Combining Data and Physical Models for Probabilistic Analysis: A Bayesian Augmented Space Learning Perspective

Fangqi Hong[a], Pengfei Wei[a,*], Jingwen Song[b], Matthias G.R. Faes[c], Marcos A. Valdebenito[c], Michael Beer[d,e,f]

[a]*School of Power and Energy, Northwestern Polytechnical University, Xi'an 710072, China*
[b]*School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China*
[c]*Chair for Reliability Engineering, TU Dortmund University, Leonhard-Euler Strasse 5, 44227, Dortmund, Germany*
[d]*Institute for Risk and Reliability, Leibniz Universität Hannover, Hannover 30167, Germany*
[e]*Institute for Risk and Uncertainty, University of Liverpool, Liverpool L69 7ZF, UK*
[f]*International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, China*

Abstract

The traditional methods for probabilistic analysis of physical systems often follow a non-intrusive scheme with, random samples for stochastic model parameters generated in the outer loop, and for each sample, physical model (described by PDEs) solved in the inner loop using, e.g., finite element method (FEM). Two of the biggest challenges when applying probabilistic methods are the high computational burden due to the repeated calls of the expensive-to-estimate computational models, and the difficulties of integrating the numerical errors from both loops. To overcome these challenges, we present a new framework for transforming the PDEs with stochastic parameters into equivalent deterministic PDEs, and then devise a statistical inference method, called Bayesian Augmented Space Learning (BASL), for inferring the probabilistic descriptors of the model responses with the combination of measurement data and physical models. With the two sources of information available, only a one-step Bayesian inference needs to be performed, and the numerical errors are summarized by posterior variances. The method is then further extended to the case where the values of the parameters of the test pieces for measurement are not precisely known. The effectiveness of the proposed methods is demonstrated with academic and real-world physical models.

Keywords: Bayesian learning; Probabilistic analysis; Gaussian process regression; Augmented space; Parameter identification.

1. Introduction

Uncertainty Quantification (UQ) of physical system responses subjected to non-deterministic model parameters and excitations is a typical, yet challenging, task in many aspects of engineering practices (such as structural reliability analysis, structural health monitoring and system identification) and scientific disciplines (e.g., computational physics and chemistry). UQ informs the analyst about effects of natural randomness and/or epistemic uncertainties on the physical quantities to be predicted, allowing for robust and reliable prediction and decision-making. To date, three categories of models, including the probabilistic model, the non-probabilistic models (e.g., the interval model and fuzzy set theory), and the imprecise

*Corresponding author at School of Power and Energy, Northwestern Polytechnical University, Xian 710072, China
Email address: pengfeiwei@nwpu.edu.cn (Pengfei Wei)

probabilistic models (e.g., the probability-box model and the evidence theory), have been developed for meeting alternative distinct scenarios for uncertainty characterization, and one can refer to Refs. [1], [2] for review and comparison, and also Ref. [3] for review of propagation methods. Thanks to its maturity, probabilistic techniques benefit from their capability of handling alternative types of uncertainties in a theoretical rigorous way and plenty of simulation methods for analysis. For simplicity, only the probabilistic UQ is of concern in this work. For most studies in this aspect, it is assumed that the information available for probabilistic analysis includes the probability models (of model parameters, external excitations, etc., collectively referred to as model inputs) and a physical model (described by PDEs and solved by, e.g., finite element method, FEM) [4, 5, 6]. The objective of this work is to devise a Bayesian machine learning scheme to infer the probabilistic characteristics of the model response by combining measured data (generated from, e.g. experiments) and the physical models described by PDEs. Thus it is assumed that, besides the above two sources of information, also measured data (either complete or incomplete) on the physical process is available. For simplicity, only linear PDEs are considered, as the nonlinear problems would demand specific treatments under this new framework.

A large share of the state-of-the-art developments for probabilistic analysis follow a non-intrusive and double-loop scheme. Following such a scheme, the physical computational model that describes the physical behaviour is built by the modellers in the inner loop. Then, this model is treated by the probabilistic analysts in the outer loop, that is, the physical model is evaluated for different realizations of the uncertain parameters [7]. The main advantage of the non-intrusive scheme is that the separation of physical and probabilistic analysis makes it easier for the communication between modellers and analysts. The most popular developments in this aspect are the Monte Carlo simulation and its improved variants [8, 9], such as importance sampling [10] and Markov Chain Monte Carlo (MCMC) simulation [11]. This popularity is mainly due to the guarantee of estimation accuracy against the required number of model calls. To alleviate the corresponding computational effort, many variance reduction techniques for making a better trade-off between accuracy and required number of model calls (see e.g., [12, 13, 14]) have been developed. Nonetheless, the double-loop scheme is not broken, and repeated calls of the computational models, each of which can be expensive, are required. Other non-intrusive techniques, such as the moment based methods [15, 16], polynomial chaos approximation [17, 18], and artificial neural networks [19], have also been developed. However, such methods may not have the accuracy guarantees as rigorous as the traditional Monte Carlo simulation. Further, still, repeated calls of the expensive simulators are still unavoidable. These two points can cause hesitation among analysts to apply these methods in engineering practice. As a final point, it needs to be pointed out that it is difficult to integrate the numerical errors in the physical analysis loop. As such, in the probabilistic analysis, inaccurate data might be used to infer results and feed the decision process. It should be noted that developments of intrusive methods, e.g., based on polynomial chaos approximation [20], are also available for UQ of physical systems, but generally lack systematic treatments.

Traditional numerical methods for solving the physical models (described by PDEs) include Finite Element Method (FEM), Boundary Element Method (BEM) [21], meshfree and particle methods [22], etc. Each of these methods can be considered for developing the computational model used in the above mentioned non-intrusive and double-loop scheme. The past decade has witnessed a new trend of extending the machine learning algorithms for solving the physical models with a meshfree approach. The state-of-the-art developments in this context can be grouped as either non-Bayesian or Bayesian methods. As for non-Bayesian methods, a typical development is the Physics-Informed Neural Network (PINN)[23], which integrates measurement data and physical models, is realized with an artificial neural network model. Improved variants of this technique can further be trained by minimizing the empirical errors from both measurement data and (randomly or actively designed) PDE grids. One can refer to Ref. [24] for a comprehensive review of the state-of-the-art developments on this topic. The PINN framework has also been extended to cope with the probabilistic analysis of model responses (see, e.g., Ref. [25] for an example). Concerning Bayesian methods, attempts at solving the governing PDEs with a Bayesian scheme can be regarded as a research branch of the topic "Bayesian probabilistic numerical methods" [26][27]. This class of methods aims at constructing (subjective) probability distributions that can summary both the prediction and the prediction accuracy for a specific numerical task. In this context, the posterior mean produces the estimates of the model quantity, while the variation of the distributions measure the prediction accuracy. As far as the ODEs/PDEs are concerned, Bayesian techniques, based on either non-parametric models (e.g., Gaussian Process Regression, GPR) [28, 29, 30] or parametric models (e.g., Bayesian neural network) [31], have both been developed. Instead of minimizing the empirical errors, these Bayesian ODEs/PDEs solvers formulate the problem of interest as a statistical inference one, and then solve it either by a full Bayesian inference scheme or a Maximum Likelihood Estimation (MLE) procedure. Herein, the numerical error is treated as epistemic uncertainty and measured by a subjective probability distribution. These schemes also have the ability to integrate the measurement data and physical model (see e.g., Ref. [32] for an example). In this work, the Bayesian ODEs/PDEs solver based on GPR model will be utilized.

As stated above, a typical formulation of the probabilistic analysis problem follows a double-loop scheme. It is expected that breaking this double loop will bring a significant gain in computational efficiency. Inspired by our previous work of introducing the augmented space for breaking the double loops for imprecise probabilistic analysis [33, 34], we first introduce a reformulation of the problem as an equivalent deterministic PDE. This PDE is then solved by a Bayesian PDE solver with the combination of both the labeled measurement data (generated by measurement and/or calling high-fidelity models) and PDE grids. This enables the derivation of the posterior features for any probabilistic characteristic (e.g., statistical moments and probability distribution functions) of the model responses. As such, also measures of numerical errors are derived on top. The developed method is named Bayesian Augmented Space Learning (BASL). Finally, next to this "forward" approach, BASL is further extended to the case where the parameters values of the

test pieces for measurements are unknown or not precisely known, but measurement data are available. For this case, the unknown parameters values also need to be inferred (known as "inverse problem"), and the devised BASL method enables solving the "forward" and "inverse" problems in one step.

The rest of this paper is organized as follows. The following section presents the formulation of problems to be solved, the motivations of the developments and a review of the GPR model to be utilized for inference. Section 3 presents the core developments, i.e., the BASL method, for forward probabilistic analysis, followed by a extension to the inverse problem in section 4. Section 5 presents benchmark studies for demonstrating the proposed methods, and section 5 gives conclusions and prospects.

2. Problem Statement and Review of Methods

2.1. Problem formulation

For most cases in this paper, we use the notations following Ref. [32]. For reasons of simplicity, only linear PDEs are considered. Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^\top$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)^\top$ denote the n-dimensional spatio-temporal variables and the m-dimensional random model parameters respectively, and assume that the response function $y(\boldsymbol{x})$ of a physical system of interest is governed by a PDE defined as:

$$\mathcal{F}_{\boldsymbol{X}}\left[y|\boldsymbol{\theta}\right](\boldsymbol{x}) := \sum_{i=1}^{L} c_i(\boldsymbol{\theta}, \boldsymbol{x}) \nabla_{\boldsymbol{\alpha}_i} y(\boldsymbol{x}) = b(\boldsymbol{\theta}, \boldsymbol{x}), \tag{1}$$

where the differential operator $\nabla_{\boldsymbol{\alpha}_i}$, with $\boldsymbol{\alpha}_i = (\alpha_{i,1}, \alpha_{i,2}, \cdots, \alpha_{i,d})$ being the orders of partial differentials, is performed with respect to only the spatio-temporal variables $\boldsymbol{x}$, and is defined as:

$$\nabla_{\boldsymbol{\alpha}_i} y := \frac{\partial^{\alpha_{i,1}}}{\partial x_1^{\alpha_{i,1}}} \cdots \frac{\partial^{\alpha_{i,d}}}{\partial x_d^{\alpha_{i,1}}} y, \tag{2}$$

where $c_i(\boldsymbol{\theta}, \boldsymbol{x})$ and $b(\boldsymbol{\theta}, \boldsymbol{x})$ are the linear coefficients and the non-homogeneous term (e.g., a random excitation), respectively, both of which are explicit expressions of the spatio-temporal vector $\boldsymbol{x}$ and the model parameter vector $\boldsymbol{\theta}$. One notes that the linear differential operator $\mathcal{F}_{\boldsymbol{X}}\left[y|\boldsymbol{\theta}\right](\boldsymbol{x})$ in Eq. (1) is taken with respect to only $\boldsymbol{x}$ on the condition of the value of $\boldsymbol{\theta}$. When the value of $\boldsymbol{\theta}$ is precisely known, Eq. (1) reduces into a regular deterministic PDE. Physically speaking, the model response function is a function of the spatio-temporal vector $\boldsymbol{x}$, which is also affected by the values of $\boldsymbol{\theta}$.

The probabilistic analysis problem is then formulated as the evaluation of certain probabilistic descriptors which synthesize the effect of uncertainty (e.g., distribution function, statistical moments, and the probability of related events) of the model response $y(\boldsymbol{x})$, given the probability distribution $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ associated with $\boldsymbol{\theta}$ and the PDE in Eq. (1) accompanied with proper initial/boundary conditions and eventually measurement data on the response $y(\boldsymbol{x})$. Specifically, we focus on the estimation of the first-order and second-order moments of

$y(\boldsymbol{x})$, but indeed, the method also applies to other probabilistic descriptors such as the distribution function of the response.

As an example, we consider a linear structural system subjected to stochastic dynamical loading. The displacement of the system at location $\boldsymbol{x}$ and time instant t is in this case governed by the PDE:

$$M(\boldsymbol{x}) \frac{\partial^2 y(\boldsymbol{x},t)}{\partial t^2} + C(\boldsymbol{x}) \frac{\partial y(\boldsymbol{x},t)}{\partial t} + K(\boldsymbol{x}) y(\boldsymbol{x},t) = b(\boldsymbol{x},t,\boldsymbol{\xi}), \tag{3}$$

where $M(\boldsymbol{x})$, $C(\boldsymbol{x})$ and $K(\boldsymbol{x})$ are the mass, damping and stiffness functions respectively, and can be assumed to be random variables. As such, for a specific existing structure, these three are constants, but for a population of existing structures or a future structure to be built, their values may show randomness across this population. Further, $b(\boldsymbol{x},t,\boldsymbol{\xi})$ refers to a random dynamic excitation indexed by the random vector $\boldsymbol{\xi}$. With the above setting, the vector of random model parameters is $\boldsymbol{\theta} = \left(M, C, K, \boldsymbol{\xi}^\top\right)^\top$, and the spatio-temporal variable vector consists of the spatial variables $\boldsymbol{x}$ and the temporary variable t. Close inspection reveals that Eq. (3) can be reformulated as the form of Eq. (1).

The state-of-the-art developments for addressing the above problems are mostly implemented in a double-loop and non-intrusive scheme. Take Monte Carlo simulation as an example, the random samples are generated for $\boldsymbol{\theta}$ in the outer loop, and then for each sample $\boldsymbol{\theta^{(i)}}$, the deterministic PDE is solved in the inner loop by, e.g., FEM, with $\boldsymbol{\theta}$ being fixed at $\boldsymbol{\theta^{(i)}}$, yielding a set of samples $y^{(i)}(\boldsymbol{x})$ of the response for further analysis. Most attempts on this aspect aim at reducing the number of required FEM analysis with the promise of numerical accuracy of the outer-loop estimation. Despite the great progress in this aspect, repeated FEM analysis can not be avoided, leading to high computational cost. Even more, it is usually difficult to monitor the two sources of numerical errors, i.e., the one caused by the numerical PDE solver in the inner loop and the one caused by the numerical integration in the outer loop, posing an unpredictable risk to the related analysis and decision making. The aim of this work is to overcome the above two limitations by introducing a new scheme with Bayesian machine learning in the joint space of the spatio-temporal variables and the random parameters , and this way to obtain the probabilistic descriptors of the solution of a stochastic PDE, such as the expectation and variance of response, and two Bayesian numerical analysis tools will be utilized for numerical implementation.

We assume that the available information for inference includes two sources, i.e., the labeled data $\mathcal{D}_1 = (\vartheta_1, \mathcal{X}_1, \mathcal{Y}_1)$ measured at the boundary conditions/initial conditions/physical process, or generated by calling a high-fidelity computational model, and the PDE data $\mathcal{D}_2 = \left(\vartheta_2, \mathcal{X}_2, \mathcal{Y}_2^{\mathcal{F}}\right)$ generated by random sampling or active design. For the measurement data $\mathcal{D}_1$, we denote $\vartheta_1 = \left(\boldsymbol{\theta}^{(1)\top}, \cdots, \boldsymbol{\theta}^{(N)\top}\right)^\top$ and $\mathcal{X}_1 = \left(\boldsymbol{x}^{(1)\top}, \cdots, \boldsymbol{x}^{(N)\top}\right)^\top$, and $\mathcal{Y}_1$ is the column vector consisting of the response values measured at $\boldsymbol{\theta} = \vartheta_1$ and $\boldsymbol{x} = \mathcal{X}_1$, i.e., $\mathcal{Y}_1 = y_{\boldsymbol{\theta}=\vartheta_1}(\mathcal{X}_1) := \boldsymbol{y}_{1:N} = \left(y^{(1)}, y^{(2)}, \cdots, y^{(N)}\right)^\top$. In real-world applications, $\mathcal{D}_1$ can be generated by measurements using design-of-experiments or derived from information in operational condi-

tions, and provides part of the information for inferring the results. For a specific test piece, the response value at a (or a set of) location(s) can be measured, and the corresponding model parameters for this piece are deterministic values. It is then logic to assume that, for $\mathcal{D}_1$ generated from one unique test piece, the values of spatio-temporal variables are exactly known, but the exact values of ϑ_1 (which is deterministic) can be (precisely or imprecisely) known or unknown. For example, considering the one-dimensional heat transfer problem, the random model parameter is assumed to be the heat-transfer coefficient. For a given test piece, the labeled data $\mathcal{D}_1$ is generated by measurement at the boundary condition where $x = 0$. For this specific test piece, the heat-transfer coefficient is known to be deterministic, but its value can be either known (measured) or unknown (unmeasured). For the later case, the estimation of the parameter values for the test piece under investigation is termed as inverse uncertainty quantification or parameter identification or model updating [35, 36], which plays an important role for structural health monitoring (SHM).

