



HAL
open science

Bayesian calibration with adaptive model discrepancy

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

► **To cite this version:**

Nicolas Leoni, Olivier Le Maitre, Maria Giovanna Rodio, Pietro Marco Congedo. Bayesian calibration with adaptive model discrepancy. [Research Report] Inria Saclay - Île de France; CEA - Commissariat à l'énergie atomique et aux énergies alternatives; CNRS. 2022. hal-04323574v1

HAL Id: hal-04323574

<https://inria.hal.science/hal-04323574v1>

Submitted on 24 Oct 2022 (v1), last revised 5 Dec 2023 (v2)

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

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

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)

8 and their cost, usually lower than full-scale experiments. The accuracy and reli-
9 ability of computer code predictions are consequently of paramount importance
10 today [1].

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

19 All models are imperfect representations of reality; they involve *model er-*
20 *ror*, or *model discrepancy*, affecting the quality of predictions. The pioneering
21 article of Kennedy and O’Hagan (KOH) [3] was the first to introduce model
22 discrepancy in the calibration process. Their Bayesian framework represents
23 the model discrepancy with Gaussian processes. It has been widely applied in
24 various fields such as aerodynamics [4, 5, 6, 7], fluid mechanics [8, 9] and solid
25 mechanics ([10, 11]). Compared to standard calibration, inferring the model
26 discrepancy jointly with the parameters yields more objective and less informa-
27 tive parameter posteriors. The work of Kennedy and O’Hagan [3] has triggered
28 a whole body of calibration literature dealing with computer models with high-
29 dimensional outputs [12, 13, 14] and with computer model validation [15, 16].
30 The nature of model discrepancy and its treatment in the calibration process
31 was also extensively discussed in [17, 18, 19]. For instance, different calibration
32 methods with model discrepancy were recently proposed in [20], [21] and [22].

33 Although the calibration with model discrepancy has encountered tremen-
34 dous successes, multiple questions remain unanswered. One question concerns
35 the choice of representation of the model error. A Gaussian process is added
36 to the model output in KOH to represent the discrepancy. Other works use
37 constant bias, random walks [23], or deterministic functions [24]. For Gaussian
38 process models, the user must provide the prior mean, the covariance function
39 structure, and the prior of its hyperparameter. Setting these quantities is not
40 an obvious task and may significantly impact the calibration results. A frequen-
41 tist approach can avoid this issue [25, 26]. Another approach to calibration in
42 the presence of model error embeds representations to add uncertainties in the
43 model parameters [27, 28].

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

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

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

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

88 2. Bayesian calibration of computer codes

89 2.1. General framework and reference solution

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

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

100 Let f be a computer model that provides predictions of y . The inputs for f
 101 includes the conditions \mathbf{x} and other model parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. The latter
 102 are usually imperfectly known, and the goal of the calibration procedure is to
 103 learn their value from the observation \mathbf{y}_{obs} . In the example, $\boldsymbol{\theta}$ might consists of
 104 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

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

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

As discussed below, a Bayesian procedure can be employed to learn the pa-
 rameters $\boldsymbol{\theta}$ 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 pre-
 dictions and observations) that do not follow the assumed measurement error
 distribution. Such inconsistencies may have two causes: an incorrect measure-
 ment 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 un-
 biased 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 irre-
 ducible 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:

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

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

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

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

150 Despite its apparent simplicity, this calibration can not be solved precisely
 151 in most applications. The reason is that it requires an accurate estimation of
 152 the posterior density, which might exhibit substantial variations and many local
 153 maxima over the joint space $\Theta \times \Psi$. Further, for calibrations involving multi-
 154 ple quantities and experimental configurations, the number of hyperparameters
 155 might become significant, making the estimation in high dimensions very com-
 156 plex. Monte-Carlo techniques [34] and other advanced sampling techniques have
 157 been employed to generate samples from the high dimensional posterior distri-
 158 butions and approximate the required integrals. The following section discusses
 159 some works proposed in the literature on direct full Bayesian calibration. How-
 160 ever, the so-called modular approaches ([29]) have gained popularity as alter-
 161 natives to full Bayesian calibration. These approaches, discussed in section 2.3,
 162 reduce the dimensionality of the sample space by estimating point values for
 163 some hyperparameters.

164 2.2. Fully Bayesian approaches

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

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

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

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

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

209 *2.3. Modular approaches*

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

220 Since the calibration is sequential, the calibration problem for the modules
 221 must be adapted, and, in particular, assumptions regarding subsequent modules
 222 not yet calibrated are needed. Since the posterior distribution of the parameters
 223 $\boldsymbol{\theta}$ is arguably the goal of the calibration, the last module estimates $p(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \boldsymbol{\psi} =$
 224 $\hat{\boldsymbol{\psi}})$ where the estimator $\hat{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$ have been estimated in previous modules such
 225 that $p(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}) \approx p_{\text{Bayes}}(\boldsymbol{\theta})$. We now distinguish between two classes of
 226 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:

$$\hat{\boldsymbol{\psi}}_1 = \arg \max_{\boldsymbol{\psi}} p(\mathbf{y}_{\text{obs}}|\boldsymbol{\psi}, \boldsymbol{\theta}_0). \quad (9)$$

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

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

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:

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

and the following estimation for the parameters posterior distribution:

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

237 Note also that an approximation of $p(\mathbf{y}_{\text{obs}}|\boldsymbol{\psi})$ is sometimes used to reduce the
 238 computational cost; Kennedy and O’Hagan propose to use a Gaussian approx-
 239 imation in [45]. Compared to the first class of methods, the computation of
 240 $p(\mathbf{y}_{\text{obs}}|\boldsymbol{\psi})$ is generally more costly than estimating a single parameter value
 241 $\boldsymbol{\theta}_0$. Compared to the first class, the second class of methods benefits from

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

244 The two classes of modular methods are compared in [19], where the second
 245 class estimates higher model discrepancies, showing the benefit of including the
 246 parameter uncertainty. In general, the modular approaches provide fairly accu-
 247 rate parameter posteriors. Nonetheless, due to the single point estimate for ψ ,
 248 they do not account for the complete uncertainty about the model discrepancy
 249 term. In what follows, we propose a calibration technique where the uncertainty
 250 about the values of ψ is estimated to include all relevant uncertainties in the
 251 calibration framework.

