Machine Learning for Computational Science and Engineering (2026) 2:24
https://doi.org/10.1007/s44379-026-00072-0

RESEARCH

Bayesian model updating with controlled Gaussian process predictive uncertainty

Marcos A. Valdebenito[1] · Chao Dang[1] · Peipei Li[2] · Zhouzhou Song[1] · Pengfei Wei[3] · Sylvia Keßler[4] · Matthias G. R. Faes[1,5]

Received: 27 March 2026 / Accepted: 22 May 2026

Abstract

Bayesian model updating is widely used in engineering for the identification of unknown parameters of computational models based on noisy experimental measurements. In practical applications, however, the repeated evaluation of computational models may be demanding, which renders direct Bayesian inference prohibitive as it relies often on extensive sampling procedures. Surrogate models, particularly Gaussian process regression, are therefore frequently employed due to their capability to provide both predictions of the model response and a probabilistic characterization of the approximation uncertainty. Nevertheless, the uncertainty introduced by the surrogate model is not always treated consistently within the Bayesian inference procedure. This contribution proposes a framework for Bayesian model updating that explicitly propagates the predictive uncertainty induced by Gaussian process surrogates into the likelihood function via analytical marginalization. In addition, the proposed framework allows the simultaneous identification of the measurement noise characteristics. Moreover, an active learning strategy is developed in order to adaptively refine the surrogate model while controlling the uncertainty in the estimation of the model evidence. The resulting formulation provides a unified framework in which surrogate uncertainty, measurement noise, and adaptive learning are treated in a consistent probabilistic manner within Bayesian inference. The approach is illustrated by means of a simple dynamical system.

Keywords Bayesian model updating · Gaussian process surrogate · Surrogate uncertainty quantification · Active learning · Model evidence

1 Introduction

Bayesian model updating is widely employed in engineering for the calibration of computational models and the identification of unknown parameters based on experimental measurements (see, e.g., [1–3]). The result of the updating procedure is the posterior distribution of the unknown parameters, which combines prior knowledge about the system with the information contained in the measurements. In practice, however, the application of Bayesian updating requires repeated forward executions of the computational

✉ Zhouzhou Song
zhouzhou.song@tu-dortmund.de

Marcos A. Valdebenito
marcos.valdebenito@tu-dortmund.de

Chao Dang
chao.dang@tu-dortmund.de

Peipei Li
li.peipei@bjut.edu.cn

Pengfei Wei
pengfeiwei@nwpu.edu.cn

Sylvia Keßler
sylvia.kessler@hsu-hh.de

Matthias G. R. Faes
matthias.faes@tu-dortmund.de

[1] Chair for Reliability Engineering, TU Dortmund University, Leonhard-Euler-Str. 5, 44227 Dortmund, NRW, Germany

[2] College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100124, China

[3] School of Power and Energy, Northwestern Polytechnical University, Xi'an 710072, China

[4] Chair of Engineering Materials and Building Preservation, Helmut-Schmidt-University/University of the Federal Armed Forces Hamburg, Hamburg 22043, Germany

[5] International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, China

Published online: 05 June 2026

model, which may correspond to high-fidelity numerical simulations such as finite element models [4]. This renders direct Bayesian updating computationally demanding. Several strategies have therefore been proposed to alleviate this difficulty. A first class of approaches relies on local approximations of the posterior distribution, for instance through maximum a posteriori estimation combined with a Gaussian approximation [5]. Such methods are computationally efficient and easy to interpret. However, their applicability may be limited when the posterior distribution exhibits complex characteristics, such as strong nonlinearity or multimodality. A second class of approaches comprises simulation-based inference techniques, including Markov chain Monte Carlo methods (see, e.g., [6, 7]), transitional Markov chain Monte Carlo (see, e.g., [8, 9]), measure changes [10], or Bayesian updating based on structural reliability methods [11]. These approaches possess the capability to explore complex posterior distributions, but they may require a large number of forward model evaluations, often reaching tens of thousands of simulations. Finally, another widely used strategy consists in the application of surrogate models, in which the expensive computational model is replaced by a computationally inexpensive approximation constructed from a limited number of model evaluations (see, e.g., [12, 13]). In this context, the main challenges are the efficient construction of the surrogate model itself and the consistent treatment of the approximation error introduced by the surrogate representation.

In recent years, surrogate modelling has become closely connected with developments in machine learning, particularly in the context of data–driven approximation of complex input–output relationships arising in computational science and engineering. At the same time, surrogate modelling remains closely rooted in statistical modelling, typically operating in regimes characterized by limited and carefully designed data sets. From this perspective, probabilistic machine learning methods offer a natural framework for the construction of surrogate models. A prominent example of such approaches is Gaussian process regression (GPR) [14], which has gained considerable attention in the context of Bayesian updating. GPR possesses several attractive characteristics for this purpose. In particular, it provides a highly flexible nonparametric model capable of representing complex nonlinear relationships between parameters and model responses, while simultaneously delivering a probabilistic measure of prediction uncertainty. Different strategies have therefore been proposed for incorporating Gaussian process surrogates into Bayesian updating. In some approaches, the Gaussian process is employed as a direct surrogate of the computational model in order to replace expensive forward simulations (see, e.g., [15, 16]). Other formulations employ Gaussian processes to approximate specific components of the Bayesian updating formulation, such as the likelihood function or its logarithm (see, e.g., [17, 18]). Further developments attempt to explicitly propagate the epistemic uncertainty associated with the Gaussian process approximation through the Bayesian inference procedure (see, e.g., [19–21]), whereas alternative approaches rely on numerical treatments in which sample paths of the Gaussian process are simulated (see, e.g., [22]). In addition, several contributions consider the treatment and identification of the measurement noise affecting the experimental observations used in the updating process (see, e.g., [16]). It should be noted that the above discussion is not intended to provide an exhaustive survey of the literature, but rather to illustrate that the integration of Gaussian process surrogates within Bayesian updating is currently an active area of research.

Despite the considerable progress achieved in the integration of GPR within Bayesian model updating, several challenges remain. In particular, some existing approaches construct surrogate models for the likelihood function or its logarithm. While such strategies may be effective in certain situations, the likelihood may exhibit highly nonlinear and sharply peaked behaviour, which may render its accurate approximation difficult. From this perspective, it may be preferable to construct surrogate models of the computational response itself, which typically possesses smoother characteristics. Furthermore, the presence of measurement noise in the experimental observations should be explicitly incorporated into the updating formulation and treated as part of the identification problem (see, e.g., [5]). Another important aspect concerns the uncertainty introduced by the Gaussian process surrogate. In some implementations, this uncertainty is handled through numerical strategies, for instance by simulating sample paths of the Gaussian process, which may increase the computational burden. It is therefore desirable to develop formulations in which the uncertainty associated with the surrogate model can be consistently propagated through the Bayesian updating procedure, ideally in closed form, thereby avoiding additional simulation steps. Therefore, this contribution proposes a framework for the integration of GPR within Bayesian model updating in which the uncertainty introduced by the surrogate model is explicitly taken into account. In the proposed formulation, GPR is employed to approximate the response of the model, and the resulting predictive uncertainty is analytically marginalized into the likelihood function. As a consequence, the model evidence becomes a random variable induced by the epistemic uncertainty of the GPR approximation. Analytical expressions are derived for the expected value of the evidence together with an upper bound of its variance, both obtained with respect to the surrogate uncertainty introduced by the GPR. These expressions build upon previous developments related to the estimation of probability of failure in structural reliability problems using GPR surrogates [23]. Based on this characterization of the evidence, an adaptive learning strategy is introduced to sequentially enrich the training set

of the surrogate model while controlling the uncertainty in the estimation of the evidence. In this manner, the proposed methodology enables an efficient and consistent Bayesian updating procedure based on GPR, allowing the simultaneous identification of both model parameters and parameters characterizing measurement noise. The GPR surrogate is exclusively employed to approximate the structural response of the model, while the parameters governing the measurement noise are identified separately within the Bayesian framework. The present work therefore constitutes a step towards a fully probabilistic interpretation of Bayesian updating with GPR surrogates. It should be noted, however, that the hyperparameters of the GPR are treated as deterministic quantities obtained through maximum likelihood estimation, and their uncertainty is not explicitly propagated within the present formulation.

The rest of this paper is organized as follows. Sections 2 and 3 present the basic aspects and assumptions for Bayesian model updating and Gaussian process regression considered in this work, respectively. Then, Section 4 analyzes how to integrate the prediction uncertainty of the Gaussian process into Bayesian updating in a consistent way. Practical implementation of such an approach is discussed in Section 5. The application of the proposed framework is illustrated in Section 6 while the paper closes with conclusions and outlook in Section 7. It is emphasized that the present contribution focuses on the methodological integration of Gaussian process surrogate modelling within a Bayesian updating framework, together with the probabilistic treatment of surrogate-induced uncertainty. For this reason, the numerical investigations of Section 6 are conducted using a synthetic example specifically designed to assess the proposed methodology under controlled conditions, whereas experimental validation is left as a subject for future research.

2 Bayesian model updating under measurement uncertainty

2.1 Parameter space and forward model

Many practical problems in engineering involve predicting the behaviour of a physical system by means of a computational model that represents that system. In this context, the model is characterized by a set of parameters that describe physical properties, boundary conditions, or other quantities governing the system response. Let $\mathbf{x} \in \Omega_{\mathbf{x}} \subseteq \mathbb{R}^{n_x}$ denote the vector of model parameters, where n_x is the dimension of the parameter space and $\Omega_{\mathbf{x}}$ is the admissible parameter domain. Note that $\Omega_{\mathbf{x}}$ may incorporate physical, geometrical, or constitutive constraints on the parameter vector. The corresponding model response is denoted by $\mathbf{y} \in \mathbb{R}^{n_y}$, where n_y is the dimension of the response space. The relation between parameters and responses is established through a deterministic forward model, which can be expressed as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}), \tag{1}$$

where $\mathbf{f} : \Omega_{\mathbf{x}} \to \mathbb{R}^{n_y}$ denotes the model mapping. The evaluation of $\mathbf{f}$ may involve the execution of a complex numerical model, for instance a finite element model (see, e.g., [4]), followed by the extraction of selected response quantities from the full numerical solution. Therefore, $\mathbf{f}$ should not be understood as the underlying finite element solver itself, but rather as the complete input–output interface that transforms the model parameters $\mathbf{x}$ into the responses of interest $\mathbf{y}$. The evaluation of $\mathbf{f}(\mathbf{x})$ may be computationally demanding, since each model call may require the solution of a high-fidelity numerical problem. As a consequence, repeated evaluations of the forward model may become prohibitive when a large number of model calls is required, as it is the case in Bayesian updating.

2.2 Observation model

In practical applications, information about the physical system is obtained through repeated measurements of the quantities of interest. Let $\mathbf{y}_o^{(j)} \in \mathbb{R}^{n_y}$, $j = 1, \ldots, n_o$, denote the j-th observation vector, where n_o is the number of measurements. Each vector $\mathbf{y}_o^{(j)}$ contains the measured values of the response quantities corresponding to the computational model introduced in the previous subsection. Under the assumption of additive measurement noise (see, e.g., [24]), the observations are expressed as

$$\mathbf{y}_o^{(j)} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}^{(j)}, \qquad j = 1, \ldots, n_o, \tag{2}$$

where $\mathbf{f}(\mathbf{x})$ denotes the deterministic forward model and $\boldsymbol{\varepsilon}^{(j)} \in \mathbb{R}^{n_y}$ is the measurement error associated with the j-th observation. The measurement errors are assumed to be independent realizations of a multivariate Gaussian distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}_\varepsilon(\boldsymbol{\sigma})$, that is

$$\boldsymbol{\varepsilon}^{(j)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon(\boldsymbol{\sigma})), \tag{3}$$

where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The covariance matrix of the measurement noise is defined as

$$\boldsymbol{\Sigma}_\varepsilon(\boldsymbol{\sigma}) = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_{n_y}^2), \tag{4}$$

where $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_{n_y}]^{\mathrm{T}}$ collects the standard deviations associated with the response components and takes values in an admissible set $\Omega_{\boldsymbol{\sigma}} \subseteq \mathbb{R}_+^{n_y}$, $(\cdot)^{\mathrm{T}}$ denotes transpose, and $\mathrm{diag}(\cdot)$ denotes the diagonal matrix constructed from its vector argument. The diagonal structure of $\boldsymbol{\Sigma}_\varepsilon(\boldsymbol{\sigma})$ implies that

the measurement errors are assumed to be statistically independent across the different observation channels. Such an assumption is often reasonable when the measured quantities are obtained from sensors that do not exhibit significant cross-correlation. At the same time, different values of σ_k are admitted, which allows one to represent distinct noise levels in the different response components. This is important in practice, since the measured quantities may correspond to different physical magnitudes and may therefore possess different levels of experimental uncertainty. Furthermore, the measurement noise is assumed to possess zero mean. This assumption reflects the hypothesis that the computational model is able to reproduce the observed system behaviour without systematic discrepancy. In other words, no explicit bias term is introduced in the observation model, and the discrepancy between measured and predicted responses is attributed exclusively to random measurement errors.

Conditional on $(\mathbf{x}, \boldsymbol{\sigma})$, each observation vector follows

$$\mathbf{y}_o^{(j)} \mid \mathbf{x}, \boldsymbol{\sigma} \sim \mathcal{N}\big(\mathbf{f}(\mathbf{x}), \boldsymbol{\Sigma}_\varepsilon(\boldsymbol{\sigma})\big). \tag{5}$$

2.3 Bayesian formulation of the inverse problem

The objective of Bayesian updating is to estimate the unknown model parameters based on the available experimental observations. In the present setting, both the parameter vector $\mathbf{x}$ and the noise parameters $\boldsymbol{\sigma}$ are considered unknown. Therefore, these quantities are treated as random variables, and prior probability distributions are assigned to them, denoted by $p_{\mathbf{X}}(\mathbf{x})$ and $p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})$, respectively. These distributions represent the available knowledge about the parameters before the observations are taken into account. In general, such information may originate from previous experiments, engineering judgment, or physical constraints, and may therefore possess a certain degree of subjectivity.