The second set of data, i.e., the PDE data $\mathcal{D}_2 = \left(\vartheta_2, \mathcal{X}_2, \mathcal{Y}_2^{\mathcal{F}}\right)$, is directly obtained from the PDEs of the physical system, where the superscript $\mathcal{F}$ indicates the differential operator defined by Eq. (1). Given the design points $\vartheta_2 = \left(\boldsymbol{\theta}^{(N+1)\top}, \cdots, \boldsymbol{\theta}^{(N+M)\top}\right)^\top$ and $\mathcal{X}_2 = \left(\boldsymbol{x}^{(N+1)\top}, \cdots, \boldsymbol{x}^{(N+M)\top}\right)^\top$, the labels $\mathcal{Y}_2^{\mathcal{F}}$ are computed from the non-homogeneous term, i.e., $\mathcal{Y}_2^{\mathcal{F}} = b\left(\vartheta_2, \mathcal{X}_2\right) := \boldsymbol{y}_{N+1:N+M}^{\mathcal{F}}$. One notes that, for the Bayesian inference solution of the stated problem, the data ϑ_2 and $\mathcal{X}_2$ can generated by either random sampling or active design [32]. With the above setting, the role of the data set $\mathcal{D}_2$ is enforcing that the physics of the problem are observed when performing Bayesian learning. In this sense, the data set $\mathcal{D}_2$ is actually related with the formulation of the physical problem. In this work, these design points are generated by a random design strategy such as Latin-hypercube sampling (LHS).

2.2. Review of Gaussian Process Regression

The numerical analysis tasks that need to be treated in this work include a multi-dimensional cubature and a PDE solution, both of which are driven by the GPR model. Thus, a brief review of the GPR model is necessary. Given an implicit deterministic model function $y\left(\boldsymbol{x}\right)$, the prior assumption is to impose a Gaussian process on the value of y across the space of $\boldsymbol{x}$. Denote the prior mean function and prior covariance function of the Gaussian process as $m\left(\boldsymbol{x}\right)$ and $\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)$ respectively. Then, the prior assumption is formulated as $\hat{y}\left(\boldsymbol{x}\right) \sim \mathcal{GP}\left(m\left(\boldsymbol{x}\right), \kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right)$. The covariance function $\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)$ is also called kernel function, which is required to be symmetric and positive definite. Many forms of kernel functions, such as the squared exponential kernel and the Matérn kernel, have been developed. One notes that the form of the kernel determines the functional behaviour that can be fitted by the GPR model as its eigenfunctions form the functional basis for regression, and the preselection of kernel form before training also reflects the prior information imposed on the inference [37, 38]. Without loss of generality, the squared exponential kernel, expressed as:

$$\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \sigma_0^2 \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \Sigma^{-1} \boldsymbol{x}\right), \tag{4}$$

which is suitable for fitting smooth functions, is utilized throughout this work, where σ_0^2 and $\Sigma = \text{diag}\left(\sigma_1^2, \cdots, \sigma_n^2\right)$ are hyper-parameters. One should note that the proposed method is applicable to any form of kernels.

Given a set of labeled data $\mathcal{D} = (\mathcal{X}, \mathcal{Y})$ of size N, assume that $\mathcal{Y}$ follows N-dimensional Gaussian distribution, i.e., $\mathcal{Y} \sim \mathcal{N}\left(m\left(\mathcal{X}\right), \mathcal{K}\right)$, where $\mathcal{K} = \kappa\left(\mathcal{X}, \mathcal{X}\right)$ with the (i,j)-th element being the covariance between the i-th and j-th training points. Further, a likelihood function can be formulated as the N-dimensional normal joint density function, i.e.,

$$\mathcal{L}\left(\mathcal{D}\right) = \frac{1}{\left(2\pi\right)^{N/2}\left|\sigma_0^2 \kappa\left(\mathcal{X}\right)\right|^{1/2}} \exp\left(-\frac{1}{2\sigma_0^2}\left(\mathcal{Y} - m\left(\mathcal{X}\right)\right)^{\top} \kappa\left(\mathcal{X}\right)^{-1}\left(\mathcal{Y} - m\left(\mathcal{X}\right)\right)\right), \tag{5}$$

whose hyper-parameters can be evaluated by maximizing this likelihood. With all these ready, the conditional (or posterior) distribution of the response value $\hat{y}\left(\boldsymbol{x}\right)$ given two non-training point $\boldsymbol{x}$ and $\boldsymbol{x}'$ is assumed to be Gaussian with mean and covariance expressed as [37]:

$$\mu_Y\left(\boldsymbol{x}\right) = \mathbb{E}_{\mathcal{D}}\left[\hat{y}\left(\boldsymbol{x}\right)\right] = m\left(\boldsymbol{x}\right) + \boldsymbol{\kappa}\left(\mathcal{X}, \boldsymbol{x}\right)^{\top} \mathcal{K}^{-1}\left(\mathcal{Y} - m\left(\mathcal{X}\right)\right), \tag{6}$$

and

$$\text{cov}_Y\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \mathbb{COV}_{\mathcal{D}}\left[\hat{y}\left(\boldsymbol{x}\right), \hat{y}\left(\boldsymbol{x}'\right)\right] = \kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right) - \boldsymbol{\kappa}\left(\mathcal{X}, \boldsymbol{x}\right)^{\top} \mathcal{K}^{-1} \boldsymbol{\kappa}\left(\mathcal{X}, \boldsymbol{x}'\right), \tag{7}$$

respectively, where $\boldsymbol{\kappa}\left(\mathcal{X}, \boldsymbol{x}\right)$ is a N-dimensional column vector with the i-th component being the prior covariance of the response between the i-th training point and $\boldsymbol{x}$, $\mathbb{E}_{\mathcal{D}}\left[\cdot\right]$ and $\mathbb{COV}_{\mathcal{D}}\left[\cdot\right]$ are posterior expectation and covariance operators with respect to the GPR model trained using data $\mathcal{D}$. Thus, the subscripts $\mathcal{D}$ in these two operators are used for indicating that they are both defined conditioned on $\mathcal{D}$.

3. Statistical Learning in Augmented Space

3.1. Fundamentals

We then go back to the probabilistic analysis problem formulated by Eq. (1). The limitations of the traditional non-intrusive probabilistic analysis, i.e., the high computational burden and the lack of accuracy guarantee caused by the incapability of monitoring the FEM errors, are mainly caused by the double-loop scheme, which needs to be broken.

As has been stated in the last section, the response y is an explicit function of $\boldsymbol{x}$, but its value is also affected by the realisation of the random model parameters $\boldsymbol{\theta}$. Imagine now there is a auxiliary function $y_a\left(\boldsymbol{\theta}, \boldsymbol{x}\right)$ with two arguments, which also meets the PDE in Eq. (1), i.e.,

$$\mathcal{F}_{\boldsymbol{X}}\left[y_a\right]\left(\boldsymbol{\theta}, \boldsymbol{x}\right) := \sum_{i=1}^{L} c_i\left(\boldsymbol{\theta}, \boldsymbol{x}\right) \nabla_{\boldsymbol{\alpha}_i} y_a\left(\boldsymbol{\theta}, \boldsymbol{x}\right) = b\left(\boldsymbol{\theta}, \boldsymbol{x}\right) \tag{8}$$

One notes that with the problem reformulated by Eq. (8), both $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are regarded as spatio-temporal variables, and the PDE in Eq. (8) can then be interpreted as a deterministic PDE, involving no non-deterministic parameters, in the augmented space of $\boldsymbol{\theta}$ and $\boldsymbol{x}$. Besides, it can be easily found that, with $\boldsymbol{\theta}$ and $\boldsymbol{x}$ being both taken as spatio-temporal variables, the expression of Eq. (8) is still a linear combination of a set of partial differential terms, indicating that the linearity of the PDE is preserved despite the above equivalent transformation. Indeed, even for nonlinear PDE, the equivalent transformation of the problem into a deterministic PDE still applies, and the mathematical characteristics of the resultant PDE also keep the same with the original one. This provides a solid foundation for treating the probabilistic analysis of nonlinear physical systems, although this is beyond the scope of the current contribution. The equivalence of the probabilistic problem formulated by Eq. (1) and the deterministic PDE solution problem expressed in Eq. (8) is demonstrated with Lemma 1.

Lemma 1. *The probabilistic descriptors (density function, moments, etc.) of $y(\boldsymbol{x})$ generated by integrating with respect to $\boldsymbol{\theta}$ from $y_a(\boldsymbol{\theta}, \boldsymbol{x})$ are the same as those defined by the PDE in Eq.(1) with random parameters, if the probability distribution of $\boldsymbol{\theta}$ is set to be the same in both problems.*

Proof. Given a realization $\boldsymbol{\theta}^* \sim f(\boldsymbol{\theta})$, Eq. (1) degrades into a deterministic PDE, and share absolutely the same expression and meaning with Eq. (8), thus the solution $y(\boldsymbol{x})$ of Eq. (1) conditional on $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ is absolutely the same with a cross section of the solution $y_a(\boldsymbol{\theta}^*, \boldsymbol{x})$ of Eq. (8) at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Therefore,

$$\int_{\Omega} y(\boldsymbol{x}) f(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} = \int_{\Omega} y_a(\boldsymbol{\theta}, \boldsymbol{x}) f(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \tag{9}$$

holds for any subregion Ω of the probability space of $\boldsymbol{\theta}$. If Ω is set to be the probability space of $\boldsymbol{\theta}$, Eq. (9) is the expectation of $y(\boldsymbol{x})$; if Ω is set to be $(-\infty, \boldsymbol{a}]$, Eq. (9) is the cumulative distribution function (CDF) of $y(\boldsymbol{x})$ at $\boldsymbol{a}$. With the CDFs equalling to each other at all points of $\boldsymbol{\theta}$, it follows by definition of any probabilistic measure that these are equal too. □

For more intuitively illustrating Lemma 1, consider a simple ODE formulated as $\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} - \pi^2\theta^2 y = 0$ with boundary conditions $\frac{\mathrm{d}y}{\mathrm{d}x}|_{x=0} = \pi\theta$ and $y|_{x=0} = 0$, where θ denotes the random parameter. Using Lemma 1, the solution of the equivalent deterministic ODE is solved as $y_a(\theta, x) = 0.5\exp(\pi\theta x) - 0.5\exp(-\pi\theta x)$, as schematically shown in Figure 1. For some specific values of θ, such as 0.5 and 1, the solutions of the stochastic ODE and the sections of $y_a(\theta, x)$ are displayed in Figure 1 for comparison. As can be seen, they share exactly the same formulation, indicating the correctness of Lemma 1. Based on Lemma 1, the probabilistic analysis problem is now transformed into the solution of the deterministic PDE in Eq. (8) in the augmented space. For probabilistic analysis, we need only to solve the PDE formulated by Eq. (8), and then integrate with respect to $\boldsymbol{\theta}$. The above idea of transforming the double-loop problem into a single-loop one by reformulating the problem in the augmented space is inspired by our previous work for propagating

the imprecise probability models [33, 39], where the two-level analysis caused by the hierarchical structure of the uncertainty model is decoupled by introducing an augmented space.

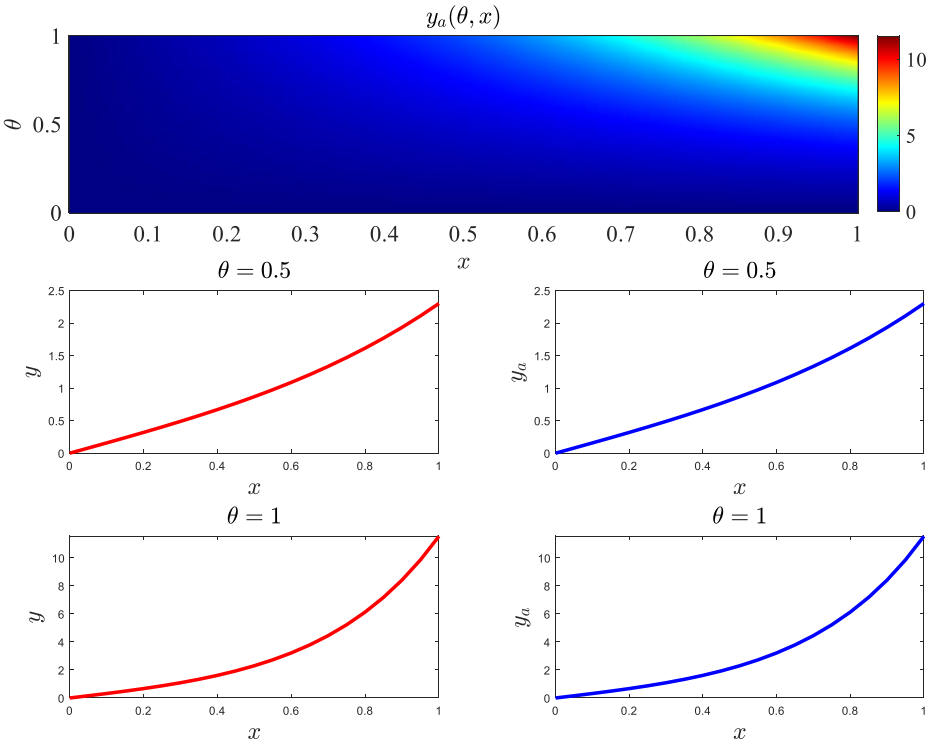

Figure 1: The solution of deterministic ODE of the illustrating example, and the comparison of the solution $y\left(x \mid \theta\right)$ of stochastic ODE and the section of $y_a\left(\theta, x\right)$ at $\theta = 0.5, 1$.

There are many numerical procedures available for solving the deterministic linear PDE of Eq. (8). Typical methods include, e.g., finite element method, physics-informed neural networks (PINN), and Bayesian PDE solver. In the setting of this work, we assume two classes of information, i.e., the labeled data $\mathcal{D}_1$ measured from sample pieces, and the PDE data $\mathcal{D}_2$. For properly informing the numerical error, the Bayesian PDE solver devised in Ref. [32], with the two groups of data as input information, is utilized, and the proposed method is termed as Bayesian Augmented Space Learning (BASL).

3.2. Bayesian Augmented Space Learning: Numerical Implementation

The numerical implementation of the BASL procedure, informed by both the measurement data $\mathcal{D}_1 = (\vartheta_1, \mathcal{X}_1, \mathcal{Y}_1)$ and the PDE data $\mathcal{D}_2 = \left(\vartheta_2, \mathcal{X}_2, \mathcal{Y}_2^{\mathcal{F}}\right)$, is introduced in this subsection following Ref. [32]. For simplicity, we assume in this subsection that the labeled data $\mathcal{D}_1$ is complete, which means that the values of $\boldsymbol{\theta}$ for each sample are exactly known. For the case of incomplete information where the values of $\boldsymbol{\theta}$ for each test pieces are not known or not precisely known, the method will be extended in the next subsection.