252 3. The Full Maximum A Posteriori method

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

260 3.1. Calibration framework

261 3.1.1. Fundamental hypotheses

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

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

Therefore, the calibration equation becomes:

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

This equation appeared previously in the work of [20], but we do not modify the
 covariance of z_{θ} as they do. Instead, we perform a joint estimation of θ and ψ
 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

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

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

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

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

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

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

$$\int_{\Theta} g(\boldsymbol{\theta}, \boldsymbol{\theta}') p(\boldsymbol{\theta}' | \mathbf{y}_{\text{obs}}) d\boldsymbol{\theta}' \propto g(\boldsymbol{\theta}, \boldsymbol{\theta}). \quad (16)$$

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

273 The second assumption supposes that $g(\boldsymbol{\theta}, \cdot)$ has a similar shape for differ-
 274 ent $\boldsymbol{\theta}$ and that the ratio between its posterior-averaged value and its maximal
 275 value (remark that $\boldsymbol{\theta} = \arg \max_{\boldsymbol{\theta}' \in \Theta} g(\boldsymbol{\theta}, \boldsymbol{\theta}')$) does not depend on $\boldsymbol{\theta}$. This is
 276 verified for example when $\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta})$ is a slowly varying function of $\boldsymbol{\theta}$ (one value
 277 of optimal hyperparameters stands out), or when, for two different values of
 278 optimal hyperparameters $\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}_1)$ and $\hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}_2)$, the conditional densities
 279 $p(\boldsymbol{\theta} | \mathbf{y}_{\text{obs}}, \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}_1))$ and $p(\boldsymbol{\theta} | \mathbf{y}_{\text{obs}}, \hat{\boldsymbol{\psi}}_{\text{FMP}}(\boldsymbol{\theta}_2))$ are close.

280 3.1.2. Posterior approximation and predictive density

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

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

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

283 Note that for most covariance functions, $c_{\boldsymbol{\psi}}(x, x')$ is a continuous function
 284 of $\boldsymbol{\psi}$ for all (x, x') . By the Maximum Theorem [46, 47] it can be shown that
 285 p_{FMP} is a continuous function of $\boldsymbol{\theta}$. With the additional assumption that for
 286 each value of $\boldsymbol{\theta}$ there is a unique value of optimal hyperparameters (which is

287 not true in general), we have that $\hat{\psi}_{\text{FMP}}$ is a continuous function of $\boldsymbol{\theta}$. These
 288 continuity properties pave the way for calibration techniques based on surrogate
 289 modelling of the FMP quantities that will be explored in subsequent work.

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

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

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

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

and

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

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

291 Equation (20) provides a decomposition of the predictive error into two contribu-
 292 tions. The first contribution, named "uncertainty in the corrected model",
 293 is the variability of the corrected predictions. It is significant when the corrected
 294 predictions vary with $\boldsymbol{\theta}$, indicating that the model discrepancy term cannot ade-
 295 quately correct the predictions. The second term, named "residual uncertainty,"
 296 is the part of the posterior uncertainty that the corrected model cannot explain.
 297 It corresponds to the remaining uncertainty in calibration using a finite amount
 298 of noisy observations.

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

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

308 3.2. Marginal likelihood of the FMP approximation

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

Assume that, in an alternative approximation method, the functional rela-
 tionship $\hat{\boldsymbol{\psi}}_h(\boldsymbol{\theta}) = h(\boldsymbol{\theta})$ is proposed, with h a generic function. The posterior of
 the model parameters would become $p_h(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) \propto p(\boldsymbol{\theta})p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{\psi} = \boldsymbol{\psi}_h(\boldsymbol{\theta}))$.
 This setting also encompasses single-point estimation methods, such as the KOH
 or cross-validation estimator of the hyperparameters. We have:

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

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

322 3.3. Elementary examples

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

328 A Gaussian posterior is unlikely in practice because priors on hyperparam-
 329 eters are rarely Gaussian. However, the Gaussian case is very instructive, and
 330 posterior approximations by Gaussian distribution are standard when the pos-
 331 teriors become increasingly sharp around their maximum as the number of ob-
 332 servations increases. Similarly, the Gaussian mixture can approximate more
 333 complex distributions, and this example helps us understand the case of a mul-
 334 timodal posterior distribution where alternative explanations of the observations
 335 are plausible.