Let $\mathcal{D}_o$ denote the set of observed data, defined as $\mathcal{D}_o = \left\{\mathbf{y}_o^{(j)}\right\}_{j=1}^{n_o}$, that is, the collection of the n_o observation vectors introduced in the previous subsection. According to Bayes' theorem, the posterior probability density of the parameters $(\mathbf{x}, \boldsymbol{\sigma})$ conditional on the observed data $\mathcal{D}_o$ is given by

$$p(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o) = \frac{\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o)\, p_{\mathbf{X}}(\mathbf{x})\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})}{c}, \tag{6}$$

where $p(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o)$ denotes the posterior probability density, $\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o)$ is the likelihood function, and c is the model evidence. The posterior distribution represents the updated knowledge about the unknown parameters after the observational information has been incorporated. In this way, the prior knowledge about $(\mathbf{x}, \boldsymbol{\sigma})$ is combined with the information contained in the observed data through the likelihood function. The likelihood function $\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o)$ is defined as the joint probability density of observing the data set $\mathcal{D}_o$ conditional on the model parameters $\mathbf{x}$ and on the noise level $\boldsymbol{\sigma}$, that is, $\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o) = p(\mathcal{D}_o \mid \mathbf{x}, \boldsymbol{\sigma})$. In the subsequent developments, the likelihood is interpreted as a function of the parameters for fixed observations. The mathematical definition of the likelihood follows from the observation model introduced in the previous subsection. Since the observation vectors $\mathbf{y}_o^{(j)}$ are assumed to be statistically independent and the measurement errors are Gaussian with diagonal covariance matrix, the likelihood of the parameters $(\mathbf{x}, \boldsymbol{\sigma})$ conditional on the observed data $\mathcal{D}_o$ can be written explicitly as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o) = \prod_{k=1}^{n_y} A_k(\sigma_k)\, \varphi\left(f_k(\mathbf{x}) \mid \bar{y}_{o,k}, \sigma_k^2/n_o\right), \tag{7}$$

where $A_k \in \mathbb{R}$ is a real scalar, $\varphi(\cdot \mid a, b^2)$ is the Gaussian probability density function with mean a and variance b^2, $\bar{y}_{o,k}$ is the mean of the observed data for the k-th component, and $f_k(\mathbf{x})$ is the k-th component of the deterministic forward model $\mathbf{f}(\mathbf{x})$. A detailed derivation of this expression, along with the precise definition of the factor A_k, can be found in Appendix A.1. Nevertheless, Eq. 7 admits a transparent statistical interpretation. For each response component k, the likelihood evaluates how well the model prediction $f_k(\mathbf{x})$ agrees with the empirical mean $\bar{y}_{o,k}$ of the available observations. The Gaussian probability density function therefore quantifies the plausibility of the prediction in view of the measurement data. Because the observation vectors are assumed statistically independent with Gaussian measurement errors of variance σ_k^2, the likelihood can be expressed in terms of the sample mean, whose variance equals σ_k^2/n_o. This scaling reflects the progressive reduction of effective observation uncertainty as the number of measurements increases, so that the likelihood becomes increasingly concentrated around values of $\mathbf{x}$ that produce predictions consistent with the observed average response.

The constant c appearing in Eq. 6 is the model evidence, which is defined as

$$c = \int_{\Omega_{\boldsymbol{\sigma}}} \int_{\Omega_{\mathbf{x}}} \mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o)\, p_{\mathbf{X}}(\mathbf{x})\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\sigma}. \tag{8}$$

The evidence acts as a normalizing constant that ensures that the posterior density integrates to unity. In addition, this quantity plays a central role in Bayesian model class selection and model averaging (see, e.g., [8, 25]). However, model selection is not considered in the present work, and therefore the evidence appears here only as a normalization term in Bayes' theorem.

Although Eq. 6 provides a formal expression for the posterior distribution, the practical evaluation of this quantity is usually computationally demanding. In particular, the likelihood function requires the evaluation of the forward model

$\mathbf{f}(\mathbf{x})$, which may involve the execution of a computationally expensive numerical model, such as a high-fidelity finite element simulation. Consequently, the repeated evaluations required for posterior analysis may become prohibitive in practice. This difficulty becomes particularly evident when statistical quantities of interest are to be computed from the posterior distribution. For instance, the estimation of posterior expectations or marginal distributions typically require the evaluation of integrals over the parameter space, which are commonly approximated using stochastic simulation methods such as transitional Markov chain Monte Carlo [8]. In such cases, a large number of evaluations of the likelihood function are required, and therefore the computational cost of the forward model becomes a critical bottleneck.

3 Gaussian process surrogate model

In many engineering applications, the evaluation of the deterministic forward model $\mathbf{f}(\mathbf{x})$ is computationally demanding. To alleviate this cost, the model is replaced by a Gaussian process (GP) surrogate constructed from a set of training evaluations.

Let

$$\mathcal{D}_t = \left\{ (\mathbf{x}_t{}^{(i)}, \mathbf{y}_t{}^{(i)}) \right\}_{i=1}^{n_t}, \qquad \mathbf{y}_t{}^{(i)} = \mathbf{f}(\mathbf{x}_t{}^{(i)}) \tag{9}$$

denote the training data set consisting of n_t evaluations of the forward model. Here, $\mathbf{x}_t{}^{(i)} \in \mathbb{R}^{n_x}$ denotes the input parameters and $\mathbf{y}_t{}^{(i)} \in \mathbb{R}^{n_y}$ the corresponding model response. The training inputs may be selected using any appropriate design-of-experiments strategy, such as Latin hypercube sampling or other space-filling approaches (see, e.g., [26–28]).

Since the model response is generally vector-valued, a separate Gaussian process is constructed for each component of the response vector. While this strategy considerably simplifies the surrogate construction, it neglects possible statistical dependencies between the components of the response vector. Nevertheless, this assumption is retained here for the sake of simplicity and because the considered response vectors possess moderate dimension. For high-dimensional responses, the framework could be extended by incorporating dimensionality-reduction strategies (see, e.g., [16]) or multi-output surrogate formulations (see, e.g., [29]) in order to explicitly account for output dependencies.

Let $f_k(\mathbf{x})$ denote the k-th component of the deterministic model response. In the Gaussian process framework, this function is approximated by a stochastic surrogate $F_k(\mathbf{x})$, which is modeled as a realization of a Gaussian process, i.e.,

$$F_k(\mathbf{x}) \sim \mathcal{GP}\big(\mu_k(\mathbf{x} \mid \boldsymbol{\theta}), \kappa_k(\mathbf{x}, \mathbf{x}' \mid \boldsymbol{\theta})\big), \qquad k = 1, \ldots, n_y. \tag{10}$$

Here, $\mathcal{GP}(\cdot, \cdot)$ denotes a Gaussian-process prior over functions, $\mu_k(\mathbf{x} \mid \boldsymbol{\theta})$ denotes the prior mean function, $\kappa_k(\mathbf{x}, \mathbf{x}' \mid \boldsymbol{\theta})$ denotes the prior covariance kernel, and $\boldsymbol{\theta}$ are the hyperparameters of the Gaussian process, which may contain, e.g., terms related with the mean, variance and correlation lengths. These hyperparameters are estimated from the training data set $\mathcal{D}_t$, for instance by maximum likelihood estimation. It is noted that such an approach neglects the uncertainty associated with the estimation of the hyperparameters; a fully Bayesian treatment could also assign prior distributions to $\boldsymbol{\theta}$ and infer their posterior distribution (see, e.g., [30]), although at the expense of additional computational complexity. In the following, the dependence of the mean function μ_k and covariance kernel κ_k on $\boldsymbol{\theta}$ is not denoted explicitly to keep the notation compact.

Conditioning the Gaussian process on the training data $\mathcal{D}_t$ yields the posterior predictive distribution of each response component

$$F_k(\mathbf{x}) \mid \mathcal{D}_t \sim \mathcal{GP}\big(\mu_{t,k}(\mathbf{x}), \kappa_{t,k}(\mathbf{x}, \mathbf{x}')\big), \qquad k = 1, \ldots, n_y. \tag{11}$$

For the k-th component, the posterior mean and covariance functions are given by [14]

$$\mu_{t,k}(\mathbf{x}) = \mu_k(\mathbf{x}) + \boldsymbol{\kappa}_k(\mathbf{x}, \mathbf{X}_t)\mathbf{K}_k^{-1}\left(\mathbf{y}_{t,k} - \boldsymbol{\mu}_k(\mathbf{X}_t)\right), \tag{12}$$

$$\kappa_{t,k}(\mathbf{x}, \mathbf{x}') = \kappa_k(\mathbf{x}, \mathbf{x}') - \boldsymbol{\kappa}_k(\mathbf{x}, \mathbf{X}_t)\mathbf{K}_k^{-1}\boldsymbol{\kappa}_k(\mathbf{X}_t, \mathbf{x}'), \tag{13}$$

where $\mathbf{X}_t$ is a matrix whose rows contain the input training data $\mathbf{x}_t{}^{(i)}$, and $\mathbf{K}_k$ denotes the covariance matrix of the training inputs,

$$[\mathbf{K}_k]_{ij} = \kappa_k\left(\mathbf{x}_t^{(i)}, \mathbf{x}_t^{(j)}\right), \qquad i, j = 1, \ldots, n_t. \tag{14}$$

Furthermore, $\boldsymbol{\kappa}_k(\mathbf{x}, \mathbf{X}_t)$ denotes the vector of covariances between the prediction location $\mathbf{x}$ and the training inputs, and $\mathbf{y}_{t,k} = \left[y_{t,k}^{(1)}, \ldots, y_{t,k}^{(n_t)}\right]^{\mathrm{T}}$ collects the k-th components of the training outputs. Analogously, $\boldsymbol{\mu}_k(\mathbf{X}_t) = \left[\mu_k\left(\mathbf{x}_t^{(1)}\right), \ldots, \mu_k\left(\mathbf{x}_t^{(n_t)}\right)\right]^{\mathrm{T}}$ denotes the vector of mean-function evaluations at the training input locations. In addition, it is convenient to define the predictive variance as

$$s_{t,k}^2(\mathbf{x}) = \kappa_{t,k}(\mathbf{x}, \mathbf{x}). \tag{15}$$

Note that the Gaussian process interpolates the training observations exactly in the noise-free formulation adopted here (i.e. when the covariance matrix $\mathbf{K}_k$ does not include an observation-noise term). Consequently, if $\mathbf{x} = \mathbf{x}_t{}^{(i)}$ coincides with one of the training points, the predictive mean satisfies

$\mu_{t,k}\left(\mathbf{x}_t^{(i)}\right) = y_{t,k}^{(i)}$ for $k = 1, \ldots, n_y$, and the predictive variance becomes zero.

Replacing the deterministic model $\mathbf{f}(\mathbf{x})$ by the predictive random vector $\mathbf{F}(\mathbf{x}) = [F_1(\mathbf{x}), \ldots, F_{n_y}(\mathbf{x})]^T$ introduces an additional layer of uncertainty into the inference problem, which reflects the limited number of model evaluations used to construct the surrogate model.

4 Bayesian updating with Gaussian process predictive uncertainty

4.1 Effect of the surrogate model on Bayesian updating

The introduction of the Gaussian process surrogate in the forward model has direct consequences for the probabilistic structure of the Bayesian updating problem. In particular, once the deterministic model response is replaced by its surrogate representation $\mathbf{F}(\mathbf{x})$, the likelihood function must be interpreted as a functional with respect to the epistemic uncertainty associated with the Gaussian process. Accordingly, the likelihood can be written as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F}) = p(\mathcal{D}_o \mid \mathbf{x}, \boldsymbol{\sigma}, \mathbf{F}(\mathbf{x}))\,, \tag{16}$$

which emphasizes that, for each realization of the surrogate field $\mathbf{F}(\cdot)$, a distinct likelihood function over the parameter space is obtained. In this way, the likelihood becomes a random functional induced by the epistemic uncertainty of the Gaussian process model. It should be noted that the surrogate field $\mathbf{F}(\mathbf{x})$ appearing in Eq. 16 is understood as the posterior Gaussian process conditioned on the training data set $\mathcal{D}_t$ introduced in Section 3. For the sake of notational compactness, this conditioning is not indicated explicitly in Eq. 16. More precisely, the likelihood functional should be interpreted as depending on the random vector $\mathbf{F}(\mathbf{x}) \mid \mathcal{D}_t$, whose probabilistic characterization is given by Eq. 11. Such an interpretation directly affects the normalization constant of Bayes' theorem. In the deterministic formulation presented previously, the evidence was denoted by the scalar quantity c in Eq. 8. However, due to the randomness introduced through the surrogate representation, the evidence now becomes a random variable, which is denoted in the following by the uppercase symbol C. It can be expressed as

$$C = \int_{\Omega_{\boldsymbol{\sigma}}} \int_{\Omega_{\mathbf{x}}} \mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})\; p_{\mathbf{X}}(\mathbf{x})\; p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\sigma}\,, \tag{17}$$

where the dependence of C on the realization of $\mathbf{F}(\cdot)$ is implicit. Consequently, the normalization constant inherits the epistemic uncertainty of the surrogate model and must therefore be regarded as a random variable.

By application of Bayes' theorem, the posterior distribution of the unknown parameters becomes

$$p(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F}) = \frac{\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})\; p_{\mathbf{X}}(\mathbf{x})\; p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})}{C}. \tag{18}$$

In view of Eqs. 16 and 17, it follows that the posterior distribution must be interpreted as a random field over the parameter space, indexed by the epistemic uncertainty of the Gaussian process surrogate. In fact, each realization of $\mathbf{F}(\cdot)$ induces a distinct posterior density, which implies that the Bayesian updating problem acquires an additional hierarchical level. Although such a formulation is conceptually consistent, its direct numerical treatment may become impractical in many applications. The representation and propagation of the posterior random field typically leads to considerable computational and methodological difficulties. Therefore, alternative strategies that enable a tractable treatment of surrogate-induced uncertainty are required. Such strategies are discussed in the following subsection.

4.2 Posterior formulation based on the expected likelihood

The formulation introduced in the previous subsection is probabilistically consistent but leads to a posterior random field that is difficult to manipulate in practical applications. In order to obtain a tractable formulation, it is assumed in the following that the Gaussian process surrogate is sufficiently accurate so that the expectation of the likelihood with respect to its epistemic uncertainty provides *sufficient accuracy* of the statistical information conveyed by the observations [31, 32]. At this stage, the notion of *sufficient accuracy* is introduced as a working assumption only. Its precise meaning and a practical criterion for its assessment will be discussed explicitly in Section 4.3.