The prior assumption for the inference is then that $y_a(\boldsymbol{\theta}, \boldsymbol{x})$ follows a Gaussian process in the augmented space, i.e.,

$$\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x}) \sim \mathcal{GP}\left(m(\boldsymbol{\theta}, \boldsymbol{x}), \kappa\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right)\right), \tag{10}$$

with $m(\boldsymbol{\theta}, \boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{\theta}, \boldsymbol{x})^\top \boldsymbol{\beta}$ being the prior mean and $\kappa\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right)$ being the prior covariance, which in this work is assumed to be separable, i.e., $\kappa\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right) = \sigma_0^2 \kappa_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}') \kappa_X(\boldsymbol{x}, \boldsymbol{x}')$. Note that $\boldsymbol{p}(\boldsymbol{\theta}, \boldsymbol{x})$ represents a prescribed vector-valued function which contains a basis for estimating the mean while $\boldsymbol{\beta}$ is a vector of unknown coefficients. Both $\boldsymbol{p}(\boldsymbol{\theta}, \boldsymbol{x})$ and $\boldsymbol{\beta}$ are vectors of dimension q. Both $\kappa_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}')$ and $\kappa_X(\boldsymbol{x}, \boldsymbol{x}')$ are assumed to be of normalized form without the variance term σ_0^2. For separable kernels, one can refer to the structured GPR model presented, e.g., in Ref. [40], for details.

Given the above assumption, it is known that a linear differential operation $\mathcal{F}$ on $\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x})$ also follows a Gaussian process, i.e., $\mathcal{F}_{\boldsymbol{X}}[\hat{y}_a](\boldsymbol{\theta}, \boldsymbol{x}) \sim \mathcal{GP}\left(m^{\mathcal{F}}(\boldsymbol{\theta}, \boldsymbol{x}), \kappa^{\mathcal{F}\mathcal{F}}\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right)\right)$, where the prior mean and covariance are expressed as:

$$m^{\mathcal{F}}(\boldsymbol{\theta}, \boldsymbol{x}) = \mathcal{F}_X[m_a](\boldsymbol{\theta}, \boldsymbol{x}) = \boldsymbol{p}_{\mathcal{F}}^\top(\boldsymbol{\theta}, \boldsymbol{x}) \boldsymbol{\beta}, \tag{11}$$

and

$$\kappa^{\mathcal{F}\mathcal{F}}\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right) = \left(\mathcal{F}_{\boldsymbol{X}} \times \mathcal{F}_{\boldsymbol{X}'}\right)[\kappa_a]\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right) = \sigma_0^2 \kappa_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}') \mathcal{F}_{\boldsymbol{X}}[\mathcal{F}_{\boldsymbol{X}'}[\kappa_{\boldsymbol{X}}]](\boldsymbol{x}, \boldsymbol{x}'), \tag{12}$$

respectively, where $\boldsymbol{p}_{\mathcal{F}}(\boldsymbol{\theta}, \boldsymbol{x}) = [\mathcal{F}_{\boldsymbol{X}}[p_1](\boldsymbol{\theta}, \boldsymbol{x}), \cdots, \mathcal{F}_{\boldsymbol{X}}[p_q](\boldsymbol{\theta}, \boldsymbol{x})]^\top$, and $\mathcal{F}_{\boldsymbol{X}} \times \mathcal{F}_{\boldsymbol{X}'}[\cdot]$ indicated that the argument is differentiated with respect to both $\boldsymbol{x}$ and $\boldsymbol{x}'$.

Following Ref. [32], we also define the identity operator as $\mathcal{I}_{\boldsymbol{X}}[y_a](\boldsymbol{\theta}, \boldsymbol{x}) := y_a(\boldsymbol{\theta}, \boldsymbol{x})$. Then, following the prior assumption above, it is known that $\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x})$ (or, equivalently, $\mathcal{I}_{\boldsymbol{X}}[\hat{y}_a](\boldsymbol{\theta}, \boldsymbol{x})$) and $\mathcal{F}_{\boldsymbol{X}}[\hat{y}_a](\boldsymbol{\theta}, \boldsymbol{x})$ are jointly correlated Gaussian process, and their cross-covariance function, denoted as $\kappa^{\mathcal{I}\mathcal{F}}\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right)$, can be formulated as [32]:

$$\begin{aligned} \kappa^{\mathcal{I}\mathcal{F}}\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right) &= \left(\mathcal{I}_{\boldsymbol{X}} \times \mathcal{F}_{\boldsymbol{X}'}\right)[\kappa]\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right) = \mathcal{F}_{\boldsymbol{X}'}[\kappa]\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right) \\ &= \sigma_0^2 \kappa_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}') \mathcal{F}_{\boldsymbol{X}'}[\kappa_X](\boldsymbol{x}, \boldsymbol{x}'). \end{aligned} \tag{13}$$

Given the Gaussian prior assumption, the vector of the response values $\mathcal{Y} = \left(\mathcal{Y}_1^\top, \mathcal{Y}_2^{\mathcal{F}\top}\right)^\top$ follows a $(N+M)$-dimensional Gaussian distribution, i.e.,

$$\mathcal{Y} = \left(\mathcal{Y}_1^\top, \mathcal{Y}_2^{\mathcal{F}\top}\right)^\top \sim \mathcal{N}_{N+M}\left(\mathbf{P}\boldsymbol{\beta}, \sigma_0^2 \mathcal{K}(\boldsymbol{\sigma})\right) \text{ with } \mathbf{P} = \begin{bmatrix} \boldsymbol{p}(\vartheta_1, \mathcal{X}_1) \\ \boldsymbol{p}_{\mathcal{F}}(\vartheta_2, \mathcal{X}_2) \end{bmatrix} \text{ and } \mathcal{K}(\boldsymbol{\sigma}) = \begin{bmatrix} \mathcal{K}_{\mathcal{I}\mathcal{I}}(\boldsymbol{\sigma}) & \mathcal{K}_{\mathcal{I}\mathcal{F}}(\boldsymbol{\sigma}) \\ \mathcal{K}_{\mathcal{I}\mathcal{F}}^\top(\boldsymbol{\sigma}) & \mathcal{K}_{\mathcal{F}\mathcal{F}}(\boldsymbol{\sigma}) \end{bmatrix}, \tag{14}$$

where $\mathcal{K}_{\mathcal{II}}(\boldsymbol{\sigma}) = \kappa_\Theta(\vartheta_1, \vartheta_1|\boldsymbol{\sigma}_\Theta) \odot \kappa_X(\mathcal{X}_1, \mathcal{X}_2|\boldsymbol{\sigma}_X)$, $\mathcal{K}_{\mathcal{IF}}(\boldsymbol{\sigma}) = \kappa_\Theta(\vartheta_1, \vartheta_2|\boldsymbol{\sigma}_\Theta) \odot \mathcal{F}_{\boldsymbol{X}'}[\kappa_X](\mathcal{X}_1, \mathcal{X}_2|\boldsymbol{\sigma}_X)$, $\mathcal{K}_{\mathcal{FF}}(\boldsymbol{\sigma}) = \kappa_\Theta(\vartheta_2, \vartheta_2|\boldsymbol{\sigma}_\Theta) \odot \mathcal{F}_{\boldsymbol{X}}[\mathcal{F}_{\boldsymbol{X}'}[\kappa_{\boldsymbol{X}}]](\mathcal{X}_2, \mathcal{X}_2|\boldsymbol{\sigma}_X)$, $\odot$ refers to the Hadamard product, and $\boldsymbol{\sigma} = \left(\boldsymbol{\sigma}_{\boldsymbol{\theta}}^\top, \boldsymbol{\sigma}_{\boldsymbol{x}}^\top\right)^\top$ indicates the scale hyper-parameters involved in the (squared exponential) kernels of $\boldsymbol{\theta}$ and $\boldsymbol{x}$. The likelihood function, with the hyperparameters $\boldsymbol{\beta}$, σ_0^2 and $\boldsymbol{\sigma}$, is formulated as:

$$\mathcal{L}\left(\mathcal{D}|\boldsymbol{\beta}, \sigma_0^2, \boldsymbol{\sigma}\right) = \frac{1}{(2\pi)^{(N+M)/2}\left|\sigma_0^2 \mathcal{K}(\boldsymbol{\sigma})\right|^{1/2}} \exp\left(-\frac{1}{2\sigma_0^2}(\mathcal{Y} - \mathbf{P}\boldsymbol{\beta})^\top \mathcal{K}^{-1}(\boldsymbol{\sigma})(\mathcal{Y} - \mathbf{P}\boldsymbol{\beta})\right), \tag{15}$$

and its negative log is:

$$\log \mathcal{L}\left(\mathcal{D}|\boldsymbol{\beta}, \sigma_0^2, \boldsymbol{\sigma}\right) \propto \frac{1}{2\sigma_0^2}(\mathcal{Y} - \mathbf{P}\boldsymbol{\beta})^\top \mathcal{K}^{-1}(\boldsymbol{\sigma})(\mathcal{Y} - \mathbf{P}\boldsymbol{\beta}) + \frac{1}{2}\log\left|\sigma_0^2 \mathcal{K}(\boldsymbol{\sigma})\right|. \tag{16}$$

Making the partial derivatives of $\log \mathcal{L}\left(\mathcal{D}|\boldsymbol{\beta}, \sigma_0^2, \boldsymbol{\sigma}\right)$ with respect to $\boldsymbol{\beta}$ and σ_0^2 equal to zero, the estimates of the hyperparameters $\hat{\boldsymbol{\beta}}(\boldsymbol{\sigma})$ and $\hat{\sigma}_0^2(\boldsymbol{\sigma})$ are derived as [37]:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\sigma}) = \left(\mathbf{P}^\top \mathcal{K}^{-1}(\boldsymbol{\sigma})\mathbf{P}\right)^{-1} \mathbf{P}^\top \mathcal{K}^{-1}(\boldsymbol{\sigma})\mathcal{Y} \quad \text{and} \quad \hat{\sigma}_0^2(\boldsymbol{\sigma}) = \frac{1}{N}(\mathcal{Y} - \mathbf{P}\boldsymbol{\beta})^\top \mathcal{K}^{-1}(\boldsymbol{\sigma})(\mathcal{Y} - \mathbf{P}\boldsymbol{\beta}). \tag{17}$$

Substituting Eq. (17) into Eq. (16), an objective function can be obtained as:

$$\mathcal{L}^{\text{obj}}(\boldsymbol{\sigma}) = \log \mathcal{L}\left(\mathcal{D}|\hat{\boldsymbol{\beta}}(\boldsymbol{\sigma}), \hat{\boldsymbol{\sigma}}_0^2(\boldsymbol{\sigma}), \boldsymbol{\sigma}\right). \tag{18}$$

The minimisation of $\mathcal{L}$ yields the optimal estimates $\hat{\boldsymbol{\sigma}}$ of the hyperparameters $\boldsymbol{\sigma}$. Substituting $\hat{\boldsymbol{\sigma}}$ into Eq. (17), the optimal estimates of the hyperparameters $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\sigma}}_0^2$ are also generated. The limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) quasi-Newton approximation method is utilized in this work for solving the optimisation problem formulated by Eq. (18), and one can refer to Ref. [41] for details.

With the values of the hyperparameters being estimated, the posterior mean $\mu_{Ya}(\boldsymbol{\theta}, \boldsymbol{x})$ and covariance $\text{cov}_{Ya}\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right)$ for fitting the auxiliary function $y_a(\boldsymbol{\theta}, \boldsymbol{x})$ are formulated as:

$$\mu_{Ya}(\boldsymbol{\theta}, \boldsymbol{x}) = \mathbb{E}_{\mathcal{D}}[\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x})] = \boldsymbol{p}^\top(\boldsymbol{\theta}, \boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{\gamma}^\top(\boldsymbol{\theta}, \boldsymbol{x})\mathcal{K}^{-1}(\mathcal{Y} - \mathbf{P}\boldsymbol{\beta}), \tag{19}$$

and

$$\begin{aligned} \text{cov}_{Ya}\left((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x}')\right) &= \mathbb{COV}_{\mathcal{D}}\left[\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x}), \hat{y}_a(\boldsymbol{\theta}', \boldsymbol{x}')\right] \\ &= \sigma_0^2\left[\kappa_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}')\kappa_X(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{\gamma}^\top(\boldsymbol{\theta}, \boldsymbol{x})\mathcal{K}^{-1}\boldsymbol{\gamma}(\boldsymbol{\theta}', \boldsymbol{x}')\right], \end{aligned} \tag{20}$$

where $\boldsymbol{\gamma}(\boldsymbol{\theta}, \boldsymbol{x}) = [\boldsymbol{\kappa}_\Theta(\boldsymbol{\theta}, \vartheta_1) \odot \boldsymbol{\kappa}_X(\boldsymbol{x}, \mathcal{X}_1); \boldsymbol{\kappa}_\Theta(\boldsymbol{\theta}, \vartheta_2) \odot \mathcal{F}_{\boldsymbol{X}'}[\kappa_X](\boldsymbol{x}, \mathcal{X}_2)]$. The posterior mean $\mu_{Ya}(\boldsymbol{\theta}, \boldsymbol{x})$ can be regarded as the posterior estimate of the auxiliary function $y_a(\boldsymbol{\theta}, \boldsymbol{x})$, while the posterior variance $\text{cov}_{Ya}((\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}, \boldsymbol{x}))$ summarises the numerical errors of this estimate.

3.3. Probabilistic analysis of model response

The purpose of this work is to infer the probabilistic descriptors of the model response $y(\boldsymbol{x})$. As has been revealed by Lemma 1, all these features can be inferred from the expression of $y_a(\boldsymbol{\theta}, \boldsymbol{x})$. However, not all probabilistic descriptors can be derived in closed forms. Two methods, a analytical one and a numerical simulation one, are developed in this subsection to implement probabilistic analysis, such that each probabilistic descriptor of $y(x)$ can be predicted with a subjective probability distribution.

3.3.1. The analytical method for probabilistic analysis

We take the expectation $E_y(\boldsymbol{x})$ of $y(\boldsymbol{x})$ as an example to illustrate the analytical method. Given the posterior features of $\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x})$ in Eqs. (19) and (20), the estimator of $E_y(\boldsymbol{x})$ can be expressed as [42, 43]:

$$\hat{E}_y(\boldsymbol{x}) = \int \hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x}) f_\Theta(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \tag{21}$$

, which is also Gaussian as it is expressed as an inner product of the Gaussian process $\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x})$ and the deterministic function $f_\Theta(\boldsymbol{\theta})$. The posterior mean and variance of $\hat{E}_y(\boldsymbol{x})$ are then formulated as:

$$\begin{aligned} \mu_{Ey}(\boldsymbol{x}) &= \mathbb{E}_{\mathcal{D}}\left[\mathbb{E}_\Theta\left[\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x})\right]\right] = \mathbb{E}_\Theta\left[\mathbb{E}_{\mathcal{D}}\left[\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x})\right]\right] \\ &= \left[\int \boldsymbol{p}(\boldsymbol{\theta}, \boldsymbol{x}) f_\Theta(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}\right]^\top \boldsymbol{\beta} + \left[\int \boldsymbol{\gamma}(\boldsymbol{\theta}, \boldsymbol{x}) f_\Theta(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}\right]^\top \mathcal{K}^{-1}(\mathcal{Y} - \mathbf{P}\boldsymbol{\beta}) \end{aligned} \tag{22}$$

, and

$$\begin{aligned} \sigma^2_{Ey}(\boldsymbol{x}) &= \mathbb{E}_{\mathcal{D}}\left[\left(\mathbb{E}_\Theta\left[\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x})\right] - \mu_{EY}(\boldsymbol{x})\right)^2\right] \\ &= \mathbb{E}_\Theta \mathbb{E}_{\Theta'}\left[\mathbb{E}_{\mathcal{D}}\left[\left(\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x}) - \mu_{Ya}(\boldsymbol{\theta}, \boldsymbol{x})\right)\left(\hat{y}_a(\boldsymbol{\theta}', \boldsymbol{x}) - \mu_{Ya}(\boldsymbol{\theta}', \boldsymbol{x})\right)\right]\right] \\ &= \sigma_0^2 \kappa_X(\boldsymbol{x}, \boldsymbol{x}) \int\int \kappa_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}') f_\Theta(\boldsymbol{\theta}) f_\Theta(\boldsymbol{\theta}') \, \mathrm{d}\boldsymbol{\theta}' \\ &\quad - \left[\int \boldsymbol{\gamma}(\boldsymbol{\theta}, \boldsymbol{x}) f_\Theta(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}\right]^\top \mathcal{K}^{-1} \left[\int \boldsymbol{\gamma}(\boldsymbol{\theta}', \boldsymbol{x}) f_\Theta(\boldsymbol{\theta}') \, \mathrm{d}\boldsymbol{\theta}'\right]. \end{aligned} \tag{23}$$

For specific form of kernel and density weight (e.g., squared exponential kernel accompanied with Gaussian or uniform density), the integrals involved in Eqs. (22) and (23) can be analytically expressed in closed form, and one can refer to Ref. [14] for a summary. For the squared exponential kernel and standard Gaussian density assumed in this work, one can refer to our previous work (e.g., Refs. [43] and [44]) for the closed-form expressions.

