models, Technometrics 49 (2) (2007) 138–154. doi:10.1198/004017007000000092.

[16] M. J. Bayarri, J. O. Berger, J. Cafeo, G. Garcia-Donato, F. Liu, J. Palomo, R. J. Parthasarathy, R. Paulo, J. Sacks, D. Walsh, Computer model validation with functional output, The Annals of Statistics 35 (5) (2007) 1874–1906. arXiv:0711.3271, doi:10.1214/009053607000000163.

[17] J. Brynjarsdóttir, A. O'Hagan, Learning about physical parameters: The importance of model discrepancy, Inverse Problems 30 (11) (2014) 114007. doi:10.1088/0266-5611/30/11/114007.

[18] Y. Ling, J. Mullins, S. Mahadevan, Selection of model discrepancy priors in Bayesian calibration, Journal of Computational Physics 276 (2014) 665–680. doi:10.1016/j.jcp.2014.08.005.

[19] P. Gardner, T. J. Rogers, C. Lord, R. J. Barthorpe, Learning model discrepancy: A Gaussian process and sampling-based approach, Mechanical Systems and Signal Processing 152 (2021) 107381. doi:10.1016/j.ymssp.2020.107381.

[20] M. Plumlee, Bayesian Calibration of Inexact Computer Models, Journal of the American Statistical Association 112 (519) (2017) 1274–1285. doi:10.1080/01621459.2016.1211016.

[21] M. Gu, L. Wang, Scaled Gaussian Stochastic Process for Computer Model Calibration and Prediction, SIAM/ASA Journal on Uncertainty Quantification 6 (4) (2018) 1555–1583. doi:10.1137/17M1159890.

[22] F. Xie, Y. Xu, Bayesian Projected Calibration of Computer Models, arXiv:1803.01231 [math, stat] (Feb. 2019). arXiv:1803.01231.

[23] G. B. Arhonditsis, D. Papantou, W. Zhang, G. Perhar, E. Massos, M. Shi, Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management, Journal of Marine Systems 73 (1-2) (2008) 8–30. doi:10.1016/j.jmarsys.2007.07.004.

[24] E. C. DeCarlo, S. Mahadevan, B. P. Smarslok, Bayesian Calibration of Aerothermal Models for Hypersonic Air Vehicles, in: 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, American Institute of Aeronautics and Astronautics, Boston, Massachusetts, 2013. doi:10.2514/6.2013-1683.

[25] R. Tuo, C. F. J. Wu, A theoretical framework for calibration in computer models: Parametrization, estimation and convergence properties, arXiv:1508.07155 [stat] (Aug. 2015). arXiv:1508.07155.

[26] R. Tuo, C. F. J. Wu, Prediction based on the Kennedy-O'Hagan calibration model: Asymptotic consistency and other properties, arXiv:1703.01326 [math, stat] (Mar. 2017). arXiv:1703.01326.

[27] Y. He, D. Xiu, Numerical strategy for model correction using physical constraints, Journal of Computational Physics 313 (2016) 617–634. doi:10.1016/j.jcp.2016.02.054.

[28] K. Sargsyan, X. Huan, H. N. Najm, Embedded Model Error Representation for Bayesian Model Calibration, International Journal for Uncertainty Quantification 9 (4) (2019). doi:10.1615/Int.J.UncertaintyQuantification.2019027384.

[29] F. Liu, M. J. Bayarri, J. O. Berger, Modularization in Bayesian analysis, with emphasis on analysis of computer models, Bayesian Analysis 4 (1) (2009) 119–150. doi:10.1214/09-BA404.

[30] D. D. Cox, J.-S. Park, C. E. Singer, A statistical method for tuning a computer code to a data base, Computational Statistics & Data Analysis 37 (1) (2001) 77–92. doi:10.1016/S0167-9473(00)00057-8.

[31] M. Carmassi, P. Barbillon, M. Keller, E. Parent, M. Chiodetti, Bayesian calibration of a numerical code for prediction, arXiv:1801.01810 [stat] (Mar. 2019). arXiv:1801.01810.

[32] M. L. Stein, Interpolation of Spatial Data, Springer Series in Statistics, Springer New York, New York, NY, 1999. doi:10.1007/978-1-4612-1494-6.