Let $\mathbb{E}_{\mathrm{GP}}[\cdot]$ denote the expectation operator with respect to the posterior distribution of the Gaussian process conditioned on the training data $\mathcal{D}_t$. Then, the expectation of the likelihood in Eq. 16 admits the following analytical expression (see Appendix A.2):

$$\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})] = \prod_{k=1}^{n_y} A_k(\sigma_k)\, \varphi\Big(\mu_{t,k}(\mathbf{x}) \,\Big|\, \bar{y}_{o,k}, \sigma_k^2/n_o + s_{t,k}^2(\mathbf{x})\Big)\,. \tag{19}$$

Equation 19 admits a clear probabilistic interpretation. In particular, the expected likelihood preserves the functional structure of the deterministic expression given in Eq. 7. The role of the deterministic model response $f_k(\mathbf{x})$ is now substituted by the posterior predictive mean of the Gaussian process surrogate, $\mu_{t,k}(\mathbf{x})$. In this manner, the surrogate prediction enters the likelihood in a form that is analogous to

the classical formulation, while accounting for the information provided by the training data set $\mathcal{D}_t$. At the same time, the variance associated with the Gaussian probability density function is augmented by the predictive variance of the surrogate model. More precisely, the term σ_k^2/n_o arising from the measurement noise is combined with the epistemic uncertainty $s_{t,k}^2(\mathbf{x})$ of the Gaussian process. This results in an effective broadening of the likelihood contribution in regions of the parameter space where the surrogate prediction is uncertain. Conversely, in the limit $s_{t,k}^2(\mathbf{x}) \to 0$, which corresponds to highly accurate surrogate predictions or to interpolation at training points, the expected likelihood reduces to the deterministic formulation. Therefore, Eq. 19 naturally reflects the combined influence of measurement uncertainty and surrogate-induced epistemic uncertainty. The likelihood contribution becomes wider in regions where the Gaussian process prediction is less reliable, whereas it approaches the classical expression in regions where the surrogate model provides confident predictions.

The treatment of the evidence follows an analogous reasoning. Starting from the random evidence introduced previously, its expectation with respect to the Gaussian process posterior can be written as

$$\mathbb{E}_{\mathrm{GP}}[C] = \mathbb{E}_{\mathrm{GP}}\left[\int_{\Omega_\sigma}\int_{\Omega_\mathbf{x}} \mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})\, p_{\mathbf{X}}(\mathbf{x})\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\sigma}\right]. \quad (20)$$

Since the likelihood and prior densities are non-negative functions, Tonelli's theorem (see, e.g., [23]) justifies the interchange between expectation and integration. Consequently,

$$\mathbb{E}_{\mathrm{GP}}[C] = \int_{\Omega_\sigma}\int_{\Omega_\mathbf{x}} \mathbb{E}_{\mathrm{GP}}\left[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})\right]\, p_{\mathbf{X}}(\mathbf{x})\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\sigma}. \quad (21)$$

In order to obtain a tractable deterministic representation of the inference problem, the likelihood function is replaced by its expectation with respect to the posterior distribution of the Gaussian process surrogate. This leads to the posterior density

$$p(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o, \mathcal{D}_t) \propto \mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})]\; p_{\mathbf{X}}(\mathbf{x})\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}). \quad (22)$$

It should be emphasized that the expectation of the likelihood is employed here as a tractable closure of the random posterior field introduced in Eq. 18. The adequacy of this approximation depends on the level of predictive uncertainty of the Gaussian process model, which will be quantified in the following subsection.

In view of Eq. 22, the Bayesian updating problem recovers a deterministic representation in which the epistemic uncertainty associated with the Gaussian process surrogate has been integrated out explicitly through the expectation operator $\mathbb{E}_{\mathrm{GP}}[\cdot]$. From a formal standpoint, Eq. 22 constitutes the counterpart of the classical posterior distribution obtained in the absence of surrogate modelling. However, it should be emphasized that the resulting posterior density remains conditioned on both the observation data $\mathcal{D}_o$ as well as on the training data set $\mathcal{D}_t$ used for the construction of the Gaussian process surrogate. In this manner, the information contained in the surrogate model enters the inference procedure in an implicit yet fully probabilistic form. An important consequence of this formulation is that, once the expected likelihood has been evaluated, the Bayesian updating problem assumes again a structure that is analogous to the standard setting as in Eq. 6. Therefore, posterior inference may be performed by means of any suitable strategy. The remaining issue concerns the verification that the Gaussian process surrogate provides a sufficiently accurate representation of the forward model so that the expected likelihood in Eq. 22 remains representative of the statistical information conveyed by the observations. The development of a practical criterion for assessing the adequacy of the surrogate model is addressed in the following subsection.

4.3 Criterion for surrogate accuracy based on the variability of the evidence

The introduction of the Gaussian process surrogate renders the Bayesian evidence a random variable with respect to the epistemic uncertainty associated with the surrogate model. In the present contribution, this property is exploited in order to define a practical criterion for assessing whether the surrogate model is sufficiently accurate for the purposes of Bayesian updating. Let the coefficient of variation (CoV) of the evidence be defined as

$$\mathrm{CoV}[C] = \frac{\sqrt{\mathrm{Var}_{\mathrm{GP}}[C]}}{\mathbb{E}_{\mathrm{GP}}[C]}, \quad (23)$$

where $\mathbb{E}_{\mathrm{GP}}[C]$ and $\mathrm{Var}_{\mathrm{GP}}[C]$ denote the expectation and variance of the evidence with respect to the Gaussian process posterior distribution. The surrogate model is regarded as sufficiently accurate if the quantity in Eq. 23 remains below a prescribed tolerance. This assumption is motivated by the observation that, in the limit of vanishing surrogate uncertainty, the variability of the evidence also vanishes.

4.3.1 Exact representation of the variance of the evidence

Starting from the definition of the random evidence in Eq. 17, its variance can be written as

$$\mathrm{Var}_{\mathrm{GP}}[C] = \mathbb{E}_{\mathrm{GP}}[C^2] - \left(\mathbb{E}_{\mathrm{GP}}[C]\right)^2. \quad (24)$$

The term $\mathbb{E}_{\mathrm{GP}}[C]$ on the right-hand side of the above equation is already available from Eq. 21. Meanwhile, Eq. 81 in Appendix A.3 provides a closed-form expression for $\mathbb{E}_{\mathrm{GP}}[C^2]$ with respect to the epistemic uncertainty of the Gaussian process. Nevertheless, this last expression depends on the bivariate Gaussian probability distribution, which must be computed for all pairs of parameter values $(\mathbf{x}, \mathbf{x}')$. Consequently, although an exact semi-analytical representation of $\mathrm{Var}_{\mathrm{GP}}[C]$ is available in principle, its practical numerical evaluation becomes computationally prohibitive. Indeed, as the associated integrals must be computed by sampling methods, that would demand evaluating the bivariate Gaussian distribution for all possible combinations of those samples. This motivates the introduction of a tractable approximation, which is discussed in the following subsection.

4.3.2 Upper bound for the variance of the evidence

The variance of the evidence is bounded from above, that is:

$$\mathrm{Var}_{\mathrm{GP}}[C] \leq \overline{\mathrm{Var}}_{\mathrm{GP}}[C] \tag{25}$$

where $\overline{\mathrm{Var}}_{\mathrm{GP}}[C]$ denotes the upper bound, which is defined as:

$$\overline{\mathrm{Var}}_{\mathrm{GP}}[C] = \int_{\Omega_{\boldsymbol{\sigma}}} \int_{\Omega_{\mathbf{x}}} \mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})]\, p_{\mathbf{X}}(\mathbf{x})\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, \mathrm{d}\mathbf{x}\mathrm{d}\boldsymbol{\sigma} \tag{26}$$

where $\mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})]$ denotes the variance of the likelihood with respect to the Gaussian process. Detailed derivations of Eq. 26 as well as of the latter variance are presented in Appendix A.4. The advantage of these expressions is that the epistemic uncertainty introduced by the surrogate model can be treated in closed form.

4.3.3 Surrogate-accuracy criterion

The Gaussian process surrogate is regarded as sufficiently accurate for the purposes of Bayesian updating if the upper bound for the coefficient of variation of the evidence satisfies

$$\overline{\mathrm{CoV}}[C] = \frac{\sqrt{\overline{\mathrm{Var}}_{\mathrm{GP}}[C]}}{\mathbb{E}_{\mathrm{GP}}[C]} \leq \varepsilon_{\mathrm{GP}}, \tag{27}$$

where $\varepsilon_{\mathrm{GP}}$ denotes a prescribed tolerance. In practical applications, typical values of this tolerance may range between 1% and 10%, depending on the required level of reliability of the posterior inference.

It should be emphasized that the quantity $\overline{\mathrm{CoV}}[C]$ depends on the training data set $\mathcal{D}_t$ used for the construction of the Gaussian process surrogate. Once an initial surrogate has been built, the corresponding value of $\overline{\mathrm{CoV}}[C]$ is therefore determined. In principle, the surrogate uncertainty could be reduced by a brute-force strategy consisting in generating a very large number of additional training samples and recomputing the surrogate model. However, such an approach may lead to an unnecessary increase of the computational cost, since not all regions of the parameter space contribute equally to the Bayesian updating procedure. A more efficient strategy consists in sequentially enriching the training data set by selecting new parameter points that most effectively reduce the surrogate-induced variability of the evidence. In this way, surrogate accuracy can be improved in a targeted manner, concentrating computational effort in regions that are relevant for inference. The development of such an adaptive learning strategy is the subject of the following section.

4.4 Adaptive control of surrogate uncertainty

The adaptive refinement of the Gaussian process surrogate is guided by the point-wise contribution of the parameter vector $\mathbf{x}$ to the surrogate-induced variability of the evidence. Since additional training information can only be generated through evaluations of the forward model at selected parameter locations $\mathbf{x} \in \Omega_{\mathbf{x}}$, whereas the noise parameters $\boldsymbol{\sigma} \in \Omega_{\boldsymbol{\sigma}}$ are not controllable, it is convenient to introduce a marginal measure of variability with respect to $\boldsymbol{\sigma}$. To this end, define

$$\psi(\mathbf{x}) = \int_{\Omega_{\boldsymbol{\sigma}}} \mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})]\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, \mathrm{d}\boldsymbol{\sigma}. \tag{28}$$

Following Eq. 26, the upper bound of the variance of the evidence can be expressed as

$$\overline{\mathrm{Var}}_{\mathrm{GP}}[C] = \int_{\Omega_{\mathbf{x}}} \psi(\mathbf{x})\, p_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{29}$$

The function $\psi(\mathbf{x})$ therefore represents the expected contribution of the parameter point $\mathbf{x}$ to the surrogate-induced variability of the evidence after marginalization with respect to the noise parameters. Large values of $\psi(\mathbf{x})\, p_{\mathbf{X}}(\mathbf{x})$ indicate regions of the parameter space that simultaneously possess significant prior probability mass and exhibit a strong sensitivity of the likelihood with respect to surrogate uncertainty. Consequently, enriching the training set in such regions is expected to reduce the variability of the evidence in an efficient manner [33].

Although the function $\psi(\mathbf{x})$ provides a natural measure of the local influence of surrogate uncertainty, it should be noted that large values of $\mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})]$ may arise in regions where the likelihood itself is negligible. In such regions, an improvement of surrogate accuracy is not expected to significantly affect the Bayesian updating results, since the corresponding parameter configurations contribute

only marginally to the evidence and to the posterior distribution. In order to account for this effect, a smooth relevance weight w is introduced in the form

$$w(\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o,\mathbf{F})]) = \frac{\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o,\mathbf{F})]}{\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o,\mathbf{F})]+\tau}, \tag{30}$$

where $\tau > 0$ is a scaling parameter, whose selection strategy is discussed in Section 5. It is observed that:

- if $\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o,\mathbf{F})] \gg \tau$, then $w \approx 1$, and the contribution of surrogate variance is fully retained;
- if $\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o,\mathbf{F})] \ll \tau$, then $w \approx \mathbb{E}_{\mathrm{GP}}[\mathcal{L}]/\tau$, and the corresponding contribution is strongly attenuated.

In this manner, regions that are negligible with respect to the evidence are automatically down-weighted, whereas parameter configurations that are relevant for posterior inference maintain a dominant influence.

Based on this construction, the acquisition function $a(\mathbf{x})$ is defined as

$$a(\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x}) \int_{\Omega_\sigma} \mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o,\mathbf{F})]\, w(\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma} \mid \mathcal{D}_o,\mathbf{F})])\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, \mathrm{d}\boldsymbol{\sigma}. \tag{31}$$

The function $a(\mathbf{x})$ quantifies the relevance-weighted contribution of the parameter point $\mathbf{x}$ to the variability of the evidence, taking into account both surrogate uncertainty and prior probability mass. Therefore, the adaptive enrichment of the Gaussian process surrogate is achieved by selecting new training locations corresponding to parameter vectors that maximize $a(\mathbf{x})$. In this way, the computational effort is concentrated in regions of the parameter space that are simultaneously influential for Bayesian inference and insufficiently resolved by the current surrogate model.

5 Practical implementation

5.1 Overall strategy

The objective of the proposed methodology is the solution of the Bayesian updating problem introduced in Section 2. In this context, repeated evaluations of the deterministic forward model may become computationally prohibitive. To alleviate this difficulty, the forward model is replaced by a Gaussian process surrogate constructed according to the formulation presented in Section 3. The surrogate model is built from an initial training data set $\mathcal{D}_t$ comprising a limited number of forward model evaluations.

Once the surrogate model has been constructed, it is necessary to verify whether its predictive accuracy is sufficient for the purposes of Bayesian inference. In the present work, this verification is performed by assessing the variability of the model evidence induced by the epistemic uncertainty of the Gaussian process. In particular, the expected value of the evidence, obtained by marginalization with respect to the surrogate uncertainty (see Eq. 21), together with the upper bound of its variance (see Eq. 26), are employed to evaluate the coefficient of variation criterion defined in Eq. 27. If this condition is satisfied, the surrogate model is regarded as sufficiently accurate for the updating procedure. If the surrogate accuracy criterion is not fulfilled, the training data set is sequentially enriched. New training points are selected in the parameter space by means of the acquisition function introduced in Eq. 31, which identifies regions that contribute most significantly to the surrogate-induced variability of the evidence. The forward model is evaluated at the selected locations and the Gaussian process surrogate is subsequently updated. This adaptive refinement process is repeated until the convergence condition in Eq. 27 is satisfied.