For the case with non-Gaussian uncertain model parameters $\boldsymbol{\theta}$, a transformation $\boldsymbol{\theta} = T(\boldsymbol{\zeta})$, based on, e.g., Nataf or Rosenblatt isoprobabilistic transformation [45], can be introduced by reformulating the PDE in Eq. (1) to be a one with standard Gaussian distributed parameter $\boldsymbol{\zeta}$. Take a one-dimensional random

parameter θ as an example. Let $F_\Theta(\theta)$ denote the CDF of θ, then let $F_\Theta(\theta) = \Phi(\zeta)$ with $\Phi(\cdot)$ indicating the standard Gaussian CDF, a inverse transformation can be obtained as $\theta = F_\Theta^{-1}[\Phi(\zeta)]$. Substituting the latter expression into Eq. (1), an equivalent PDE with standard Gaussian model parameters can be obtained.

3.3.2. The simulation method for probabilistic analysis

For other probabilistic descriptors of $y(\boldsymbol{x})$ except the expectation, as they are defined as nonlinear projections of $y_a(\boldsymbol{\theta}, \boldsymbol{x})$, the resultant posterior probability distribution for these descriptors are not Gaussian, making it generally intractable to derive the closed-form expressions for their posterior distribution. We take the variance of $y(\boldsymbol{x})$ as an example to illustrate a numerical simulation method for filling the above gap. In our previous work [44], the posterior features of the model response variance, with the model response function fitted by a GPR model, has been studied. It was concluded that, although the posterior distribution is not Gaussian, the posterior mean and variance can be derived. One can to refer Ref. [44] for more details. For the variance of model response, its posterior mean and variance can be formulated as:

$$\mu_{Vy}(\boldsymbol{x}) = \mathbb{V}_\Theta\left[\mu_{Ya}(\boldsymbol{\theta}, \boldsymbol{x})\right] + \mathbb{E}_\Theta\left[\sigma^2_{Ya}(\boldsymbol{\theta}, \boldsymbol{x})\right], \tag{24}$$

and

$$\sigma^2_{Vy}(\boldsymbol{x}) = 2\mathbb{E}_\Theta\mathbb{E}_{\Theta'}\left[\mathrm{cov}^2_{Ya}\left[(\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x})\right]\right] + 4\mathbb{E}_\Theta\mathbb{E}_{\Theta'}\left[\mu_{Ya}(\boldsymbol{\theta}, \boldsymbol{x})\,\mu_{Ya}(\boldsymbol{\theta}', \boldsymbol{x})\,\mathrm{cov}_{Ya}\left[(\boldsymbol{\theta}, \boldsymbol{x}), (\boldsymbol{\theta}', \boldsymbol{x})\right]\right], \tag{25}$$

where $\mathbb{E}_\Theta[\cdot]$ denotes the expectation operator with respect to $\boldsymbol{\theta}$. One can refer to Ref. [46] for theoretical proves of the above two formulations. As the above expressions can not be deduced analytically, the MCS method can be utilized to estimate them. With the MCS samples $\left\{\boldsymbol{\theta}^{(k)}\right\}(k = 1, \cdots, N_{\boldsymbol{\theta}})$ generated by, e.g., Latin-hypercube sampling, the corresponding estimators can be formulated as:

$$\mu^*_{Vy}(\boldsymbol{x}) = \frac{1}{N_{\boldsymbol{\theta}}}\sum_{i=1}^{N_{\boldsymbol{\theta}}}\mu_{Ya}\left(\boldsymbol{\theta}^{(i)}, \boldsymbol{x}\right)^2 - \left[\frac{1}{N_{\boldsymbol{\theta}}}\sum_{i=1}^{N_{\boldsymbol{\theta}}}\mu_{Ya}\left(\boldsymbol{\theta}^{(i)}, \boldsymbol{x}\right)\right]^2 + \frac{1}{N_{\boldsymbol{\theta}}}\sum_{i=1}^{N_{\boldsymbol{\theta}}}\sigma^2_{Ya}\left(\boldsymbol{\theta}^{(i)}, \boldsymbol{x}\right), \tag{26}$$

and

$$\sigma^{2*}_{Vy}(\boldsymbol{x}) = \frac{2}{N_{\boldsymbol{\theta}}}\sum_{i=1}^{N_{\boldsymbol{\theta}}}\mathrm{cov}^2_{Ya}\left[\left(\boldsymbol{\theta}_1^{(i)}, \boldsymbol{x}\right), \left(\boldsymbol{\theta}_2^{(i)}, \boldsymbol{x}\right)\right] + \frac{4}{N_{\boldsymbol{\theta}}}\sum_{i=1}^{N_{\boldsymbol{\theta}}}\mu_{Ya}\left(\boldsymbol{\theta}_1^{(i)}, \boldsymbol{x}\right)\mu_{Ya}\left(\boldsymbol{\theta}_2^{(i)}, \boldsymbol{x}\right)\mathrm{cov}_{Ya}\left[\left(\boldsymbol{\theta}_1^{(i)}, \boldsymbol{x}\right), \left(\boldsymbol{\theta}_2^{(i)}, \boldsymbol{x}\right)\right], \tag{27}$$

respectively.

For the cumulative distribution function of $y(\boldsymbol{x})$ denoted by $F_Y(y(\boldsymbol{x}))$, which can be approximated by $\hat{F}_Y(y(\boldsymbol{x})) = \int I[\hat{y}_a(\boldsymbol{\theta}, \boldsymbol{x}) < y] f_\Theta(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$ with $I(\cdot)$ being the indicator function, its posterior distribution

is also non-Gaussian, but the posterior mean and variance can be numerical evaluated by sampling from $f_\Theta(\boldsymbol{\theta})$. For example, the posterior mean of $\hat{F}_Y(y(\boldsymbol{x}))$ can be evaluated as:

$$\mathbb{E}_{\mathcal{D}}\left[\hat{F}_Y(y(\boldsymbol{x}))\right] = \int \mathbb{E}_{\mathcal{D}}\left[I\left(\hat{y}_a(\boldsymbol{\theta},\boldsymbol{x}) < y(\boldsymbol{x})\right)\right] f_\Theta(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \simeq \frac{1}{N_{\boldsymbol{\theta}}}\sum_{k=1}^{N_{\boldsymbol{\theta}}} \Phi\left(\frac{y(\boldsymbol{x}) - \mu_{Ya}\left(\boldsymbol{\theta}^{(k)},\boldsymbol{x}\right)}{\sigma_{Ya}\left(\boldsymbol{\theta}^{(k)},\boldsymbol{x}\right)}\right), \tag{28}$$

with $\boldsymbol{\theta}^{(k)}$ $(k = 1\cdots N_\theta)$ being a set of random samples generated following $f_\Theta(\boldsymbol{\theta})$.

4. Extension to Inverse Uncertainty Quantification

We then consider a more complex case where some values in ϑ_1 are unknown or not precisely known [47], due to, e.g., lack of measurement of the model parameters for specific test pieces. Here we assume that the values of ϑ_1 are perfectly identifiable. This is in accordance with the our previous setting that, for a specific test piece, the model parameters have unknown-but-deterministic values. One notes that, there are also more general inverse UQ problems where only a manifold where the parameters lay can be identified, and this is outside the scope of the current paper.

The identification of these unknown parameter values is called inverse UQ or parameter identification. In this section, the forward and inverse UQ problems will be solved in one step by extending the BASL method. For simplicity of illustration, it is assumed that all the values of ϑ_1 are unknown, and ϑ_1 is regarded as unknown parameters to be identified. In this case, the vector of the label values in $\mathcal{D}_1$, and the vector $\mathcal{Y}(\vartheta_1) = \left(\mathcal{Y}_1^\top(\vartheta_1), \mathcal{Y}_2^{\mathcal{F}\top}\right)^\top$, with ϑ_1 as arguments, still follow a $(N+M)$-dimensional normal distribution with the same mean vector and covariance as in Eq. (14), except that the matrices $\mathbf{P}$ and $\mathcal{K}$ are also functions of the unknown vector ϑ_1, which are rewritten as:

$$\mathbf{P}(\vartheta_1) = \begin{bmatrix} \boldsymbol{p}(\vartheta_1, \mathcal{X}_1) \\ \boldsymbol{p}_{\mathcal{F}}(\vartheta_2, \mathcal{X}_2) \end{bmatrix} \text{ and } \mathcal{K}(\boldsymbol{\sigma}, \vartheta_1) = \begin{bmatrix} \mathcal{K}_{\mathcal{I}\mathcal{I}}(\boldsymbol{\sigma}, \vartheta_1) & \mathcal{K}_{\mathcal{I}\mathcal{F}}(\boldsymbol{\sigma}, \vartheta_1) \\ \mathcal{K}_{\mathcal{I}\mathcal{F}}^\top(\boldsymbol{\sigma}, \vartheta_1) & \mathcal{K}_{\mathcal{F}\mathcal{F}}(\boldsymbol{\sigma}) \end{bmatrix}. \tag{29}$$

Thus, the likelihood function, with the hyperparameters $\boldsymbol{\beta}$, σ_0^2 and $\boldsymbol{\sigma}$ as well as the unknown values ϑ_1 as arguments, is reformulated as:

$$\mathcal{L}\left(\mathcal{D}|\boldsymbol{\beta}, \sigma_0^2, \boldsymbol{\sigma}, \vartheta_1\right) = \frac{1}{(2\pi)^{(N+M)/2}\left|\sigma_0^2\,\mathcal{K}(\boldsymbol{\sigma}, \vartheta_1)\right|^{1/2}} \exp\left(-\frac{1}{2\sigma_0^2}\left(\mathcal{Y} - \mathbf{P}(\vartheta_1)\boldsymbol{\beta}\right)^\top \mathcal{K}^{-1}(\boldsymbol{\sigma}, \vartheta_1)\left(\mathcal{Y} - \mathbf{P}(\vartheta_1)\boldsymbol{\beta}\right)\right). \tag{30}$$

We still use the Maximum Likelihood Estimation (MLE) procedure to estimate the values of both $\left(\boldsymbol{\beta}, \sigma_0^2, \boldsymbol{\sigma}\right)$ and ϑ_1. One notes that ϑ_1 is a vector of so-called deterministic-but-unknown parameters, and as such is categorized as epistemically uncertain. The prior belief on the value of ϑ_1 can be set to be the same with the density of $\boldsymbol{\theta}$, and the objective can be set to maximize the product of the likelihood function and

the prior density (denoted as $f(\vartheta_1)$) of ϑ_1. While the values of ϑ_1 are roughly measured and subjected to imprecision, the prior density can also be set to be a one with support being the interval representing the imprecision. Without of loss of generality, we assume that this density is of standard Gaussian type with zero mean vector and identity covariance matrix. The objective function, proportional to the negative log of the product of the likelihood function and the density function, is formulated as:

$$\begin{aligned}\mathcal{L}^{\text{obj}}\left(\boldsymbol{\beta},\boldsymbol{\sigma}_0^2,\boldsymbol{\sigma},\vartheta_1\right) = &\frac{1}{2\sigma_0^2}\left(\mathcal{Y}-\mathbf{P}\left(\vartheta_1\right)\boldsymbol{\beta}\right)^{\top}\mathcal{K}^{-1}\left(\boldsymbol{\sigma},\vartheta_1\right)\left(\mathcal{Y}-\mathbf{P}\left(\vartheta_1\right)\boldsymbol{\beta}\right)+\frac{1}{2}\log\left|\sigma_0^2\mathcal{K}\left(\boldsymbol{\sigma},\vartheta_1\right)\right| \\ &+\sum_{j=1}^{N}\frac{1}{2}\boldsymbol{\theta}^{(j)}\boldsymbol{\theta}^{(j)\top}\end{aligned}, \tag{31}$$

where $\boldsymbol{\theta}^{(j)}$ is the j-th row of ϑ_1. The second term of Eq. (31) can be interpreted as a L_2 regularisation for ϑ_1, with which, the value of $\boldsymbol{\theta}$ with higher density is preferred for ϑ_1. Similar to Eq. (17), it can be obtained:

$$\begin{aligned}\hat{\boldsymbol{\beta}}\left(\boldsymbol{\sigma},\vartheta_1\right) &= \left(\mathbf{P}^{\top}\left(\vartheta_1\right)\mathcal{K}^{-1}\left(\boldsymbol{\sigma},\vartheta_1\right)\mathbf{P}\left(\vartheta_1\right)\right)^{-1}\mathbf{P}^{\top}\left(\vartheta_1\right)\mathcal{K}^{-1}\left(\boldsymbol{\sigma},\vartheta_1\right)\mathcal{Y} \\ \hat{\sigma}_0^2\left(\boldsymbol{\sigma},\vartheta_1\right) &= \frac{1}{N}\left(\mathcal{Y}-\mathbf{P}\left(\vartheta_1\right)\boldsymbol{\beta}\right)^{\top}\mathcal{K}^{-1}\left(\boldsymbol{\sigma},\vartheta_1\right)\left(\mathcal{Y}-\mathbf{P}\left(\vartheta_1\right)\boldsymbol{\beta}\right)\end{aligned}, \tag{32}$$

by substituting which to Eq. (29), the objective function, as only $\boldsymbol{\sigma}$ and ϑ_1 as arguments, is reformulated as:

$$\mathcal{L}^{\text{obj}}\left(\boldsymbol{\sigma},\vartheta_1\right) = \mathcal{L}^{\text{obj}}\left(\mathcal{D}|\hat{\boldsymbol{\beta}}\left(\boldsymbol{\sigma},\vartheta_1\right),\hat{\boldsymbol{\sigma}}_0^2\left(\boldsymbol{\sigma},\vartheta_1\right),\boldsymbol{\sigma},\vartheta_1\right). \tag{33}$$

Similarly, the LBFGS algorithm is used for solving this optimisation problem. It should be noted that, with the above procedure, a MLE estimate $\hat{\vartheta}_1$ is also obtained for ϑ_1 as a byproduct. In probabilistic UQ, this is termed as inverse UQ, which aims at inferring the values or the probability distribution of the model parameters, given the measurements of model response. This problem is mostly treated with Bayesian model updating [35, 36]. With the developed method in this subsection, the forward and inverse problems are combined to one problem, and solved in one step.

Besides, the values of ϑ_1 can also be inferred in a full Bayesian scheme, following which the estimator error can be measured by the posterior variance of ϑ_1. By substituting the MLE estimations $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_0^2$, and $\hat{\boldsymbol{\sigma}}$ of the hyperparameters to Eq. (30), and then multiplying the result by the prior density of ϑ_1, a non-normalized posterior density for ϑ_1 can be obtained, and the random samples following this posterior density can be generated by either rejection sampling or MCMC algorithms [48]. The error analysis for the estimator of ϑ_1 is not the concern of this work, but may be of special interest when applying the BASL method to, e.g., structural health monitoring, thus we only show the MLE estimation of ϑ_1 in the benchmark study.

The flowchart of the proposed BASL method for probabilistic analysis is given in Figure 2. The stochastic PDE is first transformed to a deterministic PDE based on Lemma 1. Then, given the prior GP model

assumption of solution of the deterministic PDE, the measurement data $\mathcal{D}_1$ and the PDE data $\mathcal{D}_2$, a Bayesian statistical inference method is applied to derive the posterior distribution for the solution $y_a(\boldsymbol{\theta}, \boldsymbol{x})$ of the deterministic PDE solution. Once, the solution of deterministic PDE is obtained, its probabilistic descriptors can be computed by analytical or simulation method reported in Subsection 3.3. Further, the above procedures of BASL can also be extended to identify the deterministic values of the random parameters for each test piece, i.e., the inverse UQ problem, as reported in Section 4.