336 *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 $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\psi})$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_\theta, \boldsymbol{\mu}_\psi)$ is the posterior mean and $\boldsymbol{\Lambda} = \begin{pmatrix} \mathbf{V}_\theta & \mathbf{C}_{\theta,\psi}^T \\ \mathbf{C}_{\theta,\psi} & \mathbf{V}_\psi \end{pmatrix}$ is the posterior covariance, symmetric and positive definite. The joint posterior distribution writes:

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

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

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

which directly gives:

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

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 $\boldsymbol{\psi}$:

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

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

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

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

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

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

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

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

339 This order can be understood as the following: let $\{\lambda_i\}_{1 \leq i \leq p}$ be the eigenvalues
 340 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
 341 decreasing order. Then, for all $i \leq p$, we have $\lambda_i \leq \lambda'_i$. It is seen that p_{KOH}
 342 remains Gaussian, with correct mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$, but a reduced covariance compared
 343 to p_{Bayes} .

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

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

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

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

we obtain

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

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

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

353 3.3.2. Mixture of Gaussians

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

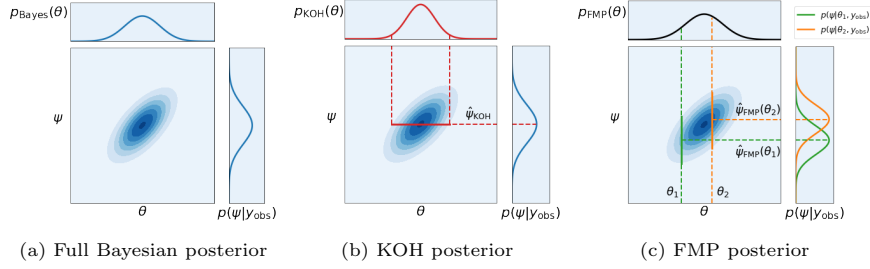


Figure 1: Joint posteriors of parameters and hyperparameters: Gaussian case.

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

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

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 $p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}_{\text{obs}}) \approx \pi_i p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}_{\text{obs}}, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)$. By linearity, the true marginal density of the parameters is the linear combination of marginals:

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

The KOH estimation first computes the hyperparameter marginal distribution:

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

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

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

So, the KOH posterior is:

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

with

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

354 The solution of the KOH estimation is Gaussian with reduced variance matrix.
355 Besides, the selection of the mode is driven by the criteria defining i_{KOH} , which
356 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:

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

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

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

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

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

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

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

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

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

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

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

with the normalizing constant

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

365 The distribution p_{FMP} is a linear combination of the m marginals associated
 366 with the original Gaussian mixture. The variances of the different peaks are
 367 correctly estimated, but the estimation introduces a bias on the weights of
 368 the Gaussian mixture: the FMP estimates the original weights π_i to be $\pi'_i =$
 369 $\frac{1}{k} \pi_i \frac{\sqrt{|\mathbf{V}_{i,\boldsymbol{\theta}}|}}{\sqrt{|\boldsymbol{\Lambda}_i|}}$ where k is a normalizing constant such that $\sum_{i=1}^m \pi'_i = 1$. Specifi-
 370 cally, the FMP method emphasizes the importance of the Gaussians with low
 371 hyperparameter variance in the mixture.

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

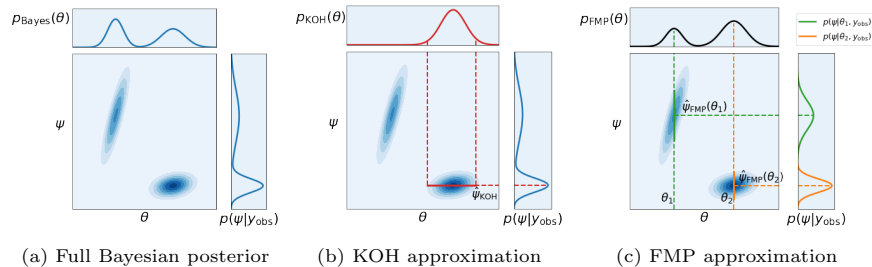


Figure 2: Joint posterior of parameters and hyperparameters: well separated Gaussian mixture case.

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

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

392 Figure 3b shows the case where the Θ -projections of the Gaussians overlap.
 393 This situation is challenging for the two approximation methods since they rely
 394 on a point-mass approximation of the marginals of ψ , which is not verified here.
 395 In the example shown, the FMP approximation performs slightly better with
 396 a posterior variance of θ underestimated but closer to the reference than for
 397 KOH. This situation occurs when the distribution of the model discrepancy is
 398 at odds with the observations, and no hyperparameters value stands out; it calls
 399 for a change in the prior of z_θ (e.g., selecting another covariance structure).

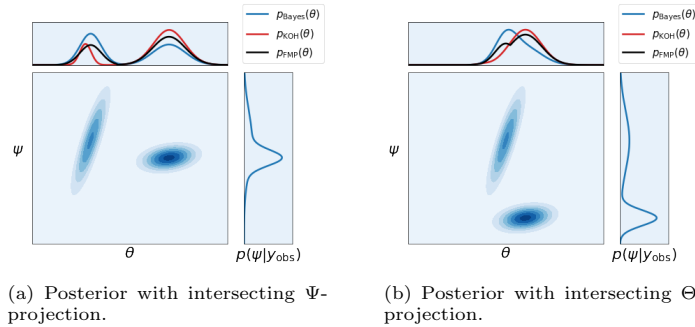


Figure 3: Comparison of KOH and FMP estimations: case of Gaussian mixtures without modes separation.

400 4. Applications

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

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

413 *4.1. Calibration of an inadequate model*

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

420 *4.1.1. Problem formulation*

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

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

421 The input variable x and the model parameter θ lie within the restricted range
 422 $(x, \theta) \in [0, 1] \times [-0.5, 1.5]$. We use observations of the true function at 8 points
 423 uniformly spaced in the interval $[0, 1]$. The observations are noisy with a centered
 424 Gaussian noise with standard deviation $\sigma_\epsilon = 0.1$. The observation noise
 425 is not known a priori but is learned along with the other hyperparameters of
 426 the model error.