Once the surrogate model is deemed sufficiently accurate, Bayesian updating is performed using the posterior formulation given in Eq. 22, in which the likelihood function has been analytically marginalized with respect to the surrogate uncertainty. In this manner, the updating procedure consistently accounts for measurement uncertainty, prior information, and the epistemic uncertainty associated with the surrogate representation.

5.2 Monte Carlo estimation and posterior approximation

The practical application of the proposed framework described above requires the numerical evaluation of several multidimensional integrals arising in the characterization of the evidence as well as in the subsequent posterior inference. In the present work, these integrals are approximated by means of plain Monte Carlo simulation (see, e.g., [34]) based on sampling from the prior distributions of the model parameters and noise variables. For each sample, the relevant integrand quantities are evaluated using the Gaussian process surrogate model. Convergence of the quantities with respect to Monte Carlo sampling is monitored through its coefficient of variation, which should be below a predefined threshold $\varepsilon_{\mathrm{MC}}$. Once the surrogate model satisfies the convergence condition introduced in Eq. 27, the same prior-based Monte Carlo sample set is employed for posterior approximation. In particular, the expected likelihood values evaluated at the prior samples are used to construct importance weights proportional to the numerator of the posterior density in Eq. 22. Normalization of these weights provides a discrete approximation of the posterior distribution supported on the prior sample

set. Posterior samples may then be obtained by means of a sampling-importance-resampling procedure [35], enabling the estimation of marginal distributions, posterior moments, and credible intervals. The distribution of the importance weights is monitored in order to assess the quality of the posterior representation, for instance through indicators related to effective sample size.

It is recognized that prior-based Monte Carlo integration and posterior approximation described above are generally not optimal from the viewpoint of simulation, particularly in situations where the posterior distribution is strongly concentrated in regions of low prior probability, a difficulty that becomes more severe with increasing dimensionality of the parameter vector being identified (see, e.g., [8]). However, the present contribution does not aim at applying advanced simulation schemes for inference. Instead, the focus is placed on the consistent probabilistic treatment and control of surrogate-induced uncertainty within Bayesian updating. From this perspective, the use of prior sampling provides a simple and robust numerical framework. Moreover, since the Gaussian process surrogate can be evaluated at negligible computational cost compared with the original forward model, very large Monte Carlo sample sizes can be employed in practice, thereby mitigating the inefficiency associated with prior-based sampling. The specific implementation details are described in the following subsection.

5.3 Summary of the adaptive algorithm

The proposed Bayesian updating framework combines prior-based Monte Carlo integration, Gaussian process surrogate modelling, and an adaptive enrichment strategy aimed at controlling the epistemic uncertainty induced by the surrogate representation. The overall procedure can be summarized in the following sequence of steps.

Step 1: Problem definition. The dimensions n_x, n_y and n_o are specified together with the observed dataset $\mathcal{D}_o$, the prior probability densities $p_{\mathbf{X}}(\mathbf{x})$ and $p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})$, and the deterministic forward model $\mathbf{y} = \mathbf{f}(\mathbf{x})$. Tolerance levels $\varepsilon_{\mathrm{GP}}$ and $\varepsilon_{\mathrm{MC}}$ are prescribed for the control of surrogate-induced uncertainty and Monte Carlo integration error, respectively. In practical applications, these tolerances are typically selected in the range between 1% and 10%, depending on the desired level of accuracy of the posterior inference.

Step 2: Definition of auxiliary functions. Define the auxiliary function $\xi_1(\mathbf{x}, \boldsymbol{\sigma})$

$$\xi_1(\mathbf{x}, \boldsymbol{\sigma}) = \prod_{k=1}^{n_y} A_k(\sigma_k)\ \varphi\Big(\mu_{t,k}(\mathbf{x}) \,\big|\, \bar{y}_{o,k},\, (\sigma_k)^2/n_o + s^2_{t,k}(\mathbf{x})\Big), \tag{32}$$

which corresponds to evaluating $\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})]$ as shown in Eq. 19. In addition, define the auxiliary function $\xi_2(\mathbf{x}, \boldsymbol{\sigma})$

$$\xi_2(\mathbf{x}, \boldsymbol{\sigma}) = \prod_{k=1}^{n_y} A_k(\sigma_k)^2 \sqrt{\frac{n_o}{4\pi\sigma_k^2}}\, \varphi\big(\mu_{t,k}(\mathbf{x}) \mid \bar{y}_{o,k},\, (\sigma_k)^2/(2n_o) + s^2_{t,k}(\mathbf{x})\big), \tag{33}$$

which follows from Eqs. 89 and 91 and allows evaluating $\mathbb{E}_{\mathrm{GP}}\big[\mathcal{L}^2(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})\big]$. Finally, define the auxiliary function $\xi_3(\mathbf{x}, \boldsymbol{\sigma})$

$$\xi_3(\mathbf{x}, \boldsymbol{\sigma}) = \xi_2(\mathbf{x}, \boldsymbol{\sigma}) - (\xi_1(\mathbf{x}, \boldsymbol{\sigma}))^2, \tag{34}$$

which follows from Eq. 88 and allows evaluating $\mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})]$.

Step 3: Generation of prior sample pools and training data. Two independent Monte Carlo sample pools are generated from the prior distributions,

$$\mathcal{S}_{\mathbf{x}} = \left\{\mathbf{x}^{(i)}\right\}_{i=1}^{N} \sim p_{\mathbf{X}}(\mathbf{x}), \qquad \mathcal{S}_{\boldsymbol{\sigma}} = \left\{\boldsymbol{\sigma}^{(i)}\right\}_{i=1}^{N} \sim p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}), \tag{35}$$

where N denotes the sample size. Since the priors are assumed independent, the indexed pairs $\left(\mathbf{x}^{(i)}, \boldsymbol{\sigma}^{(i)}\right)$, $i = 1, \ldots, N$ constitute independent and identically distributed samples from the joint prior density $p_{\mathbf{X}}(\mathbf{x})p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})$. These pairs are employed for the Monte Carlo estimation of the different integrals appearing in the formulation. In addition, an auxiliary set of candidate parameter samples

$$\mathcal{C}_{\mathbf{x}} = \left\{\mathbf{x}_c^{(j)}\right\}_{j=1}^{N_C} \sim p_{\mathbf{X}}(\mathbf{x}) \tag{36}$$

is generated again according to Monte Carlo sampling. This set will later be used in the adaptive selection of new training locations for the Gaussian process surrogate. Finally, an initial training data set $\mathcal{D}_t$ to be used for constructing the Gaussian process surrogate is generated using a suitable space-filling design of experiments, as defined in Eq. 9.

Step 4: Construction of the initial surrogate model. Based on the data set $\mathcal{D}_t$, independent Gaussian process surrogate models are constructed for each component of the model response. This yields the posterior predictive mean function $\mu_{t,k}(\mathbf{x})$, given in Eq. 12, together with the corresponding predictive variance $s^2_{t,k}(\mathbf{x})$, defined in Eq. 15.

Step 5: Monte Carlo estimation of the expected value of the evidence and of the upper bound of its variance. The evidence is estimated by

$$\widehat{\mathbb{E}}_{\mathrm{GP}}[C] = \frac{1}{N}\sum_{i=1}^{N} \xi_1\left(\mathbf{x}^{(i)}, \boldsymbol{\sigma}^{(i)}\right), \tag{37}$$

which corresponds to the Monte Carlo approximation of Eq. 21. Moreover, the upper bound of the variance of the evidence is estimated as

$$\widehat{\mathrm{Var}}_{\mathrm{GP}}[C] = \frac{1}{N}\sum_{i=1}^{N} \xi_3\left(\mathbf{x}^{(i)}, \boldsymbol{\sigma}^{(i)}\right), \tag{38}$$

which follows from Eq. 26 together with the pointwise definition of $\mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})]$ given in Eq. 88.

Step 6: Control of Monte Carlo integration error. The coefficient of variation associated with the estimators in Eqs. 37 and 38 is monitored using their respective sample coefficients of variation. If any of these coefficients of variation exceeds the tolerance $\varepsilon_{\mathrm{MC}}$, the sample pools $\mathcal{S}_{\mathbf{x}}$ and $\mathcal{S}_{\boldsymbol{\sigma}}$ are enlarged, e.g., by doubling N, and the algorithm returns to step 5. In case the coefficients of variation are below the threshold $\varepsilon_{\mathrm{MC}}$, the algorithm proceeds to step 7.

Step 7: Surrogate convergence assessment. If the condition

$$\widehat{\mathrm{CoV}}[C] = \frac{\sqrt{\widehat{\mathrm{Var}}_{\mathrm{GP}}[C]}}{\widehat{\mathbb{E}}_{\mathrm{GP}}[C]} \leq \varepsilon_{\mathrm{GP}} \tag{39}$$

is satisfied twice in a row, the surrogate model is regarded as sufficiently accurate for posterior inference and the execution of the algorithm proceeds to step 10. Otherwise, the training set is enriched as described in Steps 8-9. Note that Eq. 39 corresponds to Eq. 27 evaluated with the corresponding estimators.

Step 8: Definition of scaling parameter. The scaling parameter τ associated with the acquisition function is selected as:

$$\tau = \mathrm{Q}_\alpha\left(\left\{\xi_1\left(\mathbf{x}^{(i)}, \boldsymbol{\sigma}^{(i)}\right)\right\}_{i=1}^{N}\right), \tag{40}$$

where $\mathrm{Q}_\alpha(\cdot)$ denotes the α quantile of the argument. Numerical experience indicates that $\alpha = 99\%$ is an adequate value, as this choice attenuates contributions from parameter regions possessing negligible likelihood mass.

Step 9: Adaptive selection of new training point. For each candidate parameter vector $\mathbf{x}_c^{(j)}$ in $\mathcal{C}_{\mathbf{x}}$, the following modified version of the acquisition function in Eq. 31 is evaluated:

$$\widehat{a}\left(\mathbf{x}_c^{(j)}\right) = \frac{1}{N}\sum_{i=1}^{N} \xi_3\left(\mathbf{x}_c^{(j)}, \boldsymbol{\sigma}^{(i)}\right) w\left(\xi_1\left(\mathbf{x}_c^{(j)}, \boldsymbol{\sigma}^{(i)}\right)\right), \quad j = 1, \ldots, N_C \tag{41}$$

where $w(\cdot)$ is the relevance weight in Eq. 30. Note that in the acquisition function in Eq. 41, the prior distribution $p_{\mathbf{X}}(\mathbf{x})$ present in Eq. 31 is omitted. The reason is that the next candidate point is selected out of the candidate sample pool $\mathcal{C}_{\mathbf{x}}$ that has been generated via Monte Carlo sampling according to $p_{\mathbf{X}}(\mathbf{x})$. Hence, the effect of the prior is already captured in that candidate sample pool. Indeed, the next training point is selected as

$$\mathbf{x}_{\mathrm{new}} = \arg\max_{j=1,\ldots,N_C} \widehat{a}(\mathbf{x}_c^{(j)}) \tag{42}$$

and the exact model $\mathbf{f}$ is evaluated at that new training point. Both $\mathbf{x}_{\mathrm{new}}$ and $\mathbf{f}$ $(\mathbf{x}_{\mathrm{new}})$ are added to the training data set $\mathcal{D}_t$, and $\mathbf{x}_{\mathrm{new}}$ is removed from $\mathcal{C}_{\mathbf{x}}$ and the number of remaining candidate points is updated as $N_C \leftarrow N_C - 1$. Then, the execution of the algorithm returns to step 4.

Step 10: Posterior inference. Once the surrogate and Monte Carlo convergence criteria are satisfied, posterior inference is performed on the basis of the expected likelihood values computed at the prior sample pairs $\left(\mathbf{x}^{(i)}, \boldsymbol{\sigma}^{(i)}\right), i = 1, \ldots, N$. For that purpose, importance weights are defined as

$$\omega^{(i)} = \frac{\xi_1\left(\mathbf{x}^{(i)}, \boldsymbol{\sigma}^{(i)}\right)}{\sum_{j=1}^{N} \xi_1\left(\mathbf{x}^{(j)}, \boldsymbol{\sigma}^{(j)}\right)}. \tag{43}$$

A set of N_p posterior samples is then generated by means of a sampling–importance–resampling procedure applied to the weighted population $\{\left(\mathbf{x}^{(i)}, \boldsymbol{\sigma}^{(i)}\right), \omega^{(i)}\}_{i=1}^{N}$. The resulting resampled pairs

$$\left(\mathbf{x}^{(r)}, \boldsymbol{\sigma}^{(r)}\right), \qquad r = 1, \ldots, N_p,$$

constitute approximate realizations from the joint posterior density of the model parameters and the noise standard deviations. Consequently, the same posterior sample set can be used for the estimation of statistical quantities associated with both $\mathbf{x}$ and $\boldsymbol{\sigma}$, including posterior means, standard deviations, credible intervals, or marginal probability densities. In addition, graphical representations such as histograms or kernel density estimates may be constructed directly from the resampled population.

In practice, the number of posterior samples N_p is selected significantly smaller than the prior sample size N, that is, $N_p \ll N$, since the posterior sample set is obtained from a discrete weighted approximation supported on the original Monte Carlo population. A useful indicator for selecting an appropriate value of N_p is the effective sample size,

$$N_{\mathrm{eff}} = \frac{1}{\sum_{i=1}^{N} \left(\omega^{(i)}\right)^2}, \tag{44}$$

which quantifies the number of effectively independent samples contained in the weighted population. In general, N_p should not exceed, in order of magnitude, the value of N_{eff}, since otherwise the resampled set may contain an excessive number of replicated points and may provide a poor representation of the posterior distribution.

6 Example

6.1 Description

This example considers the three-degree-of-freedom (DOF) mass–spring system shown in Fig. 1 (see, e.g., [36]). The system is intentionally simple, since such a choice allows the application of brute-force Monte Carlo simulation without further approximations. In this way, highly accurate reference results can be obtained, and the example is therefore well suited for validation of the proposed methodology.

The nominal masses are $m_1 = m_2 = m_3 = 1.0$ kg, and the nominal stiffness values are $k_1 = k_2 = k_5 = 1.5$ N/m, $k_3 = k_4 = 1.0$ N/m, and $k_6 = 3.0$ N/m. The uncertain parameter vector is defined as

$$\mathbf{x} = \left[k_1\ k_2\ k_5\right]^\top, \tag{45}$$

while all remaining parameters are kept fixed at their nominal values. It is noted that the values of k_1, k_2, and k_5 are specified here only for definition of the reference model. In the subsequent inverse problem, these quantities are assumed to be unknown and are to be identified on the basis of measurement data.