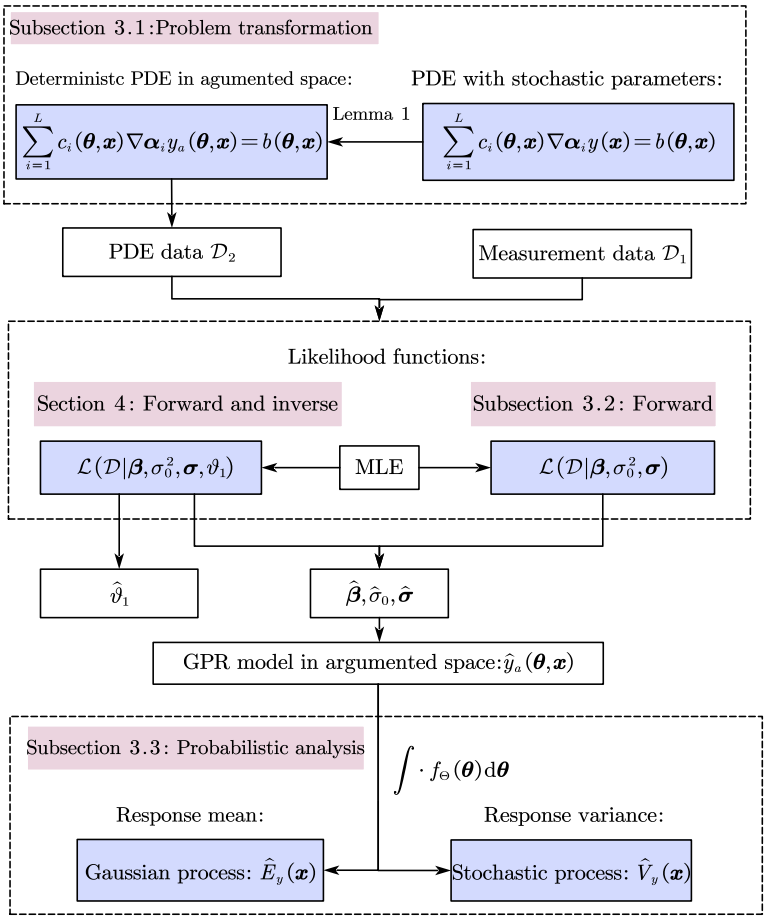

Figure 2: The flowchart of BASL for probabilistic analysis of stochastic PDE.

5. Benchmark study

In this section, several academic and engineering examples will be introduced for demonstrating the proposed methods. We assume that the measurement data $\mathcal{D}_1$ can be measured in the initial/boundary condition and/or the process, where the values of ϑ_1 can be either known or unknown, leading to two cases for each example. For simplicity, the expectation and variance of model response are considered for inference,

where, the posterior features of the expectation of model response are derived with closed-form expressions, for variance of response, its posterior features are evaluated by MCS. The benchmarks utilized in this section possesses an analytical solution, and the reference values of expectation and variance are computed by MCS with 5000 samples. For illustrating the prediction accuracy of all results, in this we report the posterior confidence intervals (CIs) with the bounds being the posterior mean minus/plus two STDs for both response expectation and variance.

5.1. A linear ODE with approximately linear response behaviour

Considering the following ordinary differential equation (ODE):

$$\mathcal{F}_X\left[y\right](\theta, x) := -\theta y\left(x\right) + \frac{\partial^2 y\left(x\right)}{\partial x^2} = 2\cos\left(x+1\right), \tag{34}$$

where $x \in [0,1]$ is the spatial variable, and $\theta \sim \mathcal{U}\left(0.64, 1.44\right)$ is a uniformly distributed model parameter. The labeled data is measured from two sample pieces with model parameter values being $\theta^{(1)} = 0.9157$ and $\theta^{(2)} = 1.1643$ respectively.

We then consider the first case where the values of $\theta^{(1)}$ and $\theta^{(2)}$ are exactly known. We assume that the measurement is conducted at three locations $x = 0, 0.2, 0.4, 0.6, 0.8,$ and 1 for each sample piece. With the above setting, we have six labeled training point in $\mathcal{D}_1$. To show the performance of the proposed method against the sizes of the two groups of data, the BASL is implemented first by selecting four points from the six measured points with the size of $\mathcal{D}_2$ varying from zero to eight, and then all the six measured points are utilized for inference.

The results for the expectation of model response, including the posterior means and CIs, are shown in Figure 3, together with the analytical solution (the red solid lines) for comparison. It is seen from Figure 3 that, with two measurements at the two end points, the resultant posterior variance always approach zero, but the posterior mean shows large bias compared to the reference value between the two end points, even when the number M of the PDE data increases, indicating that the global behavior of the expectation of model response cannot be recovered with only the four measurement points. When eight measurement points ($N = 8$) are utilized, it can be seen from Figure 3 that, with no PDE data, the results have been improved compared to the case of $N = 4$, and with $M = 4$ and $M = 8$ PDE grid points being added, the posterior mean of expectation shows perfect match with the reference values, and the posterior CIs are very narrow. As revealed by the third row of Figure 3, the corresponding results are all improved compared with the above two cases, indicating the correctness and robustness of the inference.

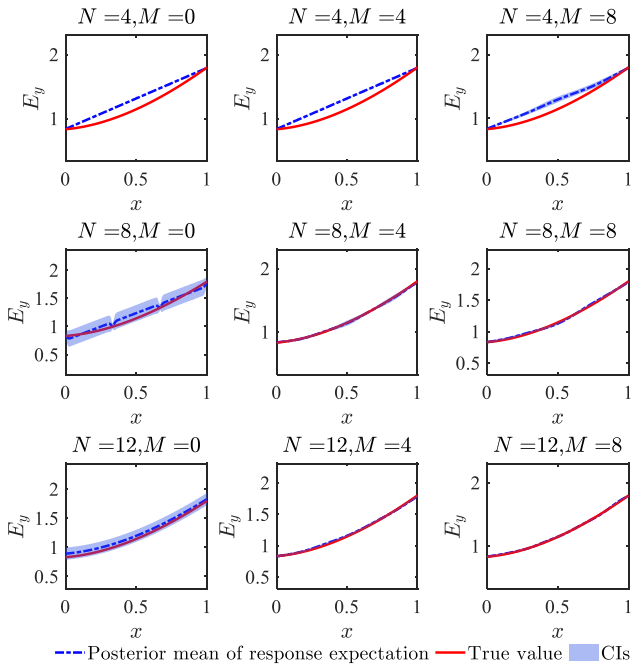

Figure 3: Results for case 1 of the first toy example with different settings of training data size: posterior features of expectation of model response, accompanied with the true curve for comparison.

Next, we check the results for the variance of response, which are reported in Figure 4. It can be seen that, the reference value of variance approaches zero almost everywhere, and all posterior means of variance inferred by the proposed method BASL with different size of measurement data and PDE data successfully capture this kind of performance. Let us start from the first row of Figure 4 with $N = 4$, when $M = 0$ and $M = 4$, the posterior variances are very small, it could probably be a coincidence as the global performance is not catched entirely (revealed by the first row of Figure 4). With more measurement data added to the training set, the posterior variance has been reduced, such as $N = 12$ with $M = 8$, indicating effectiveness of the BASL method.

We then investigate the posterior feature of the auxiliary function $y_a(\theta, x)$ with $N = 12$ and $M = 8$, which is reported in Figure 5, together with the analytical results for comparison and the two groups of training data. As can be seen, the posterior mean of $y_a(\theta, x)$ match well with the true values with small posterior Standard Deviations (STDs) almost anywhere except the end points of θ, demonstrating the effectiveness of the proposed methods for inference of the auxiliary function in the augmented space.

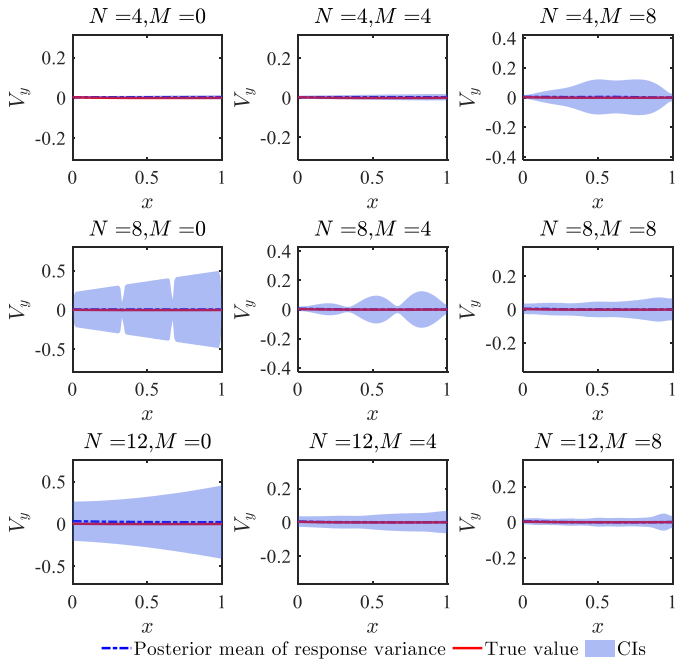

Figure 4: Results for case 1 of the first toy example with different settings of training data size: posterior features of variance of model response, accompanied with the reference curve for comparison.

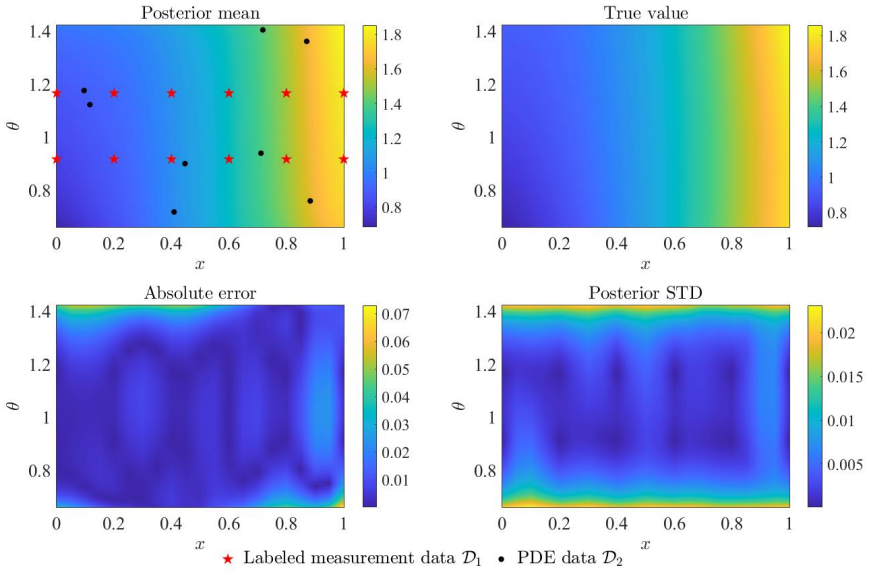

Figure 5: Results of the auxiliary function $y_a(\theta, x)$ for case 1 of the first toy example with $N = 12$ and $M = 8$: posterior mean accompanied with training data (top left), analytical result (top right), absolute error of posterior mean (left bottom), and the posterior STandard Deviation (STD).

Next, we consider the second case, where the same twelve measurement data is utilized, but the values of the $\theta^{(1)}$ and $\theta^{(2)}$ are assumed to be unknown. The posterior results for the expectation and variance of model response are then reported in Figure 6 with $M = 4$ and $M = 8$ PDE points respectively. It is shown that the posterior mean of expectation and variance function matches well with the analytical results, but the posterior variance is a little bit higher than the one for the first case (see the last row of Figure 3 and Figure 4. This is absolutely reasonable as the information of θ in the training data $\mathcal{D}_1$ is missing, resulting in higher uncertainty for the Bayesian inference of quantities related to such kind of information. A similar phenomenon exists for the posterior features of the auxiliary function $y_a(\theta, x)$, as shown by Figure 7.

For case 2, the byproducts, i.e., the MLE estimations of the $\theta^{(1)}$ and $\theta^{(2)}$ with sample size varying, are summarized in Table 1. Obviously, the identified values in each of the three implementations match well with the true values, indicating the effectiveness of the BASL method for identifying the unknown parameter values for this example.

Table 1: Results of parameter identification for case 2 of the first toy example.

Data size	MLE estimations	True values
$N = 12, M = 0$	$\theta^{(1)} = 0.9554$, $\theta^{(2)} = 1.1612$	
$N = 12, M = 4$	$\theta^{(1)} = 0.9152$, $\theta^{(2)} = 1.1589$	$\theta^{(1)} = 0.9157$, $\theta^{(2)} = 1.1189$
$N = 12, M = 8$	$\theta^{(1)} = 0.9130$, $\theta^{(2)} = 1.1612$	

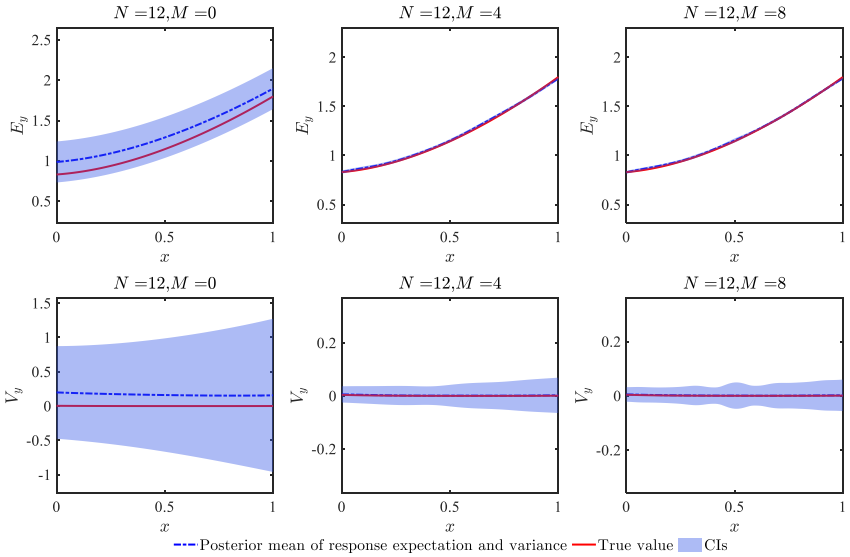

Figure 6: Results of the expectation and variance of model response for case 2 of the first toy example, accompanied with the analytical result for comparison.

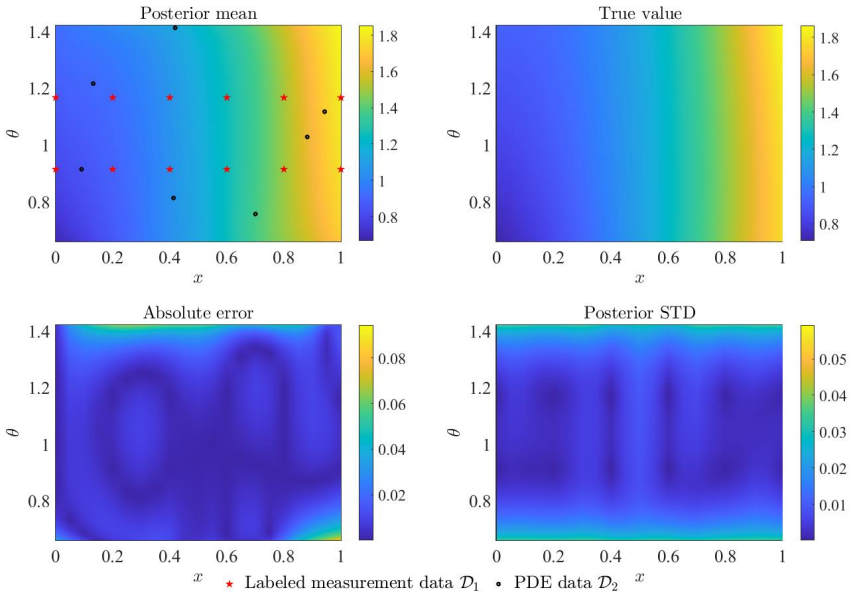

Figure 7: Posterior features of the auxiliary function $y_a(\theta, x)$ for case 2 of the first toy example with $N = 12$ and $M = 8$, accompanied with analytical results and training data.