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

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

427 Thus, a total of three hyperparameters $\psi = (\sigma, l, \sigma_\epsilon)$ are involved in the cali-
 428 bration. The prior for θ is uniform over $[-0.5, 1.5]$. The priors for σ and σ_ϵ are
 429 uniform on the interval $[1e-5, 1]$. The prior on l is also uniform, truncated on
 430 $[1e-3, 5]$.

431 Samples of the posterior distributions are obtained using Monte-Carlo Markov
 432 Chain methods with the Metropolis-Hastings algorithm [48]. For the KOH cali-
 433 bration, we first estimate the hyperparameters using (11). The MCMC method
 434 is then run for the target density $p_{\text{KOH}}(\theta)$ (see (12)). In the FMP calibration,
 435 the target density $p_{\text{FMP}}(\theta)$ is also of dimension 1, but each step of the MCMC
 436 requires the determination of the optimal hyperparameters. The cost of these
 437 optimizations is not prohibitive, thanks to the available gradients [33, Chapter
 438 5], and because the optimal hyperparameters of the previous step are generally
 439 a good starting point. For the Bayes calibration, the target density $p(\theta, \psi | \mathbf{y}_{\text{obs}})$
 440 is in dimension four and requires significantly more steps. The FMP and KOH
 441 chains have $5e5$ steps, and the Bayes chain has $2.5e6$ steps, all with a burn-
 442 in of 10%. From each chain, we extract a posterior sample of size 5000 by
 443 taking regularly-spaced visited states. This subsampling rate is large enough
 444 compared to the self-correlation length of the chains to consider the resulting
 445 samples independent. For more details about this procedure, see [49].

446 *4.1.2. Calibration Results*

447 We apply a Kernel Density Estimation method to the MCMC samples to
 448 estimate the posterior marginal densities of parameters and hyperparameters.
 449 These marginals are shown in figure 4. The parameter’s posterior marginals are
 450 bimodal with a first mode located at $\theta = -0.025$ and the second at $\theta = 1.02$.
 451 For the reference Bayes solutions, the relative importance of the two modes is
 452 in the ratio 4.3:1, computed as the relative frequency of $\theta < 0.5$. The ratio is
 453 2.75:1 for the FMP method and 69.4:1 for the KOH method. The FMP method
 454 overestimates the second mode’s importance and puts more probability mass
 455 between the two modes. The KOH estimation almost misses the second mode
 456 focusing on a single interpretation of the observations.

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

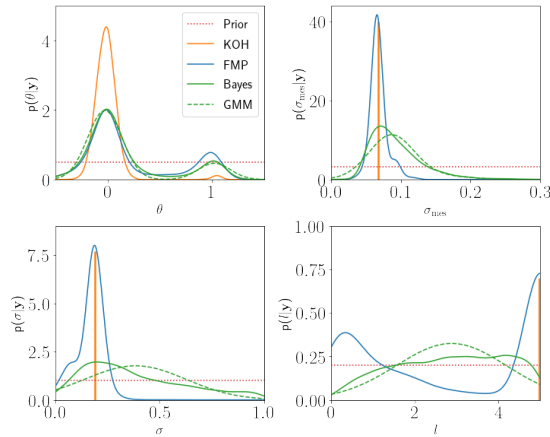


Figure 4: Prior and posterior distributions for the model parameter (top left) and hyperparameters. In the KOH calibration, hyperparameters are estimated with point mass distributions, represented with vertical lines. The Gaussian Mixture Model, fitted on the Bayes posterior, is represented with the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

469 The bimodality of the parameter posterior reflects in the posterior model

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

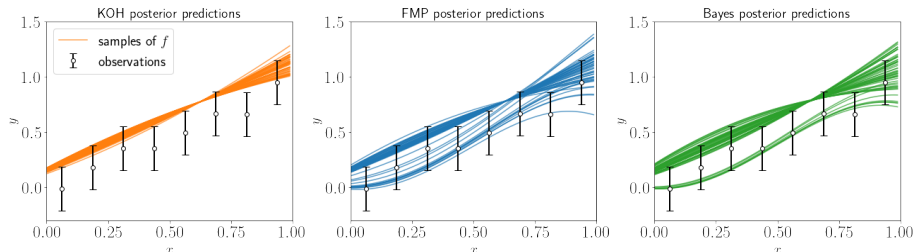


Figure 5: Posterior samples of the model predictions for all three calibration methods. 50 samples are represented on each figure, and observations are shown with $2\sigma_\epsilon$ confidence intervals. Model predictions that are quasi-linear correspond to $\theta \approx 0$, and the ones with slow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

477 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-Maximization¹. The selected form of the mixture is a weighted sum of two Gaussians

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

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

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

- 486 • The KOH estimation selects a mode according to the criterion in equa-
 487 tion (30). We check that $\frac{\pi_1}{\sqrt{|\mathbf{C}_1^{[3,3]}|}} = 70.8$ is greater than $\frac{\pi_2}{\sqrt{|\mathbf{C}_2^{[3,3]}|}} = 10.0$,
 488 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 $\boldsymbol{\mu} = (\theta, \sigma_{\text{mes}}, \Delta T, \sigma, l)$.