The response vector collects the first three natural frequencies, that is,

$$\mathbf{g}(\mathbf{x}) = \left[g_1(\mathbf{x})\ g_2(\mathbf{x})\ g_3(\mathbf{x})\right]^\top, \tag{46}$$

where $g_i(\mathbf{x})$, $i = 1, 2, 3$, denotes the i-th natural frequency of the system. Since natural frequencies are strictly positive quantities, the surrogate model is not constructed directly for $\mathbf{g}(\mathbf{x})$. Instead, a logarithmic transformation is introduced,

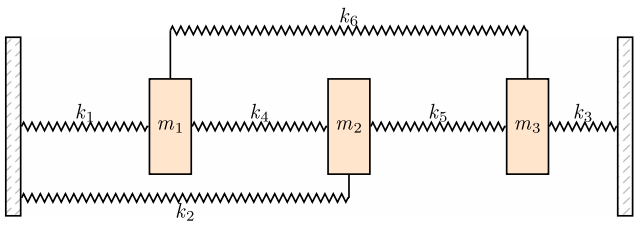

Fig. 1 Three-DOF mass–spring system. The uncertain parameters are k_1, k_2, and k_5, whereas all remaining parameters are fixed at their nominal values

which leads to

$$\mathbf{f}(\mathbf{x}) := \log \mathbf{g}(\mathbf{x}) = \left[\log g_1(\mathbf{x})\ \log g_2(\mathbf{x})\ \log g_3(\mathbf{x})\right]^\top, \tag{47}$$

where the logarithm is applied componentwise. In this way, Gaussian process regression is performed in an unconstrained space, while positivity of the predicted natural frequencies is preserved after back-transformation.

6.2 Generation of synthetic observations

For validation purposes, synthetic observations are generated by considering the reference parameter vector

$$\mathbf{x}^* = \left[1.5\ 1.5\ 1.5\right]^\top \text{N/m}, \tag{48}$$

which corresponds to the nominal values of the uncertain stiffness coefficients k_1, k_2, and k_5. A total of $n_o = 10$ independent replicate experiments are considered.

In accordance with the observation model introduced previously, the synthetic observations are generated in the transformed response space as

$$\mathbf{y}_o^{(j)} = \mathbf{f}(\mathbf{x}^*) + \boldsymbol{\varepsilon}^{(j)}, \qquad j = 1, \ldots, n_o, \tag{49}$$

where $\boldsymbol{\varepsilon}^{(j)}$ denotes an independent realization of a Gaussian random vector with zero mean and covariance matrix

$$\boldsymbol{\Sigma}_\varepsilon = \sigma^2 \mathbf{I}. \tag{50}$$

Hence, the three components of the measurement error are assumed to be mutually independent and to possess a common standard deviation σ. A numerical value of $\sigma^* = 0.05$ is considered in order to generate the synthetic observations.

It is noted that the additive Gaussian error model is introduced for the logarithm of the natural frequencies, and not for the natural frequencies themselves. Equivalently, in the original response space this corresponds to a multiplicative lognormal noise model. Such a choice is physically meaningful, since the measured natural frequencies remain strictly positive and, therefore, nonphysical negative values cannot arise.

6.3 Prior distributions

Prior information is specified for the unknown stiffness parameters collected in the vector $\mathbf{x}$ as well as for the observation-noise standard deviation σ. In the absence of detailed prior knowledge on the stiffness values, independent uniform prior distributions are assumed, that is,

$$x_i \sim \mathcal{U}(1\ \text{N/m}, 3\ \text{N/m}), \qquad i = 1, 2, 3, \tag{51}$$

where $x_1 = k_1$, $x_2 = k_2$, and $x_3 = k_5$. Regarding the observation model, only a single noise standard deviation is considered, which is denoted by σ. To reflect limited prior knowledge while ensuring positivity, a truncated half–Student–t prior with $\nu = 4$ degrees of freedom is adopted,

$$\sigma \sim \text{Half-}t_\nu(0, s), \qquad \nu = 4, \qquad \sigma \geq \sigma_{\min}, \tag{52}$$

where s denotes the scale parameter and $\sigma_{\min}$ is a prescribed lower bound. Both quantities s and $\sigma_{\min}$ are selected on the basis of the available observation data. Let $\mathbf{y}_o^{(j)} = [\, y_{o,1}^{(j)} \;\; y_{o,2}^{(j)} \;\; y_{o,3}^{(j)} \,]^\top$, $j = 1, \ldots, n_o$, denote the observation vectors. An empirical measure of dispersion is first computed as

$$\hat{\sigma}_{y_o} = \sqrt{\frac{1}{3n_o - 3} \sum_{j=1}^{n_o} \sum_{i=1}^{3} \left(y_{o,i}^{(j)} - \bar{y}_{o,i} \right)^2}, \tag{53}$$

where $\bar{y}_{o,i}$ denotes the sample mean of the ith observation component.

The scale parameter of the half–Student–t prior is then defined as

$$s = \hat{\sigma}_{y_o}. \tag{54}$$

Similarly, the lower truncation value is chosen as

$$\sigma_{\min} = c_{\min}\, \hat{\sigma}_{y_o}, \tag{55}$$

with $0 < c_{\min} \ll 1$. In this way, the prior for σ remains weakly informative and adapted to the scale of the observations through an empirical Bayes step, while unrealistically small noise levels are excluded. In this example, $c_{\min}$ is selected as 10^{-3}. The parameter σ is inferred jointly with the stiffness parameters within the Bayesian updating procedure.

6.4 Surrogate construction

The Gaussian process surrogate of the logarithm of the natural frequencies is constructed by selecting a prior mean equal to a constant function,

$$\mu_k(\mathbf{x}) = \beta_k, \tag{56}$$

where β_k is an unknown coefficient. Furthermore, the covariance structure is described by a squared exponential kernel

$$\kappa_k(\mathbf{x}, \mathbf{x}') = s_k^2 \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{x}')^\top \boldsymbol{\Lambda}_k^{-1} (\mathbf{x} - \mathbf{x}') \right), \tag{57}$$

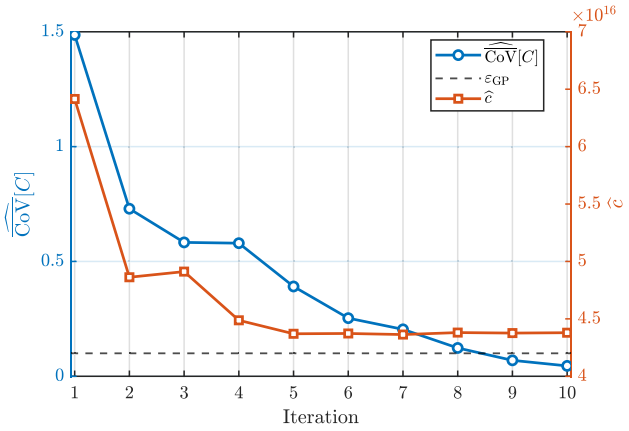

Fig. 2 Evolution of the upper bound of the coefficient of variation of the evidence $\widehat{\mathrm{CoV}}[C]$ and mean value of the evidence $\widehat{c}$ with respect to iteration number

where s_k^2 denotes the process variance and $\boldsymbol{\Lambda}_k = \mathrm{diag}(\ell_{k,1}^2, \ldots, \ell_{k,n_x}^2)$ contains the squared correlation lengths associated with each input dimension. The training data set $\mathcal{D}_t$ is initialized with $n_t = 10$ samples generated using Latin Hypercube sampling with respect to the prior distribution. Nevertheless, the training data could be initialized with any appropriate scheme such as, e.g., the Sobol' or Hammersley sequence.

6.5 Posterior inference results

For the practical implementation of the proposed approach, the following variables are chosen:

- The expected value and upper bound of the variance of the evidence are calculated using sample sets with $N = 10^7$.
- A total of $N_C = 2 \times 10^5$ candidate samples are generated to identify points for enriching the training data set.
- The convergence tolerances $\varepsilon_{\mathrm{GP}}$ and $\varepsilon_{\mathrm{MC}}$ are set equal to 10%.

The approach proposed in this work is implemented considering these variables. The results obtained are described in detail below.

In Fig. 2, it is observed that, at the first iteration (where the Gaussian process regression has been trained with $n_t = 10$ samples), the upper bound of the coefficient of variation of the evidence due to the approximation error of the Gaussian process is around 150%. As this value exceeds $\varepsilon_{\mathrm{GP}}$, the acquisition function is used to locate a new candidate design to enrich the training data set $\mathcal{D}_t$. As observed from Fig. 2, a total of 9 additional candidate designs are included in the training data set in the subsequent iterations. It is observed that the upper bound of the coefficient of variation drops

steadily, indicating that the proposed acquisition function is capable of identifying candidate samples that indeed help in improving the quality of the surrogate model. The enrichment process of the surrogate ends when the convergence criterion is met twice in a row. Regarding the evolution of the expected value of the evidence $\mathbb{E}[C]$ (which is denoted for simplicity in Fig. 2 as $\widehat{c}$), it is observed that its numerical value is fairly stable after the first few iterations.

Once the Gaussian process surrogate meets its convergence criterion, the next step is to generate posterior samples of the parameter vector $\mathbf{x}$ and the noise parameter σ. For this purpose, a total of $N_p = 2000$ samples are generated. This number seems appropriate, as the effective sample size is about $N_{\text{eff}} \approx 60 \times 10^3$. The results obtained are summarized in Table 1 under the label proposed approach (GPR). It is observed that the posterior mean for each of the four parameters being identified is close to the actual value with which the synthetic data was generated. Nevertheless, the standard deviations exhibit relatively high values while the 90% credible intervals (CI) are rather wide. Therefore, to verify the accuracy of the obtained results, posterior sampling is repeated, this time discarding the GPR surrogate and calculating the natural frequencies exactly. This leads to the results labeled as brute-force Monte Carlo in Table 1. It is observed that the results obtained with this reference approach are almost identical to those obtained with the proposed approach.

In addition to the results presented in Tables 1 and 2 presents the mean, standard deviation and 90% credible intervals for posterior natural frequencies of the system. These frequencies have been generated using the exact model (that is, solution of the associated eigenvalue problem) considering the posterior samples generated with the proposed approach and brute-force Monte Carlo. As it can be seen from Table 2, the posterior predictive summaries of the first three natural frequencies obtained with the proposed approach are in very close agreement with those delivered by brute-force Monte Carlo simulation. This indicates that the surrogate-based updating procedure preserves the predictive uncertainty in the response space.

To gain additional insight into the Bayesian updating results obtained with the proposed approach, normalized histograms and scatter plots of the posterior samples are depicted in Fig. 3. From the scatter plots, it is observed that there are strong interactions between the different values of the stiffnesses. This means that different combinations of stiffness values can explain the available noisy observations of the natural frequencies. This explains the relatively large credible intervals observed in Table 1.

To verify the observations derived from Fig. 3, the following experiment is conducted. All variables of the example are kept fixed except for:

- The number of observations, which is increased to $n_o = 20$.
- The parameter σ used to generate the synthetic observations, which is reduced to $\sigma^* = 0.02$.

This selection leads to an identification problem with more information. The results obtained are displayed in Fig. 4. From this figure, it is evident that the normalized histograms of the posterior distributions and the scatter plots are now much narrower. This is a natural consequence of the additional available information, which allows one to obtain a more informative posterior distribution.

As an additional investigation, a stress-test analysis is conducted in order to assess the behaviour of the proposed methodology under a severely restricted computational budget. For this purpose, the original configuration comprising $n_o = 10$ observations and a reference noise level equal to $\sigma = 0.05$ was considered again. However, the total number of exact model evaluations employed for the construction of the surrogate model was limited to only six. Two different strate-

Table 1 Posterior summaries of stiffness parameters k_1, k_2, and k_5 (in N/m) and noise level σ obtained using the proposed GPR–based approach and brute–force Monte Carlo simulation. The table also reports the total number of forward model evaluations required by each method as well as the estimator for the expected value of the evidence

	Proposed approach (GPR)			Brute–force Monte Carlo		
Parameter	Mean	Std. dev.	90% CI	Mean	Std. dev.	90% CI
k_1	1.410	0.276	[1.041, 1.924]	1.422	0.276	[1.046, 1.924]
k_2	1.594	0.324	[1.090, 2.145]	1.586	0.319	[1.082, 2.128]
k_5	1.466	0.179	[1.158, 1.754]	1.472	0.177	[1.174, 1.753]
σ	0.0584	0.0080	[0.0469, 0.0728]	0.0582	0.0079	[0.0469, 0.0721]
Model evaluations		19 (10 + 9)			10^7	
$\widehat{\mathbb{E}}_{\text{GP}}[C]$		4.4×10^{16}			4.4×10^{16}	

Table 2 Posterior predictive summaries of the first three natural frequencies (in Hz) obtained by propagating posterior samples generated by the proposed GPR–based Bayesian updating approach and brute–force Monte Carlo simulation through the exact forward model

Frequency	Proposed approach (GPR)			Brute–force Monte Carlo		
	Mean	Std. dev.	90% CI	Mean	Std. dev.	90% CI
g_1	0.1821	0.0033	[0.1769, 0.1875]	0.1822	0.0033	[0.1769, 0.1876]
g_2	0.3612	0.0065	[0.3506, 0.3721]	0.3613	0.0064	[0.3509, 0.3720]
g_3	0.4632	0.0057	[0.4542, 0.4727]	0.4636	0.0056	[0.4550, 0.4729]

gies were investigated. First, the proposed adaptive approach was employed using three initial training samples followed by three additional samples selected through active learning. Second, a non-adaptive strategy without active learning was considered, where six training samples were selected directly from the prior distribution. The corresponding posterior means and standard deviations are reported in Table 3, together with their relative errors with respect to the brute-force Monte Carlo results reported in Table 1. It is observed that the proposed approach with active learning still provides posterior means and standard deviations that remain reasonably close to the reference results obtained with brute-force Monte Carlo simulation. In particular, the relative errors associated with the posterior summaries are generally smaller than those obtained with the corresponding non-adaptive approach, especially for parameters k_1, k_5 and σ. Furthermore, although the upper-bound-based coefficients of variation associated with the surrogate-induced uncertainty in the evidence remain relatively large due to the extremely limited number of exact model evaluations employed in this stress-test setting, the adaptive strategy yields a substantially smaller value than the corresponding non-adaptive approach.