5.2. A linear ODE with nonlinear response behaviour

The model response function of the first example shows approximately linear behaviour compared with respect to the spatial variable. With this toy example, we consider the performance of the BASL method for problems with nonlinear response function. A physical system described by the following ODE is of interest:

$$\mathcal{F}_X[y|\theta](x) := 121y(x) + \frac{1}{\exp(\theta)}\frac{\partial y(x)}{\partial x} + \frac{\partial^2 y(x)}{\partial x^2} = 11\cos(11x+2), \tag{35}$$

where $x \in [0, 0.5]$ is a spatial variable, and θ is a random model parameter following standard normal distribution. For the labeled measurement data $\mathcal{D}_1$, it is assumed that the measurement is conducted on nine sample pieces with the values of parameters being $\boldsymbol{\theta} = (-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2)$.

For case 1, the results of the expectation and variance of model response function with varying sizes for both groups of training data are shown in Figure 8 and Figure 9, together with the analytical solutions for comparison. One notes that the analytical solution is generated by applying specific boundary conditions, which are also included in the measurement labeled data $\mathcal{D}_1$. We first check the results for the expectation of response, which are shown in Figure 8. It is seen from the first column of Figure 8 that, with no PDE points added, the inferred posterior features of the expectation possess large variance, indicating the trained GPR model fails to capture the behavior of the response expectation. When $M = 8$ PDE grid points, generated with LHS, are added, as can be seen from the second column of Figure 8, the prediction accuracy has been comprehensively improved. When $M = 15$ PDE points are added, the posterior variance is sufficiently small,

and the posterior mean matches well with the analytical solution. The results for the response variance are reported in Figure 9. As can be seen, in terms of estimation accuracy, these results show similar features with those for the response expectation, and thus we can then conclude that the BALS method performs well for the response variance of this example.

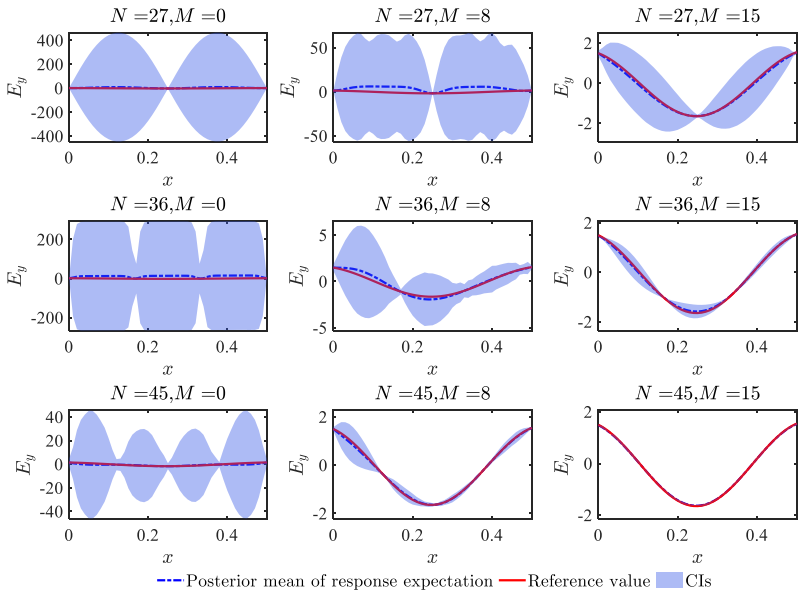

Figure 8: Posterior features of the expectation of the response function for case 1 of the second toy example.

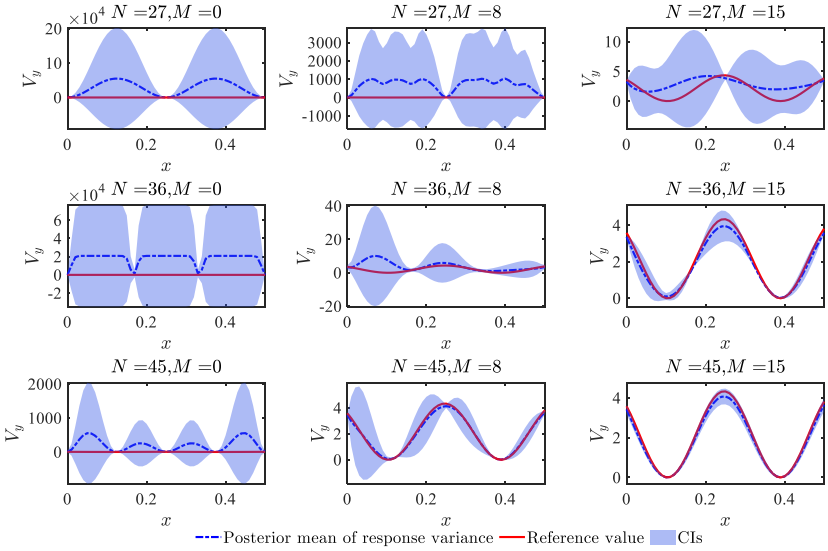

Figure 9: Posterior features of the response variance for case 1 of the second toy example.

The posterior features of the auxiliary function $y_a(\theta, x)$ inferred with $N = 45$ and $M = 15$ are then displayed in Figure 10, together with the analytical solutions (generated with the same boundary conditions involved in $\mathcal{D}_1$) and the two groups of training data. As shown by the two bottom panels, the gap between the posterior mean and the analytical solution is small almost anywhere, and the posterior STDs are also small in almost all areas bounded by $\theta \in [-2, 2]$ and $x \in [0, 1]$, revealing the correctness and robustness of the BASL method for learning the auxiliary function despite the non-linearity of the model response function.

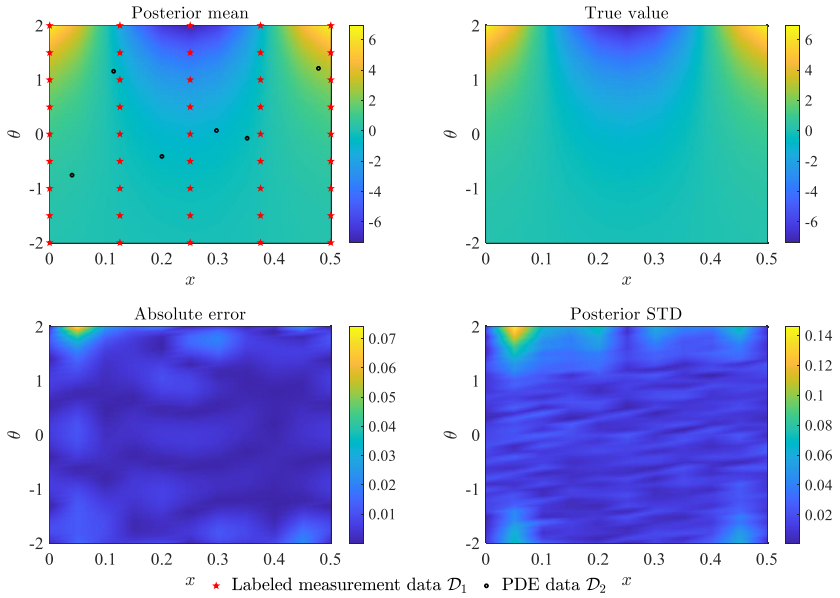

Figure 10: Posterior features of the auxiliary function $y_a(\theta, x)$ for case 1 of the second toy example.

We then implement the BASL method to case 2 where the parameters values $\theta^{(1)} \sim \theta^{(9)}$ are unknown. It is assumed that for each sample piece, five points at specific locations are measured, resulting in totally $N = 45$ measurement points for $\mathcal{D}_1$. The results with $M = 0$, $M = 8$ and $M = 15$ are then shown in Figure 11. As can bee seen, given no PDE points, the inferred results for both response expectation and variance show big difference with the reference solutions, while PDE grid points are added, the results show much better agreement with the analytical solution. As shown by the last column of Figure 11, with $N = 45$ measurement points and $M = 15$ PDE points, the inferred results for both response expectation and variance are of high accuracy and high robustness. The results of the auxiliary function $y_a(\theta, x)$ for case 2 are reported in Figure 12, which also show good agreement with the analytical solutions with very small posterior variation.

For case 2, the MLE estimates of the parameters values for the nine test pieces are displayed in Table 2. As shown, with $N = 45$ measurement points and $M = 15$ PDE points, the accuracy of the estimates is

acceptable compared with the reference values.

Table 2: Results of parameter identification for case 2 of the second toy example.

Data size	MLE estimations
$N = 45, M = 0$	$\boldsymbol{\theta}^* = (-2.1812, -1.1661, -1.1051, -0.6124, 0.0887, 0.7370, 1.2637, 2.4763, 1.7374)$
$N = 45, M = 8$	$\boldsymbol{\theta}^* = (-1.9839, -1.5611, -1.1866, -0.6117, 0.1778, 0.6251, 0.9655, 1.3044, 2.0621)$
$N = 45, M = 15$	$\boldsymbol{\theta}^* = (-1.9765, -1.4706, -1.0943, -0.5493, 0.0591, 0.4298, 0.8831, 1.5283, 2.0218)$
True value	$\boldsymbol{\theta} = (-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2)$

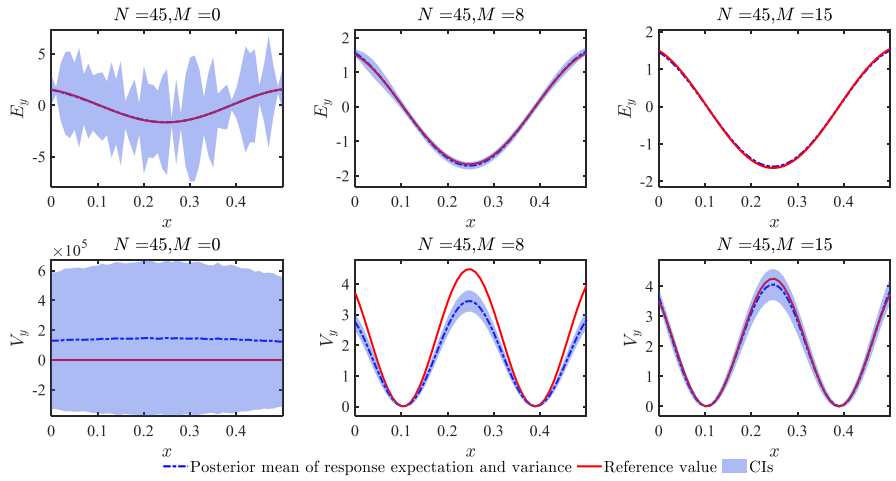

Figure 11: Posterior features of the variance of the response function for case 2 of the second toy example.

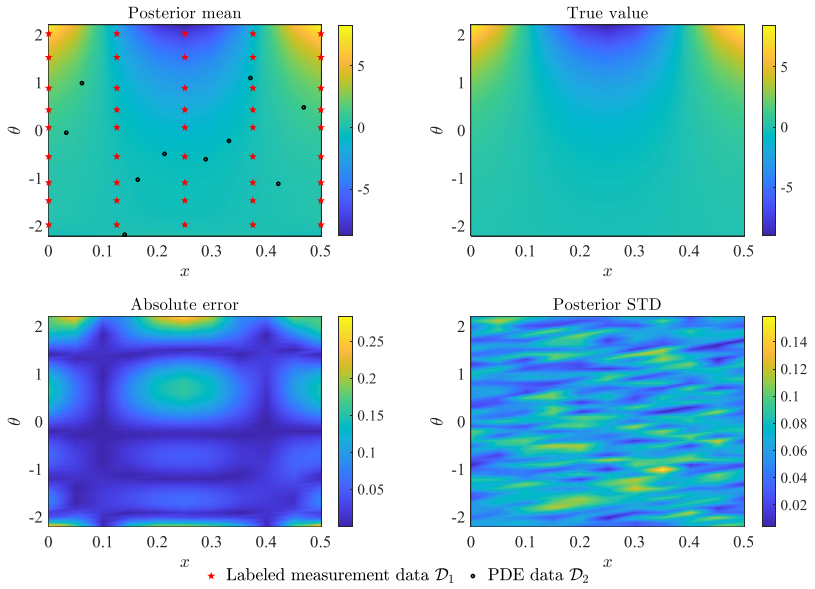

Figure 12: Posterior features of the auxiliary function $y_a(\theta, x)$ for case 2 of the second toy example.

5.3. A heat transfer problem

Consider the heat transfer process on a one-dimensional bar with length l (m), let x (m) and t (s) denote the spatial and temporal variables respectively, the governing PDE is formulated as:

$$\frac{\partial y(x,t)}{\partial t} - a^2 \frac{\partial^2 y(x,t)}{\partial x^2} = f, f = x(l-x) = \frac{f_0}{c} \tag{36}$$

, with initial condition $y(0,t) = y(1,t) = 0$ for $t > 0$, and boundary condition $y(x,0) = 0$ for $0 \leqslant x \leqslant l$, where $a = \frac{k}{c\rho}$ with k, c and ρ being the heat conductivity coefficient, specific heat capacity, and material density respectively, $f_0 = cx(l-x)$ is the intensity of heating source which varies with x. The density ρ (kg/m^3), specific heat capacity c (J/(kg $\cdot$ K)) and heat conductivity coefficient k (W/(m $\cdot$ K)) are random model parameters follow log-normal distribution, i.e., $\ln \rho \sim \mathcal{N}\left(9.0936, 0.0180^2\right)$, $\ln c \sim \mathcal{N}\left(5.8732, 0.4001^2\right)$ and $\ln k \sim \mathcal{N}\left(5.9198, 0.3853^2\right)$. The closed-form solution of this problem can be analytically generated using Fourier series. We consider the expected temperature in Celsius degree of the bar within time period from zero to 60 seconds.

The labeled data $\mathcal{D}_1$ is generated by measurement on five test pieces, for each of which, at seven spatio-temporal points. The values of the random parameters for these two pieces are set to be as displayed in Table 3. With the above setting, $N = 35$ labeled data are obtained for $\mathcal{D}_1$. With $M = 20$ PDE points being added, the inferred posterior features of the expectation $E_y(x,t)$ and variance $V_y(x,t)$ generated by BASL procedure are shown in Figure 13 and and Figure 14 respectively, together with the projections of the training data in the spatio-temporal space, and the reference solution for comparison. By checking the

absolute error between the posterior mean and the reference value in the left bottom panel of Figure 13, it is seen that the result is of high accuracy across almost the full spatio-temporal space. It is also seen from the right bottom panel that STDs are also very small across the space, indicating the robustness and high credibility of the inference results. As can be seen from Figure 14, the response variance is also estimated with high accuracy.

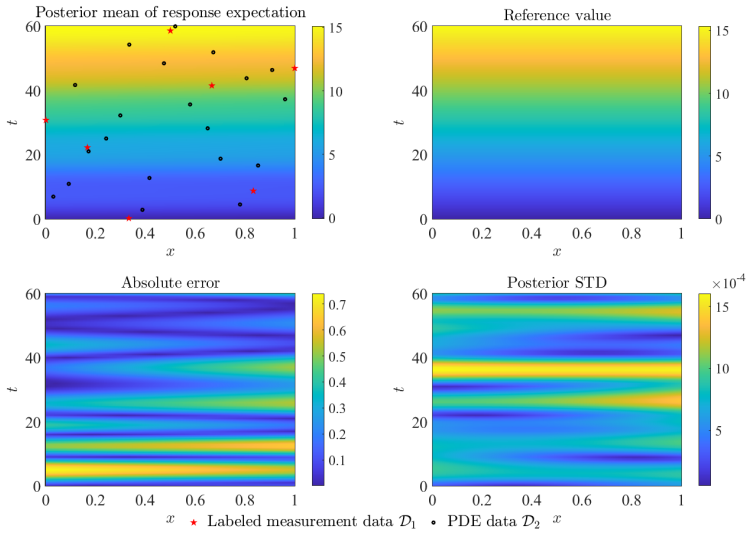

Figure 13: Posterior features of the expectation $E_y(x,t)$ for case 1 of the heat transfer problem.