- 489 • The posterior correlation between θ and $\boldsymbol{\psi}$ is low, so the variance at
490 the first mode is roughly unchanged after projection: $\mathbf{V}_\theta = \mathbf{C}_1(0, 0) =$
491 $2.90\text{e-}2$, and $\mathbf{V}_{\theta|\boldsymbol{\psi}} = \mathbf{C}_1(0, 0) - \mathbf{C}_{1,\theta,\boldsymbol{\psi}}^T (\mathbf{C}_1^{[3,3]})^{-1} \mathbf{C}_{1,\theta,\boldsymbol{\psi}} = 2.89\text{e-}2$, with
492 $\mathbf{C}_{1,\theta,\boldsymbol{\psi}} = (-1.9\text{e-}4, 1.3\text{e}3, -1.6\text{e-}3)^T$. The KOH method underestimates
493 the posterior variance due to selecting only one mode.
- 494 • The FMP weights can be computed using (32). We have $\pi'_i \propto \pi_i \frac{\sqrt{|\mathbf{C}_i|}}{\sqrt{\mathbf{C}_i(0,0)}}$,
495 which gives $\pi'_1 = 0.73$ and $\pi'_2 = 0.27$, in a proportion 2.71:1 which corre-
496 sponds to the ratio found in the FMP results.

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

501 4.2. Calibration of a boiling model

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

506 4.2.1. Problem formulation

507 The MIT Boiling model [50] simulates the boiling of fluids in contact with
508 a heated wall. The inputs of the model are experimental variables (fluid pres-
509 sure, velocity, temperature, and configuration geometry) and the wall superheat
510 ΔT_{sup} , defined as the difference between the wall temperature and the satura-
511 tion temperature of the fluid. The model output is the heat flux ϕ from the

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

519 Three calibration parameters $(\theta_1, \theta_2, \theta_3)$ are considered; the motivations for
 520 this choice and the significance of each parameter is discussed in [52]. The mea-
 521 surement uncertainty lies primarily on ΔT_{sup} , which is an input model quantity,
 522 not an output. This specificity requires an adaptation of the framework. Follow-
 523 ing [53], the sensitivity (derivatives) of \mathbf{y}_{obs} at the uncertain inputs is estimated
 524 from a polynomial fit over the observations and used to include this uncertainty
 525 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:

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

526 The vector of observations has a normal distribution, with covariance matrix
 527 $\Sigma_{\psi} + \sigma_{\text{mes}, \Delta T}^2 \mathbf{D}$, where \mathbf{D} is the diagonal matrix with diagonal coefficients equal
 528 to the derivatives of \mathbf{y}_{obs} at the observations.

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

534 As in the first example, we sample the posteriors with the Metropolis-
 535 Hastings algorithm. The KOH and FMP chains have an acceptance rate of
 536 $\approx 11\%$ and the Bayes chain $\approx 3\%$. Figure 6 shows the self-correlation of the
 537 MCMC chains. The mixing lengths of the chains are $\tau = 28$ for KOH, $\tau = 30$
 538 for FMP, and $\tau = 470$ for Bayes. To ensure sample sets of similar quality, we
 539 generate chains with respective lengths 5e5, 5e5, and 1e7 for the KOH, FMP,
 540 and Bayes methods and extract 1,000 samples from the chains by uniform sub-
 541 sampling.

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

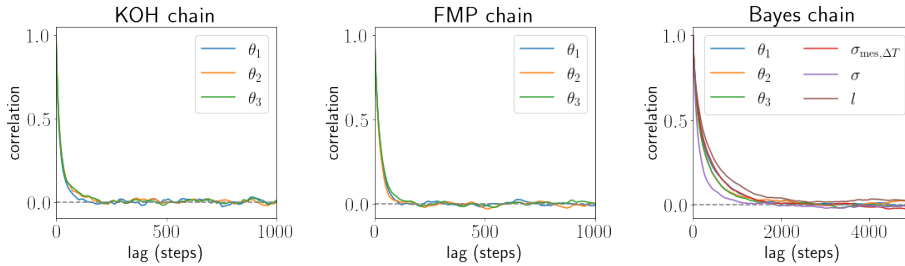


Figure 6: Self-correlation of the chains for the three methods applied to the MIT Boiling model calibration.

543 optimizations use the C++ library NLOpt², with the global algorithm MSL,
 544 using the local optimizer SBPLX. A maximal time of 10e−4s is allocated to
 545 individual optimization. For the KOH estimation of hyperparameters, the integral
 546 over Θ is computed by quadrature, using a QMC grid of size 300 over the
 547 parameters domain.

548 4.2.2. Results

549 Table 2 summarizes the results of the inference problem for the three meth-
 550 ods. The most noticeable result is that the KOH method estimates a zero
 551 model error σ and attributes all the discrepancies to measurement uncertainty.
 552 Consistently, the KOH method yields an erroneous estimation of θ_1 with an
 553 underestimated variance. On the contrary, the FMP method identifies a model
 554 error comparable to the full Bayesian calibration and, despite underestimat-
 555 ing the hyperparameters’ variances, provides a globally correct inference of the
 556 parameters and hyperparameters.

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

Table 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 $I_2 = [0, 1.5]$.