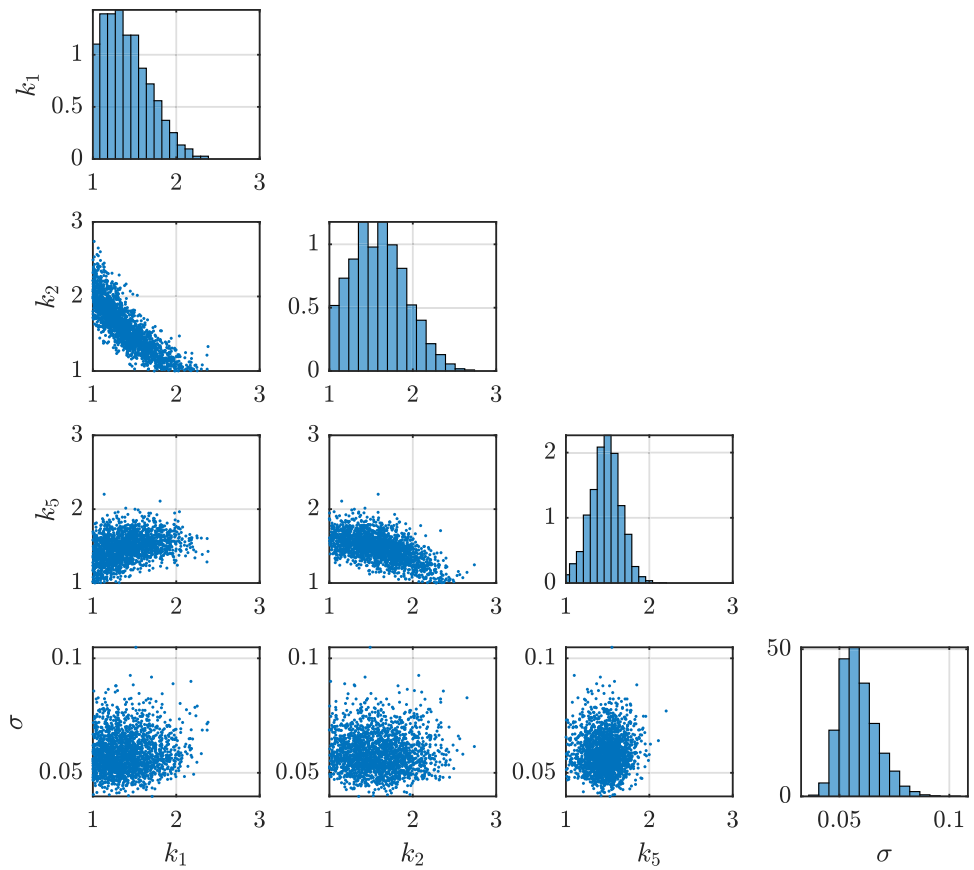

Fig. 3 Posterior samples of $\mathbf{x} = [k_1, k_2, k_5]^T$ N/m and σ. Plots on the diagonal display normalized histograms, plots outside the diagonal display scatter

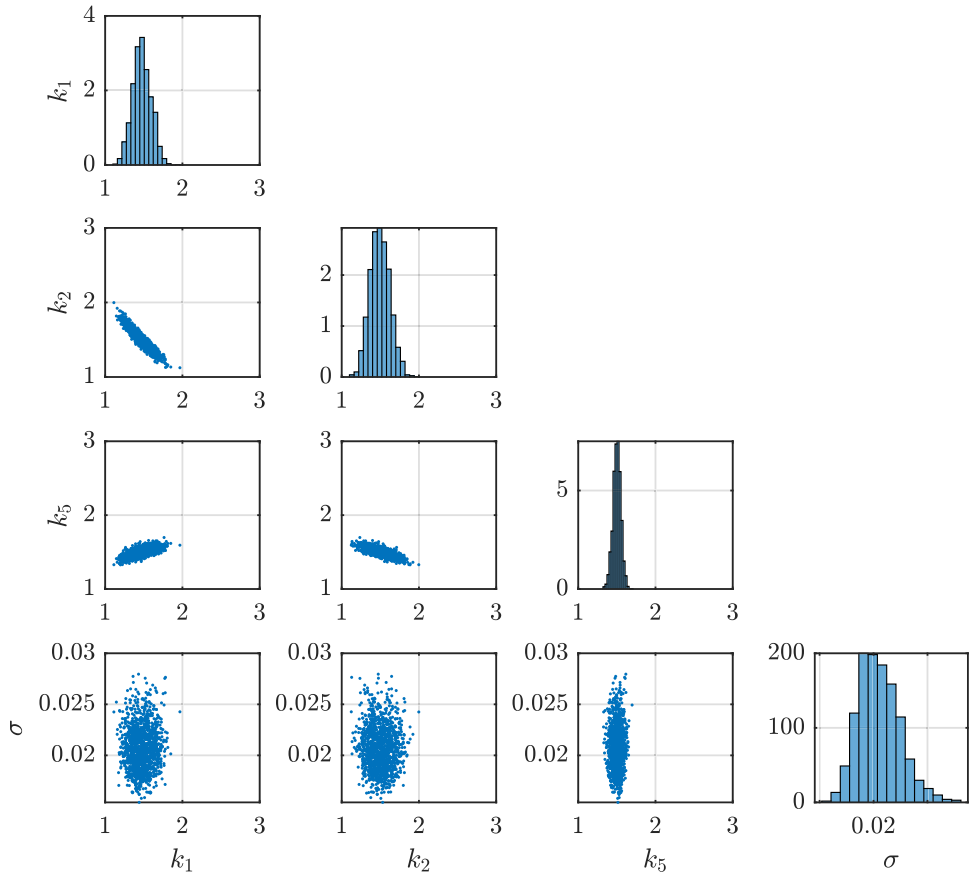

Fig. 4 Posterior samples of $\mathbf{x} = [k_1, k_2, k_5]^T$ N/m and σ for the case where $n_o = 20$ and $\sigma = 0.02$. Plots on the diagonal display normalized histograms, plots outside the diagonal display scatter

This suggests that the active learning strategy improves the accuracy of the posterior approximation and provides a more reliable control of the surrogate-induced uncertainty within the Bayesian updating procedure.

7 Conclusions and outlook

This contribution has examined the incorporation of the epistemic uncertainty associated with Gaussian process surrogate models into a Bayesian updating framework. The main idea consists in explicitly accounting for the surrogate-induced approximation error when performing inference on uncertain model parameters. To this end, the quality of the Gaussian process approximation has been controlled by monitoring the accuracy of the estimated model evidence, which serves as a global indicator of the adequacy of the surrogate representation within the Bayesian setting.

The numerical results suggest that the proposed strategy constitutes an appropriate and practically viable approach.

Table 3 Posterior summaries obtained with the proposed approach and with the corresponding non-adaptive GPR approach using the same number of exact model evaluations. Relative errors reported in parentheses are computed with respect to the brute-force Monte Carlo posterior summaries reported in Table 1

Parameter	Proposed approach		Without active learning	
	Mean	Std. dev.	Mean	Std. dev.
k_1	1.411 (0.77%)	0.272 (1.45%)	1.369 (3.73%)	0.241 (12.68%)
k_2	1.544 (2.65%)	0.329 (3.13%)	1.559 (1.70%)	0.339 (6.27%)
k_5	1.465 (0.48%)	0.187 (5.65%)	1.396 (5.16%)	0.218 (23.16%)
σ	0.0581 (0.17%)	0.0079 (0.00%)	0.0583 (0.17%)	0.0080 (1.27%)
Model evaluations	6 (3 + 3)		6	
$\widehat{\mathbb{E}}_{\mathrm{GP}}[C]$	6.9×10^{16}		5.7×10^{16}	
$\widehat{\mathrm{CoV}}[C]$	341%		691%	

In the considered example, accurate characterization of the posterior distribution was achieved using only a limited number of training samples for constructing the surrogate model. These findings support the general message of the present work, namely that the use of surrogate models in Bayesian updating inevitably introduces additional uncertainty, which should be explicitly propagated and controlled within the inference procedure.

From a methodological perspective, the proposed formulation can be regarded as a step towards a fully Bayesian treatment of surrogate-based inference. However, the approach is not yet fully Bayesian in all aspects. In the present study, the hyperparameters of the Gaussian process model have been determined by means of maximum likelihood estimation. A more rigorous treatment would require the consideration of prior distributions for these hyperparameters and their joint inference together with the physical parameters of interest. Despite this limitation, the resulting framework remains attractive in practical applications, since the uncertainty associated with the Gaussian process prediction can be analytically marginalized in the evaluation of posterior expectations, while conservative upper bounds can be derived for the variance of the model evidence.

Several directions for future research can be identified. First, independent Gaussian process models have been employed for the approximation of the individual response components. Although this assumption simplifies the surrogate construction and remains suitable for response vectors possessing moderate dimension, it may become inefficient in situations where the number of response components n_y is large, as it may occur, e.g., for time-dependent response signals. The development of multi-output or co-Kriging surrogate models therefore represents a natural extension, which may lead to improved predictive performance and a more consistent representation of the underlying system behaviour. Alternatively, dimensionality-reduction strategies based on principal component analysis could also be incorporated before surrogate construction, thereby enabling an efficient treatment of high-dimensional response vectors (see, e.g., [16]). Second, plain Monte Carlo simulation has been adopted for the estimation of the model evidence as well as for sampling from the posterior distribution. While the computational cost of evaluating the surrogate model is typically low, the efficiency of crude Monte Carlo methods deteriorates as the dimension of the model parameter space increases, particularly due to the concentration of the posterior distribution in localized regions of the parameter space possessing low prior probability. In view of high-dimensional identification problems, the integration of more advanced simulation techniques, such as transitional Markov chain Monte Carlo or related sequential sampling strategies, appears to be a promising research direction (see, e.g., [37–39]). Finally, the adaptive construction of the surrogate model has been based on a comparatively simple acquisition strategy relying on Monte Carlo evaluation of candidate points. Although such an approach is adequate for problems of moderate dimensionality, its efficiency may also deteriorate with increasing dimension of the model parameter space, as regions possessing high acquisition value become progressively more difficult to identify through random sampling. Future work should therefore investigate the use of dedicated optimization algorithms for the solution of the associated acquisition problem, with the aim of improving the robustness and scalability of the overall methodology.

It is envisioned that the developments outlined above may further enhance the applicability of surrogate-based Bayesian updating in complex engineering systems, where accurate and computationally efficient treatment of multiple sources of uncertainty is of primary importance.

Appendix A: Derivation of formulas associated with likelihood and evidence

A.1 Derivation of likelihood

Under the assumptions listed in Section 2.3, the likelihood is given by

$$\mathcal{L}(\mathbf{x},\boldsymbol{\sigma}\,|\,\mathcal{D}_o)=(2\pi)^{-\frac{n_o n_y}{2}}\left(\prod_{k=1}^{n_y}\sigma_k^{-n_o}\right)\exp\left(-\frac{1}{2}\sum_{j=1}^{n_o}\sum_{k=1}^{n_y}\frac{\left(y_{o,k}^{(j)}-f_k(\mathbf{x})\right)^2}{\sigma_k^2}\right), \tag{58}$$

where $y_{o,k}^{(j)}$ denotes the k-th component of the j-th observation vector $\mathbf{y}_o^{(j)}$, and $f_k(\mathbf{x})$ denotes the k-th component of the forward model response $\mathbf{f}(\mathbf{x})$. It is convenient to rewrite the likelihood in terms of sufficient statistics of the observation sample. For each response component $k = 1, \ldots, n_y$, let

$$\bar{y}_{o,k}=\frac{1}{n_o}\sum_{j=1}^{n_o}y_{o,k}^{(j)},\qquad S_k=\sum_{j=1}^{n_o}\left(y_{o,k}^{(j)}-\bar{y}_{o,k}\right)^2. \tag{59}$$

Then, the standard identity

$$\sum_{j=1}^{n_o}\left(y_{o,k}^{(j)}-f_k(\mathbf{x})\right)^2=S_k+n_o\left(\bar{y}_{o,k}-f_k(\mathbf{x})\right)^2 \tag{60}$$

holds. By substitution of Eq. 60 into Eq. 58, the likelihood can be factorized with respect to the response components as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \,|\, \mathcal{D}_o) = \prod_{k=1}^{n_y} (2\pi\sigma_k^2)^{-n_o/2} \exp\left(-\frac{S_k}{2\sigma_k^2}\right) \exp\left(-\frac{n_o}{2\sigma_k^2}\left(\bar{y}_{o,k} - f_k(\mathbf{x})\right)^2\right). \tag{61}$$

Since

$$\frac{n_o}{2\sigma_k^2}\left(\bar{y}_{o,k} - f_k(\mathbf{x})\right)^2 = \frac{\left(\bar{y}_{o,k} - f_k(\mathbf{x})\right)^2}{2(\sigma_k^2/n_o)},$$

each factor in Eq. 61 can also be expressed in terms of the one-dimensional Gaussian probability density function introduced, leading to:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o) = \prod_{k=1}^{n_y} A_k(\sigma_k)\, \varphi\left(f_k(\mathbf{x}) \mid \bar{y}_{o,k}, \sigma_k^2/n_o\right), \tag{62}$$

where

$$A_k(\sigma_k) = (2\pi\sigma_k^2)^{-n_o/2} \exp\left(-\frac{S_k}{2\sigma_k^2}\right) \sqrt{\frac{2\pi\sigma_k^2}{n_o}}, \tag{63}$$

and where $\varphi(\cdot \mid a, b^2)$ is the Gaussian probability density function with mean a and variance b^2. Equation 62 shows that, for each response component, the information contained in the repeated observations is conveyed through the sample mean $\bar{y}_{o,k}$ and the sample scatter S_k. In this way, the likelihood becomes proportional to a Gaussian density evaluated at the model response $f_k(\mathbf{x})$, with variance σ_k^2/n_o. This representation is particularly convenient for the purposes of this work.

