Data-driven and Active Learning of Variance-based Sensitivity Indices with Bayesian Probabilistic Integration

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

[a]*Northwestern Polytechnical University, Xi'an 710072, China*
[b]*Advanced Research Laboratories, Tokyo City University, 1-28-1 Tamazutsumi, Setagaya-ku, Tokyo 158-8557, Japan*
[c]*Institute for Risk and Reliability, Leibniz Universität Hannover, Hannover 30167, Germany*
[d]*Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Av. Padre Hurtado 750, 2562340 Viña del Mar, Chile*
[e]*Department of Mechanical Engineering, KU Leuven, Jan DeNayerlaan 5, St.-Katelijne-Waver 2860,Belgium*
[f]*Institute for Risk and Uncertainty, University of Liverpool, Liverpool L69 7ZF, UK*
[g]*International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, China*

Abstract

Variance-based sensitivity indices play an important role in scientific computation and data mining, thus the significance of developing numerical methods for efficient and reliable estimation of these sensitivity indices based on (expensive) computer simulators and/or data cannot be emphasized too much. In this article, the estimation of these sensitivity indices is treated as a statistical inference problem. Two principle lemmas are first proposed as rules of thumb for making the inference. After that, the posterior features for all the (partial) variance terms involved in the main and total effect indices are analytically derived (not in closed form) based on Bayesian Probabilistic Integration (BPI). This forms a data-driven method for estimating the sensitivity indices as well as the involved discretization errors. Further, to improve the efficiency of the developed method for expensive simulators, an acquisition function, named Posterior Variance Contribution (PVC), is utilized for realizing optimal designs of experiments, based on which an adaptive BPI method is established. The application of this framework is illustrated for the calculation of the main and total effect indices, but the proposed two principle lemmas also apply to the calculation of interaction effect indices. The performance of the development is demonstrated by a illustrative numerical example and three engineering benchmarks with finite element models.

Keywords: Variance-based sensitivity; Gaussian process regression; Bayesian probabilistic integration; Data-driven; Adaptive experiment design; Posterior variance contribution.

1. Introduction

Nowadays, owing to the rapid development of computation power, scientific computation based on computer simulators (e.g., finite element models) has been widely utilized in both academic research and engineering practice for predicting the behavior of complex systems or structures and aiding the design of new products. However, due to the uncertainties of various sources, the researchers and practitioners have found

---

*Corresponding author at School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University, Xi'an 710072, China
Email address: `pengfeiwei@nwpu.edu.cn` (Pengfei Wei)

it difficult to achieve accurate and robust predictions with the deterministic simulators, and performing uncertainty quantification to properly incorporate those uncertainties in the model predictions has been a common trend in scientific computing [1, 2], and especially, in structural dynamics. As an important sub-task of uncertainty quantification, Sensitivity Analysis (SA) plays an important role in model developments and refinement as it informs the main sources of model prediction uncertainties [3, 4, 5]. This information is extremely useful for directing the future data collection (with the target of effectively reducing the model prediction uncertainty), and for specifying the subset of influential model parameters to be calibrated in finite element (FE) model updating [6].

Specifically, SA aims at attributing the uncertainty present in the model output to the input variables, and this way to measure the contribution of each input variable to the uncertainty of model outputs [7]. Three groups of SA methods have been developed, i.e., local SA, regional SA, and Global SA (GSA), one can refer to Refs.[4] and [5] for comprehensive reviews and comparisons of these methods. The local method measures the sensitivity of each input variable using the local partial derivatives, and it is widely used in the area of structural reliability for measuring the effects of the distribution parameters of input variables on the failure probability [8, 9]. The regional SA aims at quantifying the effects/contributions of the subregions of the distribution support of each input variables to the uncertainty of model outputs, and it can be especially useful for reduction of epistemic uncertainty [10]. The GSA indices are usually defined as the expected change of the statistical features (e.g., variance and density function) of model response when the input variables are fixed over their full supports, thus summarize the overall contribution of the uncertainty present in the input variables to those of the model outputs.