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

² <https://nlopt.readthedocs.io/en/latest/>

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

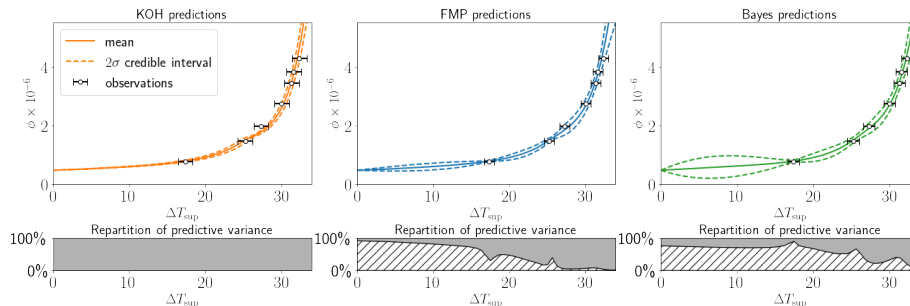


Figure 7: Posterior predictions of the corrected model ($f + z_{\theta}$) with uncertainty. Observations are shown with inferred error bars of length $2\mathbb{E}_{\theta}[\sigma_{\text{mes},\Delta T}]$. The bottom plots show the decomposition of the predictive variance in residual uncertainty (stripped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

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

584 5. Conclusions

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

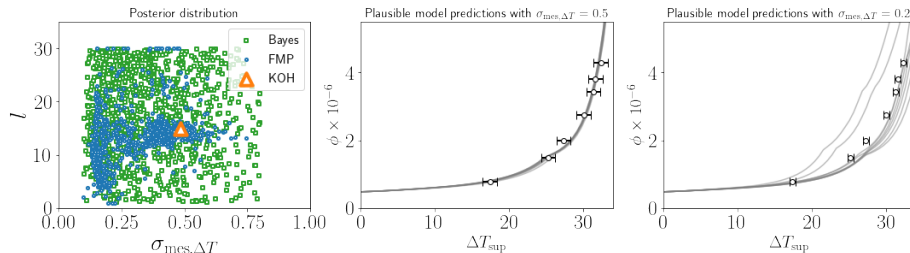


Figure 8: Samples in the σ_ϵ and l plane for the three methods (left plot). Samples of the FMP model predictions conditioned on a zero (middle) and non-zero model error (right).

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

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

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

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

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

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

627 SUPPLEMENTARY MATERIAL

628 The implementation of the MIT Boiling model, as well as the Kennel mea-
629 surements data set used in section 4.2, are available online at [https://github.](https://github.com/nleoni95/MITB)
630 [com/nleoni95/MITB](https://github.com/nleoni95/MITB).

631 Acknowledgements

632 This work was partly funded by the French Innovation Agency for De-
633 fense (Agence de l’Innovation de Défense, AID). N. Leoni would also like to
634 thank Ravikishore Kommajosyula for providing the boiling model, and Kathryn
635 Maupin for helpful discussions.

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

638 References

- 639 [1] W. L. Oberkampf, C. J. Roy, *Verification and Validation in Scientific*
640 *Computing*, Cambridge University Press, Cambridge, 2010. doi:10.1017/
641 CB09780511760396.
- 642 [2] E. T. Jaynes, G. L. Bretthorst, *Probability Theory: The Logic of Science*,
643 Cambridge University Press, Cambridge, UK; New York, NY, 2003.
- 644 [3] M. C. Kennedy, A. O’Hagan, Bayesian calibration of computer models,
645 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*
646 63 (3) (2001) 425–464. doi:10.1111/1467-9868.00294.
- 647 [4] W. Edeling, P. Cinnella, R. Dwight, H. Bijl, Bayesian estimates of pa-
648 rameter variability in the $k-\epsilon$ turbulence model, *Journal of Computational*
649 *Physics* 258 (2014) 73–94. doi:10.1016/j.jcp.2013.10.027.
- 650 [5] H. Xiao, P. Cinnella, Quantification of model uncertainty in RANS sim-
651 ulations: A review, *Progress in Aerospace Sciences* 108 (2019) 1–31.
652 doi:10.1016/j.paerosci.2018.10.001.
- 653 [6] B. Nadiga, C. Jiang, D. Livescu, Leveraging Bayesian analysis to improve
654 accuracy of approximate models, *Journal of Computational Physics* 394
655 (2019) 280–297. doi:10.1016/j.jcp.2019.05.015.