[33] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning, MIT Press, Cambridge, Mass, 2006.

[34] C. P. Robert, G. Casella, Monte Carlo Statistical Methods, 2nd Edition, Springer Texts in Statistics, Springer, New York, NY, 2010.

[35] J. Yuan, S. H. Ng, A sequential approach for stochastic computer model calibration and prediction, Reliability Engineering & System Safety 111 (2013) 273–286. doi:10.1016/j.ress.2012.11.004.

[36] J. Yuan, V. Nian, B. Su, Q. Meng, A simultaneous calibration and parameter ranking method for building energy models, Applied Energy 206 (2017) 657–666. doi:10.1016/j.apenergy.2017.08.220.

[37] J. Yuan, S. H. Ng, An Integrated Method for Simultaneous Calibration and Parameter Selection in Computer Models, ACM Transactions on Modeling and Computer Simulation 30 (1) (2020) 7:1–7:23. doi:10.1145/3364217.

[38] R. B. Gramacy, D. Bingham, J. P. Holloway, M. J. Grosskopf, C. C. Kuranz, E. Rutter, M. Trantham, R. P. Drake, Calibrating a large computer experiment simulating radiative shock hydrodynamics, The Annals of Applied Statistics 9 (3) (2015) 1141–1168. arXiv:1410.3293, doi:10.1214/15-AOAS850.

[39] X. Wu, T. Kozlowski, H. Meidani, K. Shirvan, Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian Process, Part 2: Application to TRACE, Nuclear Engineering and Design 335 (2018) 417–431. doi:10.1016/j.nucengdes.2018.06.003.

[40] V. R. Joseph, S. N. Melkote, Statistical Adjustments to Engineering Models, Journal of Quality Technology 41 (4) (2009) 362–375. doi:10.1080/00224065.2009.11917791.

[41] R. K. W. Wong, C. B. Storlie, T. C. M. Lee, A frequentist approach to computer model calibration, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79 (2) (2017) 635–648. doi:10.1111/rssb.12182.

[42] K. Rumsey, G. Huerta, J. Brown, L. Hund, Dealing with Measurement Uncertainties as Nuisance Parameters in Bayesian Model Calibration, SIAM/ASA Journal on Uncertainty Quantification (2020) 1287–1309doi:10.1137/19M1283707.

[43] J. L. Loeppky, D. Bingham, W. J. Welch, Computer Model Calibration or Tuning in Practice, Tech. rep. (2006).

[44] P. D. Arendt, D. W. Apley, W. Chen, Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability, Journal of Mechanical Design 134 (10) (2012) 100908. doi:10.1115/1.4007390.

[45] M. C. Kennedy, A. O'Hagan, Supplementary details on Bayesian Calibration of Computer Models (2001).

[46] C. Berge, Topological Spaces: Including a Treatment of Multi-Valued Functions, Vector Spaces, and Convexity, Macmillan Co., New York, 1963.

[47] G. Tian, J. Zhou, The maximum Theorem and the existence of Nash equilibrium of (generalized) games without lower semicontinuities, Journal of Mathematical Analysis and Applications 166 (2) (1992) 351–364. doi:10.1016/0022-247X(92)90302-T.

[48] C. Andrieu, J. Thoms, A tutorial on adaptive MCMC, Statistics and Computing 18 (4) (2008) 343–373. doi:10.1007/s11222-008-9110-y.

[49] D. Gamerman, H. F. Lopes, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd Edition, no. 68 in Texts in Statistical Science Series, Taylor & Francis, Boca Raton, 2006.

[50] R. Kommajosyula, Development and assessment of a physics-based model for subcooled flow boiling with application to CFD, Ph.D. thesis, Massachusetts Institute of Technology (2020).

[51] W. E. Kennel, Local boiling of water and superheating of high pressure steam in annuli, Thesis, Massachusetts Institute of Technology (1949).

[52] N. Leoni, Bayesian inference of model error for the calibration of two-phase CFD codes, Ph.D. thesis, Institut Polytechnique de Paris (2022).

[53] A. Mchutchon, C. Rasmussen, Gaussian Process Training with Input Noise, in: Advances in Neural Information Processing Systems, Vol. 24, Curran Associates, Inc., 2011.