A.2 Expectation of the likelihood under the Gaussian process posterior

This appendix presents the derivation of the expected likelihood with respect to the Gaussian process posterior predictive distribution. The starting point is the likelihood in Eq. 7, namely

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o) = \prod_{k=1}^{n_y} A_k(\sigma_k)\, \varphi\Big(f_k(\mathbf{x}) \,\Big|\, \bar{y}_{o,k}, \sigma_k^2/n_o\Big). \tag{64}$$

Now the exact model $f_k(\mathbf{x})$ is replaced by its Gaussian counterpart $F_k(\mathbf{x})$. Recalling the posterior predictive distribution of its k-th response component $F_k(\mathbf{x})$ conditional on the training data $\mathcal{D}_t$ is Gaussian, that is, $F_k(\mathbf{x}) \mid \mathcal{D}_t \sim \mathcal{N}\Big(\mu_{t,k}(\mathbf{x}), s_{t,k}^2(\mathbf{x})\Big)$, the expectation of the likelihood with respect to the Gaussian process posterior is therefore:

$$\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})] = \prod_{k=1}^{n_y} A_k(\sigma_k) \int \varphi\Big(F_k \,\Big|\, \bar{y}_{o,k}, \sigma_k^2/n_o\Big)\, \varphi\Big(F_k \,\Big|\, \mu_{t,k}(\mathbf{x}), s_{t,k}^2(\mathbf{x})\Big)\, \mathrm{d}F_k, \tag{65}$$

where independence of the predictive components has been used. The remaining integral corresponds to the convolution of two one-dimensional Gaussian probability density functions. Using the identity of the Gaussian density and completing the square with respect to f_k, one obtains (see, e.g., [38, 40])

$$\int \varphi\Big(F_k \,\Big|\, \bar{y}_{o,k}, \sigma_k^2/n_o\Big)\, \varphi\Big(F_k \,\Big|\, \mu_{t,k}(\mathbf{x}), s_{t,k}^2(\mathbf{x})\Big)\, \mathrm{d}F_k = \varphi\Big(\mu_{t,k}(\mathbf{x}) \,\Big|\, \bar{y}_{o,k}, \sigma_k^2/n_o + s_{t,k}^2(\mathbf{x})\Big). \tag{66}$$

Substitution of the last expression into Eq. 65 finally leads to

$$\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})] = \prod_{k=1}^{n_y} A_k(\sigma_k)\, \varphi\Big(\mu_{t,k}(\mathbf{x}) \,\Big|\, \bar{y}_{o,k}, \sigma_k^2/n_o + s_{t,k}^2(\mathbf{x})\Big), \tag{67}$$

which corresponds to Eq. 19.

A.3 Expectation of the square of the evidence under the Gaussian process posterior

The expected value of the square of the evidence with respect to the Gaussian process surrogate is

$$\mathbb{E}_{\mathrm{GP}}[C^2] = \mathbb{E}_{\mathrm{GP}}\Bigg[\int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}} \mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})\, p_{\mathbf{X}}(\mathbf{x}) p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\sigma} \times \int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}} \mathcal{L}(\mathbf{x}', \boldsymbol{\sigma}' \mid \mathcal{D}_o, \mathbf{F}')\, p_{\mathbf{X}}(\mathbf{x}') p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}')\, \mathrm{d}\mathbf{x}'\, \mathrm{d}\boldsymbol{\sigma}'\Bigg]. \tag{68}$$

where $\mathbf{F}' = \mathbf{F}(\mathbf{x}')$. Since the integrand is non-negative, Tonelli's theorem allows the interchange of expectation and integration, which yields

$$\mathbb{E}_{\mathrm{GP}}[C^2] = \int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}}\int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}} \mathbb{E}_{\mathrm{GP}}\left[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})\, \mathcal{L}(\mathbf{x}', \boldsymbol{\sigma}' \mid \mathcal{D}_o, \mathbf{F}')\right] \times p_{\mathbf{X}}(\mathbf{x}) p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}) p_{\mathbf{X}}(\mathbf{x}') p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}')\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\sigma}\, \mathrm{d}\mathbf{x}'\, \mathrm{d}\boldsymbol{\sigma}'. \tag{69}$$

From this last equation, the challenge is to compute the expectation of likelihoods $\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})$ and $\mathcal{L}(\mathbf{x}', \boldsymbol{\sigma}' \mid \mathcal{D}_o, \mathbf{F}')$ with respect to the Gaussian process. To solve this

expectation, it is recalled from Eq. 64 that the likelihood can be represented as:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F}) = \prod_{k=1}^{n_y} \mathcal{L}_k(\mathbf{x}, \sigma_k \mid \mathcal{D}_o, F_k), \tag{70}$$

with

$$\mathcal{L}_k(\mathbf{x}, \sigma_k \mid \mathcal{D}_o, F_k) = A_k(\sigma_k)\, \varphi\Big(F_k(\mathbf{x}) \,\Big|\, \bar{y}_{o,k}, \sigma_k^2/n_o\Big). \tag{71}$$

Now consider two parameter points $\mathbf{x}$ and $\mathbf{x}'$. For each response component k, the Gaussian process posterior implies that the random vector

$$\begin{bmatrix} F_k(\mathbf{x}) \\ F_k(\mathbf{x}') \end{bmatrix} \Bigg| \mathcal{D}_t \sim \mathcal{N}\big(\boldsymbol{\mu}_{t,k}, \boldsymbol{\Sigma}_{t,k}\big), \tag{72}$$

with mean vector

$$\boldsymbol{\mu}_{t,k} = \begin{bmatrix} \mu_{t,k}(\mathbf{x}) \\ \mu_{t,k}(\mathbf{x}') \end{bmatrix}, \tag{73}$$

and covariance matrix

$$\boldsymbol{\Sigma}_{t,k} = \begin{bmatrix} s_{t,k}^2(\mathbf{x}) & \kappa_{t,k}(\mathbf{x}, \mathbf{x}') \\ \kappa_{t,k}(\mathbf{x}, \mathbf{x}') & s_{t,k}^2(\mathbf{x}') \end{bmatrix}. \tag{74}$$

Using Eq. 71, the mixed Gaussian process moment of the k-th likelihood factor can be written as

$$\begin{aligned} &\mathbb{E}_{\mathrm{GP}}\big[\mathcal{L}_k(\mathbf{x}, \sigma_k \mid \mathcal{D}_o, F_k)\, \mathcal{L}_k(\mathbf{x}', \sigma_k' \mid \mathcal{D}_o, F_k)\big] \\ &\quad = A_k(\sigma_k)\, A_k(\sigma_k')\, \mathbb{E}_{\mathrm{GP}}\Big[\varphi\big(F_k(\mathbf{x}) \mid \bar{y}_{o,k}, \sigma_k^2/n_o\big)\, \varphi\Big(F_k(\mathbf{x}') \,\Big|\, \bar{y}_{o,k}, \sigma_k'^2/n_o\Big)\Big]. \end{aligned} \tag{75}$$

For convenience, introduce the auxiliary random vector

$$\mathbf{Z}_k = \begin{bmatrix} F_k(\mathbf{x}) \\ F_k(\mathbf{x}') \end{bmatrix}. \tag{76}$$

From Eq. 72, this vector follows a bivariate normal distribution conditioned on the training data set $\mathcal{D}_t$. Furthermore, the product of the two univariate Gaussian probability densities appearing in Eq. 75 can be associated with a bivariate normal density in the vector $\mathbf{Z}_k$. Indeed, defining

$$\mathbf{m}_k = \begin{bmatrix} \bar{y}_{o,k} \\ \bar{y}_{o,k} \end{bmatrix}, \qquad \boldsymbol{\Sigma}_{\varepsilon,k} = \begin{bmatrix} \sigma_k^2/n_o & 0 \\ 0 & \sigma_k'^2/n_o \end{bmatrix}, \tag{77}$$

it follows directly from the definition of the multivariate normal density that

$$\varphi\Big(F_k(\mathbf{x}) \mid \bar{y}_{o,k}, \sigma_k^2/n_o\Big)\, \varphi\Big(F_k(\mathbf{x}') \mid \bar{y}_{o,k}, \sigma_k'^2/n_o\Big) = \varphi_2\big(\mathbf{Z}_k \mid \mathbf{m}_k, \boldsymbol{\Sigma}_{\varepsilon,k}\big), \tag{78}$$

where $\varphi_2\left(\cdot \mid \mathbf{m}_k, \boldsymbol{\Sigma}_{\varepsilon,k}\right)$ is the bivariate Gaussian probability density function with mean $\mathbf{m}_k$ and covariance $\boldsymbol{\Sigma}_{\varepsilon,k}$. Consequently, the expectation in Eq. 75 can be expressed as the integral of the product of two bivariate normal densities, namely

$$A_k(\sigma_k)\, A_k(\sigma_k') \int_{\mathbb{R}^2} \varphi_2\big(\mathbf{z} \mid \mathbf{m}_k, \boldsymbol{\Sigma}_{\varepsilon,k}\big)\, \varphi_2\big(\mathbf{z} \mid \boldsymbol{\mu}_{t,k}, \boldsymbol{\Sigma}_{t,k}\big)\, \mathrm{d}\mathbf{z}. \tag{79}$$

Using the standard identity for the product-integral of Gaussian densities, it follows that the mixed Gaussian process moment admits the closed-form representation

$$\begin{aligned} &\mathbb{E}_{\mathrm{GP}}\big[\mathcal{L}_k(\mathbf{x}, \sigma_k \mid \mathcal{D}_o, F_k)\, \mathcal{L}_k(\mathbf{x}', \sigma_k' \mid \mathcal{D}_o, F_k)\big] \\ &\quad = A_k(\sigma_k)\, A_k(\sigma_k')\, \varphi_2\left(\begin{bmatrix} \bar{y}_{o,k} \\ \bar{y}_{o,k} \end{bmatrix} \;\middle|\; \boldsymbol{\mu}_{t,k},\ \tilde{\boldsymbol{\Sigma}}_{t,k}\right), \end{aligned} \tag{80}$$

where $\tilde{\boldsymbol{\Sigma}}_{t,k} = \boldsymbol{\Sigma}_{\varepsilon,k} + \boldsymbol{\Sigma}_{t,k}$. Hence, the sought expected value of the square of the evidence with respect to the Gaussian process is

$$\begin{aligned} \mathbb{E}_{\mathrm{GP}}[C^2] = \int_{\Omega_{\boldsymbol{\sigma}}} \int_{\Omega_{\mathbf{x}}} \int_{\Omega_{\boldsymbol{\sigma}}} \int_{\Omega_{\mathbf{x}}} \prod_{k=1}^{n_y} A_k(\sigma_k)\, A_k(\sigma_k')\, \varphi_2\left(\begin{bmatrix} \bar{y}_{o,k} \\ \bar{y}_{o,k} \end{bmatrix} \mid \boldsymbol{\mu}_{t,k},\ \tilde{\boldsymbol{\Sigma}}_{t,k}\right) \\ \times\, p_{\mathbf{X}}(\mathbf{x})\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, p_{\mathbf{X}}(\mathbf{x}')\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}')\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\sigma}\, \mathrm{d}\mathbf{x}'\, \mathrm{d}\boldsymbol{\sigma}'. \end{aligned} \tag{81}$$

A.4 Calculation of upper bound of variance of evidence

In order to derive a tractable surrogate-accuracy indicator, an upper bound for the variance of the evidence is introduced. Starting from Eq. 24

$$\mathrm{Var}_{\mathrm{GP}}[C] = \mathbb{E}_{\mathrm{GP}}[C^2] - \big(\mathbb{E}_{\mathrm{GP}}[C]\big)^2, \tag{82}$$

and substituting Eqs. 69 and 21 yields

$$\begin{aligned} \mathrm{Var}_{\mathrm{GP}}[C] = &\int_{\Omega_{\boldsymbol{\sigma}}} \int_{\Omega_{\mathbf{x}}} \int_{\Omega_{\boldsymbol{\sigma}}} \int_{\Omega_{\mathbf{x}}} \mathbb{E}_{\mathrm{GP}}\big[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})\, \mathcal{L}(\mathbf{x}', \boldsymbol{\sigma}' \mid \mathcal{D}_o, \mathbf{F}')\big] \\ &\times\, p_{\mathbf{X}}(\mathbf{x})\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, p_{\mathbf{X}}(\mathbf{x}')\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}')\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\sigma}\, \mathrm{d}\mathbf{x}'\, \mathrm{d}\boldsymbol{\sigma}' \\ &- \left(\int_{\Omega_{\boldsymbol{\sigma}}} \int_{\Omega_{\mathbf{x}}} \mathbb{E}_{\mathrm{GP}}\big[\mathcal{L}(\mathbf{x}, \boldsymbol{\sigma} \mid \mathcal{D}_o, \mathbf{F})\big]\, p_{\mathbf{X}}(\mathbf{x})\, p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\sigma}\right)^2. \end{aligned} \tag{83}$$

This last equation can be simplified to:

$$\mathrm{Var}_{\mathrm{GP}}[C] = \int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}}\int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}} \mathrm{Cov}_{\mathrm{GP}}\big[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma}\mid\mathcal{D}_o,\mathbf{F}),\ \mathcal{L}(\mathbf{x}',\boldsymbol{\sigma}'\mid\mathcal{D}_o,\mathbf{F}')\big] \times p_{\mathbf{X}}(\mathbf{x})p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})p_{\mathbf{X}}(\mathbf{x}')p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}')\,\mathrm{d}\mathbf{x}\,\mathrm{d}\boldsymbol{\sigma}\,\mathrm{d}\mathbf{x}'\,\mathrm{d}\boldsymbol{\sigma}', \tag{84}$$

where $\mathrm{Cov}_{\mathrm{GP}}[\cdot]$ denotes the covariance of the argument with respect to the Gaussian process surrogate. Applying the Cauchy-Schwarz inequality to bound that covariance from above yields

$$\mathrm{Var}_{\mathrm{GP}}[C] \leq \int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}}\int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}} \sqrt{\mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma}\mid\mathcal{D}_o,\mathbf{F})]} \times \sqrt{\mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x}',\boldsymbol{\sigma}'\mid\mathcal{D}_o,\mathbf{F}')]} \times p_{\mathbf{X}}(\mathbf{x})p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})p_{\mathbf{X}}(\mathbf{x}')p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}')\,\mathrm{d}\mathbf{x}\,\mathrm{d}\boldsymbol{\sigma}\,\mathrm{d}\mathbf{x}'\,\mathrm{d}\boldsymbol{\sigma}'. \tag{85}$$