- 656 [7] O. A. Doronina, S. M. Murman, P. E. Hamlington, Parameter Esti-
657 mation for RANS Models Using Approximate Bayesian Computation,
658 arXiv:2011.01231 [physics] (Nov. 2020). arXiv:2011.01231.
- 659 [8] G. Damblin, M. Keller, P. Barbillon, A. Pasanisi, É. Parent, Bayesian
660 Model Selection for the Validation of Computer Codes, Quality and Re-
661 liability Engineering International 32 (6) (2016) 2043–2054. doi:10.1002/
662 qre.2036.
- 663 [9] A. del Val, O. P. Le Maître, T. E. Magin, O. Chazot, P. M. Congedo, A
664 surrogate-based optimal likelihood function for the bayesian calibration of
665 catalytic recombination in atmospheric entry protection materials, Applied
666 Mathematical Modelling 101 (2022) 791–810. doi:https://doi.org/10.
667 1016/j.apm.2021.07.019.
- 668 [10] T. Gally, P. Groche, F. Hoppe, A. Kuttich, A. Matei, M. E. Pfetsch,
669 M. Rakowitsch, S. Ulbrich, Identification of Model Uncertainty via Optimal
670 Design of Experiments applied to a Mechanical Press, arXiv:1910.08408 [cs,
671 math, stat] (Feb. 2020). arXiv:1910.08408.
- 672 [11] K. A. Maupin, L. P. Swiler, Model Discrepancy Calibration Across
673 Experimental Settings, Reliability Engineering & System Safety (2020)
674 106818doi:10.1016/j.res.2020.106818.
- 675 [12] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, R. D. Ryne, Com-
676 bining Field Data and Computer Simulations for Calibration and Pre-
677 diction, SIAM Journal on Scientific Computing 26 (2) (2004) 448–466.
678 doi:10.1137/S1064827503426693.
- 679 [13] D. Higdon, J. Gattiker, B. Williams, M. Rightley, Computer Model Calibra-
680 tion Using High-Dimensional Output, Journal of the American Statistical
681 Association 103 (482) (2008) 570–583.
- 682 [14] P. D. Arendt, D. W. Apley, W. Chen, D. Lamb, D. Gorsich, Improving
683 Identifiability in Model Calibration Using Multiple Responses, Journal of
684 Mechanical Design 134 (10) (Oct. 2012). doi:10.1115/1.4007573.
- 685 [15] M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish,
686 C.-H. Lin, J. Tu, A Framework for Validation of Computer Models, Tech-
687 nometrics 49 (2) (2007) 138–154. doi:10.1198/004017007000000092.
- 688 [16] M. J. Bayarri, J. O. Berger, J. Cafeo, G. Garcia-Donato, F. Liu, J. Palomo,
689 R. J. Parthasarathy, R. Paulo, J. Sacks, D. Walsh, Computer model vali-
690 dation with functional output, The Annals of Statistics 35 (5) (2007) 1874–
691 1906. arXiv:0711.3271, doi:10.1214/009053607000000163.
- 692 [17] J. Brynjarsdóttir, A. O’Hagan, Learning about physical parameters: The
693 importance of model discrepancy, Inverse Problems 30 (11) (2014) 114007.
694 doi:10.1088/0266-5611/30/11/114007.

- 695 [18] Y. Ling, J. Mullins, S. Mahadevan, Selection of model discrepancy priors
696 in Bayesian calibration, *Journal of Computational Physics* 276 (2014) 665–
697 680. doi:10.1016/j.jcp.2014.08.005.
- 698 [19] P. Gardner, T. J. Rogers, C. Lord, R. J. Barthorpe, Learning model dis-
699 crepancy: A Gaussian process and sampling-based approach, *Mechanical*
700 *Systems and Signal Processing* 152 (2021) 107381. doi:10.1016/j.ymsp.
701 2020.107381.
- 702 [20] M. Plumlee, Bayesian Calibration of Inexact Computer Models, *Journal*
703 *of the American Statistical Association* 112 (519) (2017) 1274–1285. doi:
704 10.1080/01621459.2016.1211016.
- 705 [21] M. Gu, L. Wang, Scaled Gaussian Stochastic Process for Computer Model
706 Calibration and Prediction, *SIAM/ASA Journal on Uncertainty Quantifi-*
707 *cation* 6 (4) (2018) 1555–1583. doi:10.1137/17M1159890.
- 708 [22] F. Xie, Y. Xu, Bayesian Projected Calibration of Computer Models,
709 arXiv:1803.01231 [math, stat] (Feb. 2019). arXiv:1803.01231.
- 710 [23] G. B. Arhonditsis, D. Papantou, W. Zhang, G. Perhar, E. Massos, M. Shi,
711 Bayesian calibration of mechanistic aquatic biogeochemical models and
712 benefits for environmental management, *Journal of Marine Systems* 73 (1-
713 2) (2008) 8–30. doi:10.1016/j.jmarsys.2007.07.004.
- 714 [24] E. C. DeCarlo, S. Mahadevan, B. P. Smarslok, Bayesian Calibration
715 of Aerothermal Models for Hypersonic Air Vehicles, in: 54th
716 AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and
717 Materials Conference, American Institute of Aeronautics and Astronautics,
718 Boston, Massachusetts, 2013. doi:10.2514/6.2013-1683.
- 719 [25] R. Tuo, C. F. J. Wu, A theoretical framework for calibration in com-
720 puter models: Parametrization, estimation and convergence properties,
721 arXiv:1508.07155 [stat] (Aug. 2015). arXiv:1508.07155.
- 722 [26] R. Tuo, C. F. J. Wu, Prediction based on the Kennedy-O’Hagan calibra-
723 tion model: Asymptotic consistency and other properties, arXiv:1703.01326
724 [math, stat] (Mar. 2017). arXiv:1703.01326.
- 725 [27] Y. He, D. Xiu, Numerical strategy for model correction using physical
726 constraints, *Journal of Computational Physics* 313 (2016) 617–634. doi:
727 10.1016/j.jcp.2016.02.054.
- 728 [28] K. Sargsyan, X. Huan, H. N. Najm, Embedded Model Error Rep-
729 resentation for Bayesian Model Calibration, *International Journal*
730 *for Uncertainty Quantification* 9 (4) (2019). doi:10.1615/Int.J.
731 *UncertaintyQuantification.2019027384.*