Among the above three groups of methods, the GSA has received the greatest attention during the past few decades, and a plenty of GSA techniques/indices have been developed for different purposes. The screening methods have been developed for screening the non-influential variables in moderate to high dimensional problems [11, 12]. The variance-based sensitivity indices [13, 14, 15], rooted in the Random sampling-high dimensional model representation (RS-HDMR) [16], aim at measuring the relative importance of the input variables by attributing the model response variance to each input variable and their interactions. Considering the setting of uncertainty reduction, a modified versions of the variance-based sensitivity indices, called W-indices, has also been developed for quantifying the effects of reducing the input uncertainty on that of model output [17]. Given that the variance is not sufficient for characterizing the uncertainty, the moment-independent sensitivity indices have also been devised for investigating the effect of each input variable on the full probability distribution of the model response [18, 19, 20]. The derivative-based sensitivity indices have also been established to realize variable screening with lower computational cost than the variance-based ones [21, 22, 23]. The global reliability sensitivity indices have been developed in the area of structural reliability, based on the variance-based indices, for measuring the contribution of input variables to the failure probability of structures [24, 25, 26]. Despite the extensive GSA indices that have been developed,

the variance-based ones continue to receive the greatest concerns of both researchers and practitioners owing to the elegant mathematical interpretations for both independent and dependent variables, as well as their ability of capturing different types of effects [7, 27, 28]. Developing efficient and robust algorithms for estimating variance-based indices is then one of the most relevant challenges for performing the GSA analysis.

The past few decades have witnessed a rapid development of numerical algorithms for variance-based sensitivity indices, and one can refer to Ref. [29] for a comprehensive review on these related developments. Generally, these methods can be divided into three classes, i.e., Fourier amplitude sensitivity test (FAST), (quasi-) Monte Carlo simulation (MCS), and surrogate models. The FAST method, developed in the area of computational chemistry [30], estimates the partial variance terms involved in the variance-based sensitivity indices based on periodic sampling and Fourier transformation, and it has been widely studied and substantially improved since its development (see e.g., Refs. [31, 32, 33, 34]). The MCS method involves first formulating the partial variance terms with multi-dimensional integrals, and then utilizing MCS, driven by simple random sampling or Latin Hypercube Sampling (LHS) [35] or Sobol's low-discrepancy sequence [36], to estimate these integrals. Following this scheme, a multitude of MCS estimators have been developed (see e.g., Refs. [37, 38, 39, 40]). The surrogate models, such as state dependent regression [41], polynomial chaos expansion [42], support vector regression [43] and Kriging, also called Gaussian Process Regression (GPR) [44, 45, 46, 47], have also been investigated for estimating the sensitivity indices. In terms of reliability of estimation, MCS is the most competitive scheme as confidence intervals can be computed for the sensitivity indices from the MCS estimators, but it also suffers from the large number of required simulator calls, which make it not applicable to computationally expensive simulators.

In recent years, Bayesian numerical analysis [48] with its different variants, such as Bayesian probabilistic optimization [49], Bayesian Probabilistic Integration (BPI) [50, 51], and Bayesian probabilistic Partial Differential Equation (PDE) solution [52], has emerged as a cutting-edge method in scientific computation. The aim of this work is therefore to extend the BPI methods for inferring the variance-based sensitivity indices from data and computer simulators. This topic has also been investigated in Ref.[46] in a full Bayesian scheme and in Ref.[53] with the so-called Bayesian MCS scheme, but in both papers, only the posterior mean and the main effect indices are investigated. In this work, both the posterior means and posterior variances will be first investigated for both the main and total variance-based indices based on BPI, following which, a data-driven BPI approach and an adaptive BPI approach will be developed for efficiently estimating the sensitivity indices. To achieve this goal, two principle lemmas are first developed for realizing the Bayesian inference, and then the posterior means and variances are both analytically derived for the sensitivity indices, where the posterior variances summarize the discretization errors for estimating these sensitivity indices. These analytical results form the basis of the data-driven BPI approach, with which the posterior features of the sensitivity indices can be inferred from any supervised learning data. To further

improve the efficiency of the algorithm for computationally expensive simulators, an adaptive experiment design strategy is ultimately introduced. The effectiveness of the proposed methods are demonstrated by numerical examples, and their applicability to real-world engineering problems as well as their engineering significance are illustrated by three engineering benchmarks with FE simulators.

The rest of this paper is organized as follows. Section 2 briefly reviews the variance-based sensitivity indices and the BPI approach, followed by the core developments in section 3, which includes the Bayesian inference of the sensitivity indices and the data-driven BPI. The adaptive BPI approach is then developed in section 4, followed by the numerical and engineering test examples in section 5. Section 6 closes the paper with conclusions.