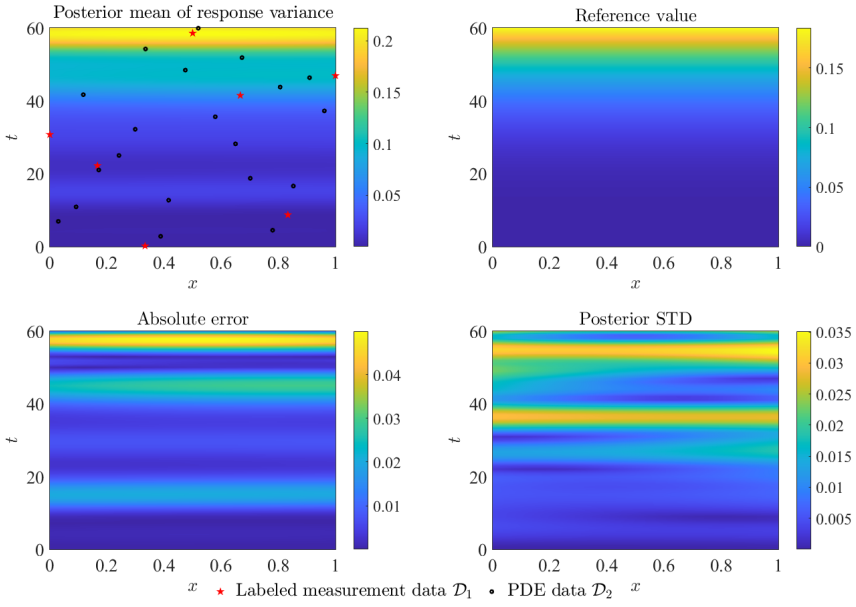

Figure 14: Posterior features of the variance $V_y(x,t)$ for case 1 of the heat transfer problem.

We then consider case 2 where all the values of the three parameters for each of the five test pieces are unknown. The results for the expectation and variance of model response are shown in Figure 15 and Figure 16, and the byproducts, i.e., the results for parameter identification, are displayed in Table 3 in comparison with the reference values. Comparing Figure 15 with Figure 13, and Figure 16 with Figure 14, it is found that, with the same training data, the posterior features for case 2 show higher errors compared with those for case 1, as revealed by the absolute errors and posterior STDs. This phenomenon is the same as that in the last two examples, and still due to the loss of parameters information in the training data $\mathcal{D}_1$. However, the results for case 2 are still of high credibility and high robustness. Checking Table 3, it is seen that, except $\rho^{(1)}$, all the other three parameter values are accurately evaluated by MLE. As for $\rho^{(1)}$, the estimation shows higher error, but this does not bring much negative effect for probabilistic analysis of model response, as indicated by the results in Figure 15 and Figure 16.

Table 3: Results of parameter identification for case 2 of heat transfer problem.

Parameters	MLE estimations	True values
$\boldsymbol{\theta}^{(1)} = \left(\rho^{(1)}, c^{(1)}, k^{(1)}\right)$	$(9020.5324, 125.1419, 293.8622)$	$(9029.1307, 124.1629, 283.1316)$
$\boldsymbol{\theta}^{(2)} = \left(\rho^{(2)}, c^{(2)}, k^{(2)}\right)$	$(8830.6821, 456.3920, 334.5399)$	$(8842.8799, 460.2669, 323.8227)$
$\boldsymbol{\theta}^{(3)} = \left(\rho^{(2)}, c^{(2)}, k^{(2)}\right)$	$(9121.4572, 339.4032, 375.7894)$	$(8919.0498, 348.2860, 367.4393)$
$\boldsymbol{\theta}^{(4)} = \left(\rho^{(2)}, c^{(2)}, k^{(2)}\right)$	$(8652.8938, 336.6059, 378.4173)$	$(9136.8939, 341.5234, 389.8148)$
$\boldsymbol{\theta}^{(5)} = \left(\rho^{(2)}, c^{(2)}, k^{(2)}\right)$	$(8932.1081, 285.2905, 456.9316)$	$(8658.0533, 285.8450, 444.4482)$

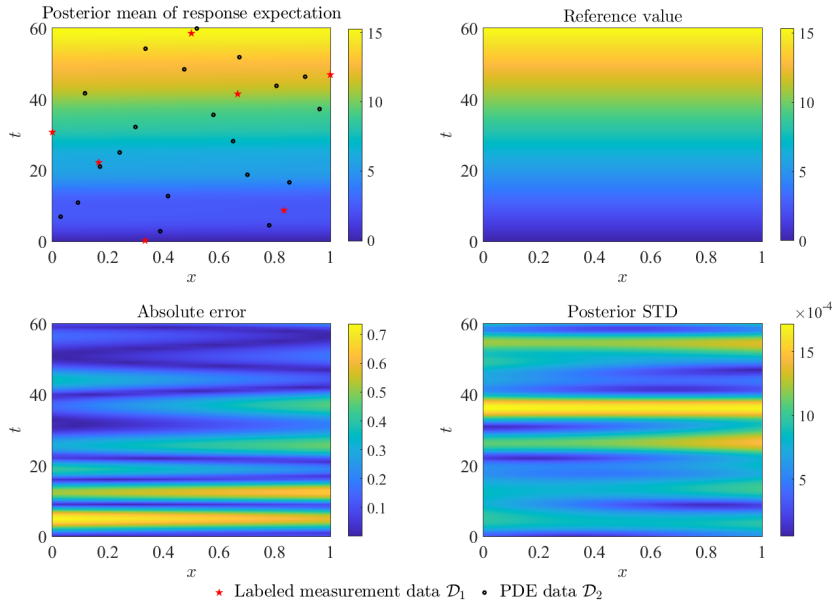

Figure 15: Results of the expectation $E_y(x,t)$ of model response for case 2 of the heat transfer problem.

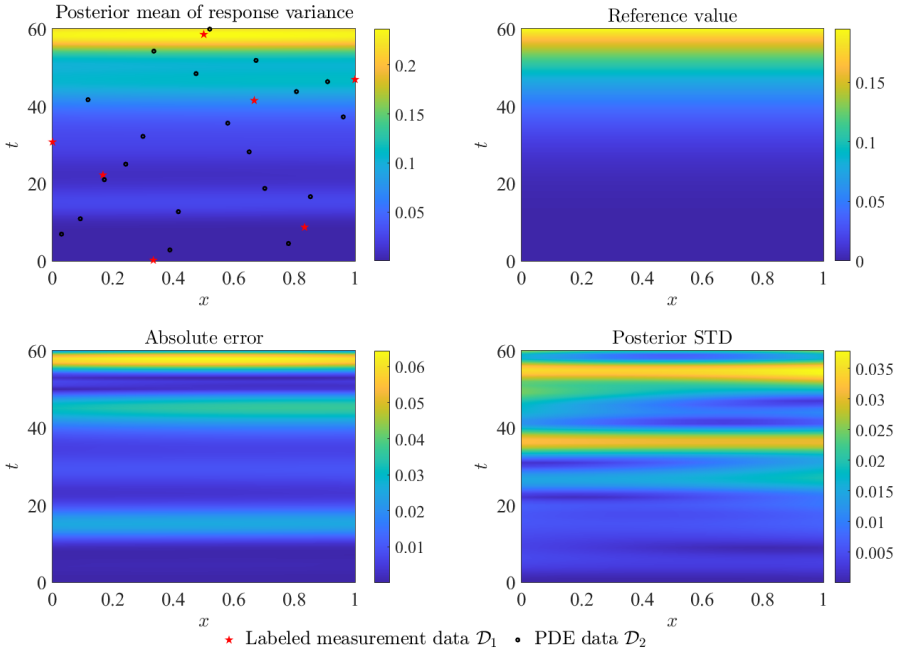

Figure 16: Results of the variance $V_y(x,t)$ of model response for case 2 of the heat transfer problem.

5.4. Deflection prediction of a plate under uniform pressure load

We then apply the BASL procedure to the deflection prediction of a plate subjected to uniform pressure, which is adapted from the Matlab PDE toolbox. The governing PDE of this problem is formulated as:

$$\nabla^2 \left(D\nabla^2\right) w = -p \tag{37}$$

, where w indicates the transverse deflection of the plate, p implies the applied pressure load, and D refer to the bending stiffness of the plate and can be calculated by $D = Eh^3/\left(12\left(1-\nu^2\right)\right)$, with E being the modulus of elasticity, ν being the Poisson's ratio, and h being the plate thickness. The plate is assumed to be square with the edge length being 10, and all the four edges clamped. The boundary condition for the clamped edges are then set to be $w = 0$ and $w' = 0$ with w' being the gradient of w in the direction normal to the boundary. All the model parameters and the load are assumed to be deterministic in the Matlab PDE toolbox, and here, we assume that the three model parameters E, ν and h as well as the uniform pressure p are all random variables with probability distribution stated in Table 4. The aim of the investigation is then to estimate the expected deflection $\mathbb{E}_{E,\nu,h,p}\left[w\left(x,y,E,\nu,h,p\right)\right]$ and its variance $\mathbb{V}_{E,\nu,h,p}\left[w\left(x,y,E,\nu,h,p\right)\right]$ of the plate given the four random inputs. To validate the results, the reference solution is computed by the non-intrusive scheme combining the MCS and FEM analysis, and to ensure the correctness of the reference solution, the MCS sample size is set to be 10^5.

We use two test pieces to generate the measurement data $\mathcal{D}_1$, and for each piece, 8 points are generated on the boundary edges with response values known to be zero, and the other five internal training points are generated by measurement, as partially indicated by the pink star points in the first panel of Figure 17 and Figure 18. The second group of data, i.e., the PDE data $\mathcal{D}_2$ consisting of 16 points, are generated by LHS design, as shown by the black solid dots in the first panel of Figure 17 and Figure 18. One notes that, for this example, one more source of information available for Bayesian learning is that, on the four boundary edges, the gradients of the response w in the direction normal to the boundary equal to zero. We pick 8 training points from such kind of boundary condition to form the third group of training data $\mathcal{D}_3$, as shown by the blue solid blocks in the first panel of Figure 17 and Figure 18. For each point in $\mathcal{D}_3$, the label is that the gradient (normal to the boundary) equals zero.

We then consider the first case where the parameter values for the two test pieces are exactly known. Based on the Gaussian prior assumption on the response function $w\left(x,y,E,\nu,h,p\right)$, the response values contained in $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ jointly follow multivariate Gaussian distribution, based on which the hyper-parameters of the GPR model are evaluated with the MLE method. The resultant posterior mean and STD of the expected deflection $\mathbb{E}_{E,\nu,h,p}\left[w\left(x,y,E,\nu,h,p\right)\right]$ and its variance $\mathbb{V}_{E,\nu,h,p}\left[w\left(x,y,E,\nu,h,p\right)\right]$ of the plate are then reported in the first and fourth panels of Figure 17 and Figure 18 respectively, together with the reference solution in the top right panel, and the corresponding absolute difference between the posterior

mean prediction and the reference solution in the bottom left panel. As can be seen from Figure 17, the relative absolution error of the BASL estimation in most area is less than 10%, and in the central region around $x = y = 5$ which is of special concern, the absolute error is less than 5%, and the posterior COV is also less than 5%, indicating the high accuracy and high robustness of the BASL method for case 1 of this example. Comparing Figure 18 with 17, it can be found that the estimation accuracy for the response variance is not as good as that for the response expectation, but is acceptable. If the accuracy for variance needs to be further improved, one can add more points (either measurement points or PDe points) to achieve this goal.

Table 4: Distribution information of the input variables of the plate model.

Input variables	Distribution type	Distribution parameters	units
E	Lognormal	$\mu = 10^6$, $\sigma = 2 \times 10^4$	Pa
ν	Lognormal	$\mu = 0.3$, $\sigma = 0.006$	–
h	Uniform	$[0.09, 0.11]$	m
p	Uniform	$[1.95, 2.05]$	N/m^2

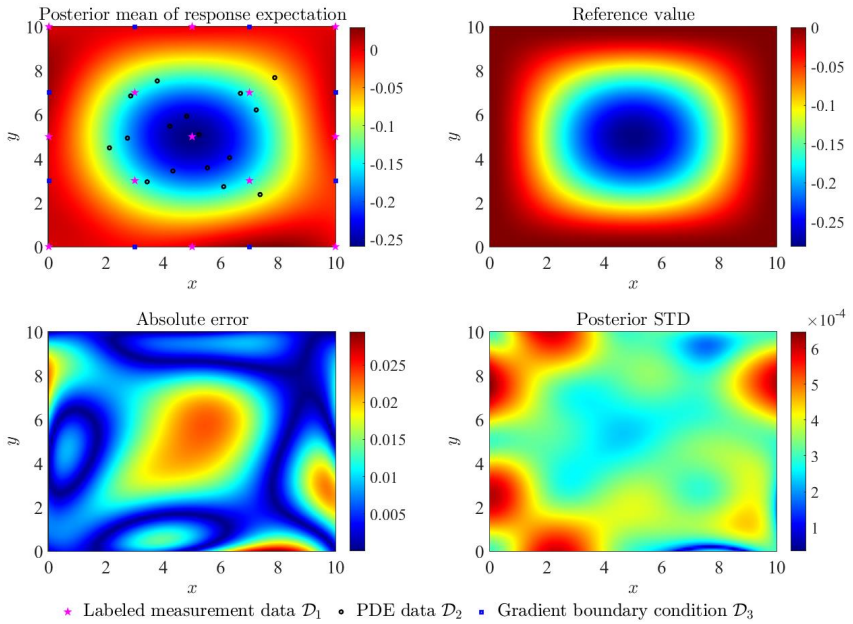

Figure 17: Results of the expected deflection $\mathbb{E}_{E,\nu,h,p}\left[w\left(x, y, E, \nu, h, p\right)\right]$ for case 1 of the plate model.

We then consider case 2 where the exact values of the two parameters E and ν for the two test pieces

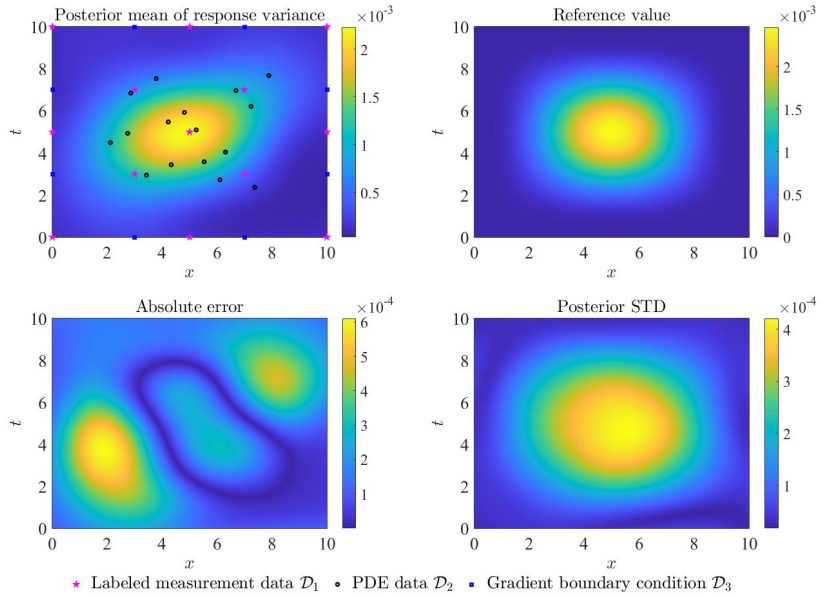

Figure 18: Results of the deflection variance $\mathbb{V}_{E,\nu,h,p}\left[w\left(x,y,E,\nu,h,p\right)\right]$ for case 1 of the plate model.

are listed in the last column of Table 5, but not known, and need also to be inferred during the probabilistic analysis process. The resultant posterior features for the expected deflection are then reported in Figure 19 and Figure 20, together with the reference solution as well as the absolute difference between the posterior mean prediction and reference solution. As can be seen, the posterior features show similar characteristics as those for case 1, demonstrating the effectiveness of the BASL method for case 2 of this example. The MLE estimation of the parameters values, as byproducts of BASL analysis, are also reported in the second column of Table 5, which show good agreement with the true values reported in the last column. With the above analysis, it can be concluded that both the probabilistic analysis and parameter identification problems have been properly addressed with only one run of BASL.