- 732 [29] F. Liu, M. J. Bayarri, J. O. Berger, Modularization in Bayesian analysis,
733 with emphasis on analysis of computer models, *Bayesian Analysis* 4 (1)
734 (2009) 119–150. doi:10.1214/09-BA404.
- 735 [30] D. D. Cox, J.-S. Park, C. E. Singer, A statistical method for tuning a
736 computer code to a data base, *Computational Statistics & Data Analysis*
737 37 (1) (2001) 77–92. doi:10.1016/S0167-9473(00)00057-8.
- 738 [31] M. Carmassi, P. Barbillon, M. Keller, E. Parent, M. Chiodetti, Bayesian
739 calibration of a numerical code for prediction, arXiv:1801.01810 [stat] (Mar.
740 2019). arXiv:1801.01810.
- 741 [32] M. L. Stein, *Interpolation of Spatial Data*, Springer Series in Statis-
742 tics, Springer New York, New York, NY, 1999. doi:10.1007/
743 978-1-4612-1494-6.
- 744 [33] C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learn-*
745 *ing*, Adaptive Computation and Machine Learning, MIT Press, Cambridge,
746 Mass, 2006.
- 747 [34] C. P. Robert, G. Casella, *Monte Carlo Statistical Methods*, 2nd Edition,
748 Springer Texts in Statistics, Springer, New York, NY, 2010.
- 749 [35] J. Yuan, S. H. Ng, A sequential approach for stochastic computer model
750 calibration and prediction, *Reliability Engineering & System Safety* 111
751 (2013) 273–286. doi:10.1016/j.ress.2012.11.004.
- 752 [36] J. Yuan, V. Nian, B. Su, Q. Meng, A simultaneous calibration and paramete-
753 rer ranking method for building energy models, *Applied Energy* 206 (2017)
754 657–666. doi:10.1016/j.apenergy.2017.08.220.
- 755 [37] J. Yuan, S. H. Ng, An Integrated Method for Simultaneous Calibration and
756 Parameter Selection in Computer Models, *ACM Transactions on Modeling*
757 *and Computer Simulation* 30 (1) (2020) 7:1–7:23. doi:10.1145/3364217.
- 758 [38] R. B. Gramacy, D. Bingham, J. P. Holloway, M. J. Grosskopf, C. C. Kur-
759 ranz, E. Rutter, M. Trantham, R. P. Drake, Calibrating a large com-
760 puter experiment simulating radiative shock hydrodynamics, *The An-*
761 *nals of Applied Statistics* 9 (3) (2015) 1141–1168. arXiv:1410.3293,
762 doi:10.1214/15-AOAS850.
- 763 [39] X. Wu, T. Kozłowski, H. Meidani, K. Shirvan, Inverse uncertainty quantifi-
764 cation using the modular Bayesian approach based on Gaussian Process,
765 Part 2: Application to TRACE, *Nuclear Engineering and Design* 335 (2018)
766 417–431. doi:10.1016/j.nucengdes.2018.06.003.
- 767 [40] V. R. Joseph, S. N. Melkote, Statistical Adjustments to Engineering Mod-
768 els, *Journal of Quality Technology* 41 (4) (2009) 362–375. doi:10.1080/
769 00224065.2009.11917791.

- 770 [41] R. K. W. Wong, C. B. Storlie, T. C. M. Lee, A frequentist approach to
771 computer model calibration, *Journal of the Royal Statistical Society: Series*
772 *B (Statistical Methodology)* 79 (2) (2017) 635–648. doi:10.1111/rssb.
773 12182.
- 774 [42] K. Rumsey, G. Huerta, J. Brown, L. Hund, Dealing with Measurement
775 Uncertainties as Nuisance Parameters in Bayesian Model Calibration,
776 *SIAM/ASA Journal on Uncertainty Quantification* (2020) 1287–1309doi:
777 10.1137/19M1283707.
- 778 [43] J. L. Loeppky, D. Bingham, W. J. Welch, Computer Model Calibration or
779 Tuning in Practice, Tech. rep. (2006).
- 780 [44] P. D. Arendt, D. W. Apley, W. Chen, Quantification of Model Uncertainty:
781 Calibration, Model Discrepancy, and Identifiability, *Journal of Mechanical*
782 *Design* 134 (10) (2012) 100908. doi:10.1115/1.4007390.
- 783 [45] M. C. Kennedy, A. O’Hagan, Supplementary details on Bayesian Calibra-
784 tion of Computer Models (2001).
- 785 [46] C. Berge, *Topological Spaces: Including a Treatment of Multi-Valued Func-*
786 *tions, Vector Spaces, and Convexity*, Macmillan Co., New York, 1963.
- 787 [47] G. Tian, J. Zhou, The maximum Theorem and the existence of Nash
788 equilibrium of (generalized) games without lower semicontinuities, *Journal*
789 *of Mathematical Analysis and Applications* 166 (2) (1992) 351–364.
790 doi:10.1016/0022-247X(92)90302-T.
- 791 [48] C. Andrieu, J. Thoms, A tutorial on adaptive MCMC, *Statistics and Com-*
792 *puting* 18 (4) (2008) 343–373. doi:10.1007/s11222-008-9110-y.
- 793 [49] D. Gamerman, H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Sim-*
794 *ulation for Bayesian Inference*, 2nd Edition, no. 68 in *Texts in Statistical*
795 *Science Series*, Taylor & Francis, Boca Raton, 2006.
- 796 [50] R. Kommajosyula, Development and assessment of a physics-based model
797 for subcooled flow boiling with application to CFD, Ph.D. thesis, Mas-
798 sachusetts Institute of Technology (2020).
- 799 [51] W. E. Kennel, Local boiling of water and superheating of high pressure
800 steam in annuli, Thesis, Massachusetts Institute of Technology (1949).
- 801 [52] N. Leoni, Bayesian inference of model error for the calibration of two-phase
802 CFD codes, Ph.D. thesis, Institut Polytechnique de Paris (2022).
- 803 [53] A. Mchutchon, C. Rasmussen, Gaussian Process Training with Input Noise,
804 in: *Advances in Neural Information Processing Systems*, Vol. 24, Curran
805 Associates, Inc., 2011.