By further manipulation, this expression leads to the bound (see, e.g., [41])

$$\mathrm{Var}_{\mathrm{GP}}[C] \leq \left(\int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}} \sqrt{\mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma}\mid\mathcal{D}_o,\mathbf{F})]}\, p_{\mathbf{X}}(\mathbf{x})p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\boldsymbol{\sigma}\right)^2. \tag{86}$$

Equation 86 already provides a meaningful bound, since it shows that the global variability of the evidence can be controlled through the point-wise variability of the likelihood induced by the surrogate model. In other words, Eq. 86 constitutes the main theoretical result required for relating the uncertainty of the evidence with respect to the Gaussian process surrogate. For practical implementation, however, it is convenient to introduce a further simplification. Indeed, by applying Jensen's inequality to the concave function $\varphi(u) = \sqrt{u}$ with respect to the probability measure $p_{\mathbf{X}}(\mathbf{x})p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\,\mathrm{d}\mathbf{x}\mathrm{d}\boldsymbol{\sigma}$, the square of the weighted integral in Eq. (86) can be bounded from above by the corresponding weighted integral of $\mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma}\mid\mathcal{D}_o,\mathbf{F})]$. This yields the upper bound $\overline{\mathrm{Var}}_{\mathrm{GP}}[C]$, namely

$$\mathrm{Var}_{\mathrm{GP}}[C] \leq \overline{\mathrm{Var}}_{\mathrm{GP}}[C] = \int_{\Omega_{\boldsymbol{\sigma}}}\int_{\Omega_{\mathbf{x}}} \mathrm{Var}_{\mathrm{GP}}\big[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma}\mid\mathcal{D}_o,\mathbf{F})\big]\, p_{\mathbf{X}}(\mathbf{x})p_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma})\,\mathrm{d}\mathbf{x}\mathrm{d}\boldsymbol{\sigma}. \tag{87}$$

Hence, the problem reduces to the evaluation of the point-wise Gaussian process variance of the likelihood,

$$\mathrm{Var}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma}\mid\mathcal{D}_o,\mathbf{F})] = \mathbb{E}_{\mathrm{GP}}\Big[\mathcal{L}^2(\mathbf{x},\boldsymbol{\sigma}\mid\mathcal{D}_o,\mathbf{F})\Big] - \left(\mathbb{E}_{\mathrm{GP}}[\mathcal{L}(\mathbf{x},\boldsymbol{\sigma}\mid\mathcal{D}_o,\mathbf{F})]\right)^2. \tag{88}$$

The expected value of the likelihood appearing in Eq. 88 has already been derived in closed form in Eq. 19. Therefore, it remains to evaluate the expected value of the square of the likelihood. Recalling Eq. 70, it is readily observed that,

$$\mathbb{E}_{\mathrm{GP}}\Big[\mathcal{L}^2(\mathbf{x},\boldsymbol{\sigma}\mid\mathcal{D}_o,\mathbf{F})\Big] = \prod_{k=1}^{n_y}\mathbb{E}_{\mathrm{GP}}\Big[\mathcal{L}_k^2(\mathbf{x},\sigma_k\mid\mathcal{D}_o,F_k)\Big]. \tag{89}$$

Now considering the definition of $\mathcal{L}_k$ in Eq. 71 and using the identity for the square of a Gaussian probability density, which is proportional to another Gaussian density with halved variance leads to:

$$\mathcal{L}_k^2(\mathbf{x},\sigma_k\mid\mathcal{D}_o,F_k) = A_k(\sigma_k)^2\sqrt{\frac{n_o}{4\pi\sigma_k^2}}\,\varphi\left(F_k(\mathbf{x})\,\middle|\,\bar{y}_{o,k},\ \frac{\sigma_k^2}{2n_o}\right). \tag{90}$$

Taking the expected value of this last Gaussian kernel with respect to the Gaussian process corresponds to solving the convolution of two one-dimensional Gaussian probability density functions (as performed in Eq. 66), which leads to the sought expression for the expected value of the square of the likelihood.

$$\mathbb{E}_{\mathrm{GP}}\big[\mathcal{L}_k^2(\mathbf{x},\sigma_k\mid\mathcal{D}_o,F_k)\big] = A_k(\sigma_k)^2\sqrt{\frac{n_o}{4\pi\sigma_k^2}}\,\varphi\left(\mu_{t,k}(\mathbf{x})\big|\bar{y}_{o,k},\ \frac{\sigma_k^2}{2n_o}+s_{t,k}^2(\mathbf{x})\right). \tag{91}$$

Equations 89 and 91, together with Eq. 19, provide a closed-form representation of the point-wise Gaussian process variance of the likelihood in Eq. 88. Substitution into Eq. 87 yields the explicit upper bound $\overline{\mathrm{Var}}_{\mathrm{GP}}[C]$ for $\mathrm{Var}_{\mathrm{GP}}[C]$ in which the surrogate uncertainty has been analytically marginalized.

Acknowledgements Sylvia Keßler gratefully acknowledges funding from dtec.bw – the Digitalization and Technology Research Center of the Bundeswehr. This research center is funded by the European Union – NextGenerationEU. Pengfei Wei gratefully acknowledges the financial support provided by the National Natural Science Foundation of China under Grant No. 52475164. Zhouzhou Song gratefully acknowledges the support of the Alexander von Humboldt Foundation through a post-doctoral fellowship. Chao Dang gratefully acknowledges funding from the German Research Foundation (DFG) under Grant No. 530326817. Matthias G.R. Faes further gratefully acknowledges support from the Alexander von Humboldt Foundation through the Henriette Herz Scouting Program.

Author Contributions M.V. wrote the original draft. C.D., P.L, Z.S, P.W., S.K., MF. edited the draft. All authors reviewed the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Kerschen G, Worden K, Vakakis AF, Golinval JC. Past, present and future of nonlinear system identification in structural dynamics. Mech Syst Signal Process. 2006;20(3):505–92. https://doi.org/10.1016/j.ymssp.2005.04.008.
2. Yuen K-V. Bayesian Methods for Structural Dynamics and Civil Engineering. John Wiley & Sons, Singapore; 2010. https://doi.org/10.1002/9780470824566.
3. Faes M, Broggi M, Patelli E, Govers Y, Mottershead J, Beer M, et al. A multivariate interval approach for inverse uncertainty quantification with limited experimental data. Mech Syst Signal Process. 2019;118:534–48. https://doi.org/10.1016/j.ymssp.2018.08.050.
4. Bathe KJ. Finite Element Procedures. New Jersey: Prentice Hall; 1996.
5. Beck JL, Katafygiotis LS. Updating models and their uncertainties I: Bayesian statistical framework. J Eng Mech. 1998;124(4):455–61. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:4(455).
6. Beck JL, Au SK. Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. J Eng Mech. 2002;128(4):380–91. https://doi.org/10.1061/(ASCE)0733-9399(2002)128:4(380).
7. Figueiredo E, Radu L, Worden K, Farrar CR. A Bayesian approach based on a Markov-chain Monte Carlo method for damage detection under unknown sources of variability. Eng Struct. 2014;80:1–10. https://doi.org/10.1016/j.engstruct.2014.08.042.
8. Ching J, Chen Y-C. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. J Eng Mech. 2007;133(7):816–32. https://doi.org/10.1061/(ASCE)0733-9399(2007)133:7(816).
9. Betz W, Papaioannou I, Straub D. Transitional Markov chain Monte Carlo: Observations and improvements. J Eng Mech. 2016;142(5):04016016. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001066.
10. Zhang J, Shields MD. Efficient Monte Carlo resampling for probability measure changes from Bayesian updating. Probab Eng Mech. 2019;55:54–66. https://doi.org/10.1016/j.probengmech.2018.10.002.
11. Straub D, Papaioannou I. Bayesian updating with structural reliability methods. J Eng Mech. 2015;141(3):04014134. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000839.
12. Goller B, Broggi M, Calvi A, Schuëller GI. A stochastic model updating technique for complex aerospace structures. Finite Elem Anal Des. 2011;47(7):739–52. https://doi.org/10.1016/j.finel.2011.02.005.
13. Nagel JB, Rieckermann J, Sudret B. Principal component analysis and sparse polynomial chaos expansions for global sensitivity analysis and model calibration: Application to urban drainage simulation. Reliabil Eng Syst Safety. 2020;195:106737. https://doi.org/10.1016/j.ress.2019.106737.
14. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Cambridge, Massachusetts: MIT Press; 2006.
15. Oladyshkin S, Mohammadi F, Kroeker I, Nowak W. Bayesian^3 active learning for the Gaussian process emulator using information theory. Entropy. 2020;22(8). https://doi.org/10.3390/e22080890.
16. Taflanidis AA, Aakash BS, Yi S-R, Conte JP. Surrogate-aided Bayesian calibration with adaptive learning strategies. Mech Syst Signal Process. 2025;237:113014. https://doi.org/10.1016/j.ymssp.2025.113014.
17. Angelikopoulos P, Papadimitriou C, Koumoutsakos P. X-TMCMC: Adaptive kriging for Bayesian inverse modeling. Comput Methods Appl Mech Eng. 2015;289:409–28. https://doi.org/10.1016/j.cma.2015.01.015.
18. Jensen HA, Esse C, Araya V, Papadimitriou C. Implementation of an adaptive meta-model for Bayesian finite element model updating in time domain. Reliabil Eng Syst Safety. 2017;160:174–90. https://doi.org/10.1016/j.ress.2016.12.005.
19. Sinsbeck M, Cooke E, Nowak W. Sequential design of computer experiments for the computation of Bayesian model evidence. SIAM/ASA Journal on Uncertainty Quantification. 2021;9(1):260–79. https://doi.org/10.1137/20M1320432.
20. Kitahara M, Dang C, Beer M. Bayesian updating with two-step parallel Bayesian optimization and quadrature. Comput Methods Appl Mech Eng. 2023;403:115735. https://doi.org/10.1016/j.cma.2022.115735.
21. Wei P, Kitahara M, Faes MGR, Beer M. Probabilistic calibration of model parameters with approximate Bayesian quadrature and active machine learning. J Reliabil Sci Eng. 2025;1(1):015003. https://doi.org/10.1088/3050-2454/ad9f62.
22. Song J, Liang Z, Wei P, Beer M. Sampling-based adaptive Bayesian quadrature for probabilistic model updating. Comput Methods Appl Mech Eng. 2025;433:117467. https://doi.org/10.1016/j.cma.2024.117467.
23. Dang C, Valdebenito MA, Faes MGR, Wei P, Beer M. Structural reliability analysis: A Bayesian perspective. Struct Saf. 2022;99:102259. https://doi.org/10.1016/j.strusafe.2022.102259.
24. Yuen K-V, Beck JL, Katafygiotis LS. Unified probabilistic approach for model updating and damage detection. J Appl Mech (ASME). 2006;73(4):555–64. https://doi.org/10.1115/1.2150235.
25. Beck JL. Bayesian updating, model class selection and robust stochastic predictions of structural response. In: De Roeck G, Degrande G, Lombaert G, Müller G, editors. 8th International Conference on Structural Dynamics (EURODYN 2011). Leuven, Belgium: EU; 2011. p. 1–6.
26. McKay MD, Conover WJ, Beckman RJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 1979;21(2):239–45.
27. Vořechovský M. Hierarchical refinement of Latin hypercube samples. Computer-Aided Civil and Infrastructure Engineering. 2015;30(5):394–411. https://doi.org/10.1111/mice.12088.
28. Shields MD, Zhang J. The generalization of Latin hypercube sampling. Reliabil Eng Syst Safety. 2016;148:96–108. https://doi.org/10.1016/j.ress.2015.12.002.
29. Poloczek M, Wang J, Frazier P. Multi-information source optimization. Advances in neural information processing systems, 2017;30

30. O'Hagan A. Bayes-Hermite quadrature. J Stat Plan Infer. 1991;29(3):245–60. https://doi.org/10.1016/0378-3758(91)90002-V.
31. Wei P. Bayesian model inference with complex posteriors: Exponential-impact-informed Bayesian quadrature. Mech Syst Signal Process. 2025;239:113333. https://doi.org/10.1016/j.ymssp.2025.113333.
32. Li P-P, Dang C, Acevedo CH, Valdebenito MA, Faes MGR. Bayesian model updating via streamlined Bayesian active learning cubature. Mechanical Systems and Signal Processing. 2026. https://doi.org/10.48550/arXiv.2509.11204.
33. Dang C, Wei P, Faes MGR, Valdebenito MA, Beer M. Parallel adaptive Bayesian quadrature for rare event estimation. Reliabil Eng Syst Safety. 2022;225:108621. https://doi.org/10.1016/j.ress.2022.108621.
34. Fishman GS. Monte Carlo: Concepts. New York, NY: Algorithms and Applications. Springer; 1996.
35. Rubin DB. Using the SIR algorithm to simulate posterior distributions. In: Bernardo JM, DeGroot MH, Lindley DV, Smith AFM, editors. Bayesian Statistics 3. New York: Oxford University Press Inc; 1988. p. 395–402.
36. Khodaparast HH, Mottershead JE, Friswell MI. Perturbation methods for the estimation of parameter variability in stochastic model updating. Mech Syst Signal Process. 2008;22(8):1751–73. https://doi.org/10.1016/j.ymssp.2008.03.001.
37. Järvenpää M, Corander J. Approximate Bayesian inference from noisy likelihoods with Gaussian process emulated MCMC. J Mach Learn Res. 2024:25(366):1–55. https://doi.org/10.48550/arXiv.2104.03942
38. Villani P, Unger J, Weiser M. Adaptive Gaussian process regression for Bayesian inverse problems. In: Proceedings of the Conference Algoritmy. 2024;214–224. https://doi.org/10.48550/arXiv.2404.19459.
39. Riccius L, Rocha IBCM, Bierkens J, Kekkonen H, van der Meer FP. Integration of active learning and MCMC sampling for efficient Bayesian calibration of mechanical properties. Eur J Mech A Solids. 2026;117:106015. https://doi.org/10.1016/j.euromechsol.2026.106015.
40. Nitzler J, Temür BZ, Koutsourelakis P-S, Wall WA. Efficient Bayesian multi-fidelity inverse analysis for expensive and non-differentiable physics-based simulations in high stochastic dimensions. Comput Methods Appl Mech Eng. 2026;448:118442. https://doi.org/10.1016/j.cma.2025.118442.
41. Song J, Wei P. Bayesian active learning for Bayesian model updating: The art of acquisition functions and beyond. Mech Syst Signal Process. 2026;251:114237. https://doi.org/10.1016/j.ymssp.2026.114237.