Table 5: Results of parameter identification for case 2 of the plate model.

Parameters	MLE estimations	True values
$\boldsymbol{\theta}^{(1)} = \left(E^{(1)}, \nu^{(1)}\right)$	$(9.8000 \times 10^5,\ 0.3060)$	$(9.8104 \times 10^5,\ 0.3056)$
$\boldsymbol{\theta}^{(2)} = \left(E^{(2)}, \nu^{(2)}\right)$	$(1.0098 \times 10^6,\ 0.2964)$	$(1.0077 \times 10^6,\ 0.2962)$

6. Conclusions and prospects

A one-step Bayesian machine learning method, termed as BASL, has been devised for probabilistic response analysis of physical models subjected to random inputs. The BASL method combines the labeled measurement data (generated by measurement or high-fidelity computational models) and the PDE data

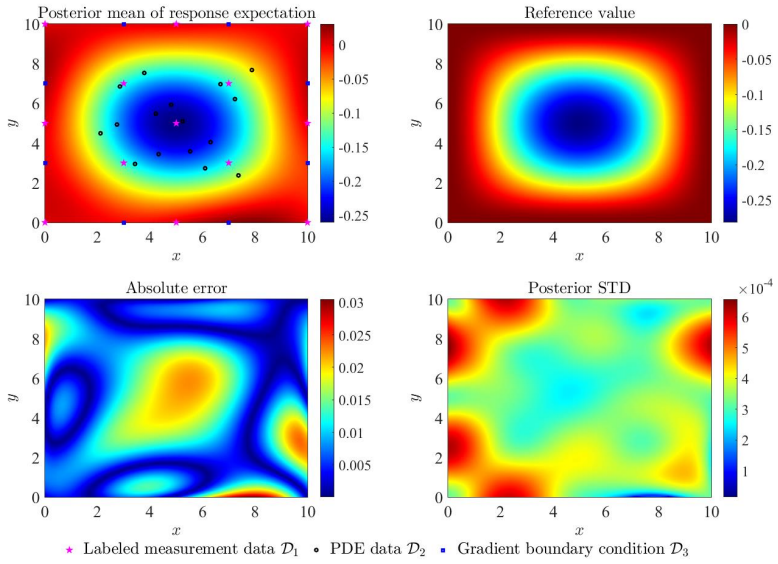

Figure 19: Results of the expected deflection $\mathbb{E}_{E,\nu,h,p}\left[w\left(x,y,E,\nu,h,p\right)\right]$ for case 2 of the plate model.

(with spatio-temporal points generated randomly or by active design) for inference. The developed method is devised by first reformulating the PDE with random parameters as an equivalent deterministic PDE based on the idea of augmented space, posing a Gaussian process prior assumption on the auxiliary function $y_a\left(\boldsymbol{\theta},\boldsymbol{x}\right)$ with both the spatio-temporal variables $\boldsymbol{x}$ and random model parameters $\boldsymbol{\theta}$ as arguments, and then inferring the hyper-parameters of the Gaussian process model by maximising the likelihood function established based on both the measurement and PDE data. With the above procedures, a posterior GPR model $\hat{y}_a\left(\boldsymbol{\theta},\boldsymbol{x}\right)$ can be generated, based on which, the posterior features for all probabilistic descriptors of model response can be derived. The BASL method is further extended to case 2, where some values of the model parameters of the test pieces used for measurement are unknown or not precisely known, resulting in the necessity for parameter identification. The extension enables the probabilistic analysis and parameters identification to be realized in one step. Finally, with the expectation and variance of model response as examples, the results of the two toy examples and two real-world physical systems demonstrate the effectiveness of the proposed methods for both cases.

The BASL method brings new opportunities for promoting the probabilistic analysis of physical model responses in a more efficient way of fusing multi-source information. Compared to the traditional double-loop and non-intrusive scheme, BASL is more numerically efficient as its main cost is caused by the MLE procedure for optimizing the hyper-parameters of the GPR model. It also allows to solve the "forward" and "backward" UQ problems in one step, which is a significant advantage over the traditional scheme. Another strength is that it merges two types of data for inference. Besides, it also brings opportunity for other types of applications. For example, developing active design strategy for generating both sources of training data

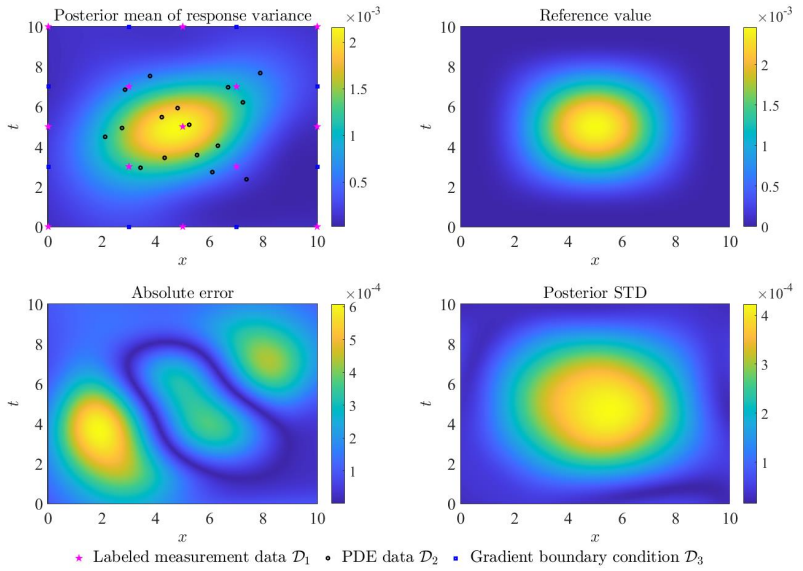

Figure 20: Results of the variance of deflection $\mathbb{V}_{E,\nu,h,p}\left[w\left(x,y,E,\nu,h,p\right)\right]$ for case 2 of the plate model.

can be appealing. For the labeled data $\mathcal{D}_1$, if it is generated from measurement, active design of the training points can be instructive for placing measurement sensors; if it is generated by calling expensive high-fidelity computational models, the number of model calls can be reduced with active design. For the PDE data $\mathcal{D}_2$, active design can also be of value for reducing the number of training points especially when only some partial response information (e.g., the maximum stress instead of the full stress field) is of interest.

It should be noted that, given the development in this work, the BASL does have some limitations. First, it is not suitable for problems with nonlinear PDEs; second, it is not applicable to problems with very high-dimensional uncertain variables due to the limitation of the GPR model in high-dimensional space; third, the method may not applicable to very large-scale problems with varying local behaviors. All these limitations will be specifically treated in our future work. Despite these limitations, the proposed method opens a door for performing "forward" and "backward" UQ in one step, with the combination of physical equations and measurement data. This scheme is completely different from the traditional double-loop and non-intrusive scheme.

Acknowledgment

Pengfei Wei acknowledges the supports of the National Natural Science Foundation of China under grant number 72171194 and the Sino-German Mobility Programme under grant number M-0175 (2021-2023). Matthias Faes acknowledges the support of the Research Foundation Flanders (FWO) under grant 12P3519N, as well as of the Humboldt foundation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this pap.

References

[1] J. C. Helton, J. D. Johnson, W. L. Oberkampf, C. J. Sallaberry, Representation of analysis results involving aleatory and epistemic uncertainty, International Journal of General Systems 39 (6) (2010) 605–646.

[2] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, Mechanical Systems and Signal Processing 37 (1-2) (2013) 4–29.

[3] M. G. Faes, M. Daub, S. Marelli, E. Patelli, M. Beer, Engineering analysis with probability boxes: A review on computational methods, Structural Safety 93 (2021) 102092.

[4] U. Tripathi, S. Garg, R. Nayek, S. Chakraborty, Physics-integrated deep learning for uncertainty quantification and reliability estimation of nonlinear dynamical systems, Probabilistic Engineering Mechanics (2023) 103419.

[5] J. Song, W.-H. Kang, Y.-J. Lee, J. Chun, Structural system reliability: Overview of theories and applications to optimization, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 7 (2) (2021) 03121001.

[6] A. Bhaduri, D. Brandyberry, M. D. Shields, P. Geubelle, L. Graham-Brady, On the usefulness of gradient information in surrogate modeling: Application to uncertainty propagation in composite material models, Probabilistic Engineering Mechanics 60 (2020) 103024.

[7] W. L. Oberkampf, C. J. Roy, Verification and validation in scientific computing, Cambridge University Press, 2010.

[8] M. H. Kalos, P. A. Whitlock, Monte carlo methods: Second revised and enlarged edition, Monte Carlo Methods: Second Revised and Enlarged Edition (2008).

[9] M. Hunt, B. Haley, M. McLennan, M. Koslowski, J. Murthy, A. Strachan, PUQ: A code for non-intrusive uncertainty propagation in computer simulations, Computer Physics Communications 194 (2015) 97–107.

[10] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Miguez, P. M. Djuric, Adaptive importance sampling: The past, the present, and the future, IEEE Signal Processing Magazine 34 (4) (2017) 60–79.

[11] S.-K. Au, On mcmc algorithm for subset simulation, Probabilistic Engineering Mechanics 43 (2016) 117–120.

[12] W. Li, R. Chen, Z. Tan, Efficient sequential monte carlo with multiple proposals and control variates, Journal of the American Statistical Association 111 (513) (2016) 298–313.

[13] C. J. Oates, M. Girolami, N. Chopin, Control functionals for monte carlo integration, Journal of The Royal Statistical Society Series B-statistical Methodology 79 (3) (2017) 695–718.

[14] F.-X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, D. Sejdinovic, et al., Probabilistic integration: A role in statistical computation?, Statistical Science 34 (1) (2019) 1–22.

[15] X. Zhang, M. D. Pandey, Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method, Structural Safety 43 (2013) 28–40.

[16] B. Keshtegar, Z. Meng, A hybrid relaxed first-order reliability method for efficient structural reliability analysis, Structural Safety 66 (2017) 84–93.

[17] H. N. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, Annual Review of Fluid Mechanics 41 (1) (2009) 35–52.

[18] S. Oladyshkin, W. Nowak, Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion, Reliability Engineering and System Safety 106 (2012) 179–190.

[19] S. Freitag, P. Edler, K. Kremer, G. Meschke, Multilevel surrogate modeling approach for optimization problems with polymorphic uncertain parameters, International Journal of Approximate Reasoning 119 (2020) 81–91.

[20] B. Turnquist, M. Owkes, multiUQ: A software package for uncertainty quantification of multiphase flows, Computer Physics Communications 268 (2021) 108088.

[21] G. Chen, J. Zhou, Boundary element methods with applications to nonlinear problems, Vol. 7, Springer Science & Business Media, 2010.

[22] S. Li, W. K. Liu, Meshfree and particle methods and their applications, Applied Mechanics Reviews 55 (1) (2002) 1–34.

[23] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics 378 (2019) 686–707.

[24] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nature Reviews Physics 3 (6) (2021) 422–440.

[25] Y. Yang, P. Perdikaris, Adversarial uncertainty quantification in physics-informed neural networks, Journal of Computational Physics 394 (2019) 136–152.

[26] M. I. Katsidoniotaki, A. F. Psaros, I. A. Kougioumtzoglou, Uncertainty quantification of nonlinear system stochastic response estimates based on the wiener path integral technique: A bayesian compressive sampling treatment, Probabilistic Engineering Mechanics 67 (2022) 103193.

[27] O. Sedehi, C. Papadimitriou, L. S. Katafygiotis, Data-driven uncertainty quantification and propagation in structural dynamics through a hierarchical bayesian framework, Probabilistic Engineering Mechanics 60 (2020) 103047.
[28] Y. Chen, B. Hosseini, H. Owhadi, A. M. Stuart, Solving and learning nonlinear pdes with gaussian processes, Journal of Computational Physics (2021) 110668.
[29] M. Raissi, P. Perdikaris, G. E. Karniadakis, Machine learning of linear differential equations using gaussian processes, Journal of Computational Physics 348 (2017) 683–693.
[30] F. Tronarp, S. Särkkä, P. Hennig, Bayesian ode solvers: the maximum a posteriori estimate, Statistics and Computing 31 (3) (2021) 1–18.
[31] L. Yang, X. Meng, G. E. Karniadakis, B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data, Journal of Computational Physics 425 (2021) 109913.
[32] J. Chen, Z. Chen, C. Zhang, C. Jeff Wu, APIK: Active physics-informed kriging model with partial differential equations, SIAM/ASA Journal on Uncertainty Quantification 10 (1) (2022) 481–506.
[33] P. Wei, J. Song, S. Bi, M. Broggi, M. Beer, Z. Lu, Z. Yue, Non-intrusive stochastic analysis with parameterized imprecise probability models: I. performance estimation, Mechanical Systems and Signal Processing 124 (2019) 349–368.
[34] M. G. Faes, M. A. Valdebenito, X. Yuan, P. Wei, M. Beer, Augmented reliability analysis for estimating imprecise first excursion probabilities in stochastic linear dynamics, Advances in Engineering Software 155 (2021) 102993.
[35] S. H. Cheung, J. L. Beck, Bayesian model updating using hybrid monte carlo simulation with application to structural dynamic models with many uncertain parameters, Journal of Engineering Mechanics-asce 135 (4) (2009) 243–255.
[36] R. Rocchetta, M. Broggi, Q. Huchet, E. Patelli, On-line bayesian model updating for structural health monitoring, Mechanical Systems and Signal Processing 103 (2018) 174–195.
[37] C. E. Rasmussen, C. Williams, Gaussian processes for machine learning, vol. 1, MIT press 39 (2006) 40–43.
[38] S. Theodoridis, Machine learning: a Bayesian and optimization perspective, Academic press, 2015.
[39] P. Wei, F. Hong, K.-K. Phoon, M. Beer, Bounds optimization of model response moments: a twin-engine bayesian active learning method, Computational Mechanics 67 (5) (2021) 1273–1292.
[40] S. Atkinson, N. Zabaras, Structured bayesian gaussian process latent variable model: Applications to data-driven dimensionality reduction and high-dimensional inversion, Journal of Computational Physics 383 (2019) 166–195.
[41] J. Nocedal, S. J. Wright, Numerical optimization, Springer, 1999.
[42] C. E. Rasmussen, Z. Ghahramani, Bayesian Monte Carlo, Advances in neural information processing systems (2003) 505–512.
[43] P. Wei, X. Zhang, M. Beer, Adaptive experiment design for probabilistic integration, Computer Methods in Applied Mechanics and Engineering 365 (2020) 113035.
[44] J. Song, P. Wei, M. Valdebenito, M. Beer, Active learning line sampling for rare event analysis, Mechanical Systems and Signal Processing 147 (2021) 107113.
[45] R. Lebrun, A. Dutfoy, Do Rosenblatt and Nataf isoprobabilistic transformations really differ?, Probabilistic Engineering Mechanics 24 (4) (2009) 577–584.
[46] J. Song, P. Wei, M. A. Valdebenito, M. Faes, M. Beer, Data-driven and active learning of variance-based sensitivity indices with bayesian probabilistic integration, Mechanical Systems and Signal Processing 163 (2022) 108106.
[47] J. B. Nagel, B. Sudret, A unified framework for multilevel uncertainty quantification in bayesian inverse problems, Probabilistic Engineering Mechanics 43 (2016) 68–84.
[48] C. P. Robert, V. Elvira, N. Tawn, C. Wu, Accelerating mcmc algorithms., Wiley Interdisciplinary Reviews: Computational Statistics 10 (5) (2018).