2. Brief Review of Related Topics

Before the introduction of the main developments, it is helpful to briefly review two important topics to be studied/utilized in this article, i.e., the variance-based sensitivity indices and the BPI. The expectation and variance operators utilized in this paper are declared in Table 1 for avoiding confusion.

Table 1: Explanations of operators utilized in this paper.

Operators	Explanations
$\mathbb{E}_{\mathcal{D}}\left[\cdot\right]$, $\mathbb{V}_{\mathcal{D}}\left[\cdot\right]$, $\mathrm{cov}_{\mathcal{D}}\left[\cdot,\cdot\right]$	Posterior expectation, variance and covariance operators with respect to the GPR model trained based on the data set $\mathcal{D}$.
$\mathbb{E}_{\boldsymbol{I}}\left[\cdot\right]$, $\mathbb{V}_{\boldsymbol{I}}\left[\cdot\right]$	Expectation and variance operators with respect to any subset $\boldsymbol{x}_{\boldsymbol{I}}$ of $\boldsymbol{x}$, where $\boldsymbol{I} \subseteq \{1, 2, \cdots, n\}$.
$\mathbb{E}_{-\boldsymbol{I}}\left[\cdot\right]$, $\mathbb{V}_{-\boldsymbol{I}}\left[\cdot\right]$	Expectation and variance operators with respect to the complementary set $\boldsymbol{x}_{-\boldsymbol{I}} = \boldsymbol{x} \setminus \boldsymbol{x}_{\boldsymbol{I}}$, where $(\cdot) \setminus (\cdot)$ indicates set subtraction.
$\mathbb{E}'_{\boldsymbol{I}}\left[\cdot\right]$, $\mathbb{V}'_{\boldsymbol{I}}\left[\cdot\right]$	Expectation and variance operators with respect to any subset $\boldsymbol{x}'_{\boldsymbol{I}}$ of $\boldsymbol{x}'$, where $\boldsymbol{x}'$ is an element-by-element independent replicate of $\boldsymbol{x}$.
$\mathbb{E}_{\boldsymbol{I}_1} \mathbb{E}'_{\boldsymbol{I}_2}\left[\cdot\right]$, $\mathbb{V}_{\boldsymbol{I}_1} \mathbb{V}'_{\boldsymbol{I}_2}\left[\cdot\right]$	Expectation and variance operators with respect to two random vectors $\boldsymbol{x}_{\boldsymbol{I}_1}$ and $\boldsymbol{x}'_{\boldsymbol{I}_2}$, where $\boldsymbol{I}_1, \boldsymbol{I}_2 \subseteq \{1, 2, \cdots, n\}$.

2.1. Variance-based sensitivity indices

In this paper, only the sensitivity indices for independent input variables are investigated. Let $y = \mathcal{M}\left(\boldsymbol{x}\right)$ denote the deterministic computer simulator, where y is the one-dimensional model output of interest, and $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ denotes the n-dimensional vector of random input variables assumed to follow independent standard normal distribution for ease of description. For a non-Gaussian random variable, the iso-probabilistic transformation can be utilized to transform it into standard normal [54]. For example, suppose the cumulative distribution function (CDF) of x is $F\left(x\right)$, then the mapping $u = \Phi^{-1}\left(F\left(x\right)\right)$ transforms

it into a standard normal variable u, where $\Phi^{-1}\left(\cdot\right)$ denotes the inverse CDF of standard normal distribution. With the above setting, the HDMR decomposition (also known as functional ANOVA decomposition) of the model function is formulated as

$$y = \mathcal{M}\left(\boldsymbol{x}\right) = \mathcal{M}_0 + \sum_{i=1}^{n} \mathcal{M}_i\left(x_i\right) + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathcal{M}_{ij}\left(\boldsymbol{x}_{ij}\right) + \cdots + \mathcal{M}_{12\cdots n}\left(\boldsymbol{x}\right) \tag{1}$$

where $\boldsymbol{x}_{ij} = \left(x_i, x_j\right)$, and

$$\begin{aligned}
\mathcal{M}_0 &= \mathbb{E}_{1:n}\left[\mathcal{M}\left(\boldsymbol{x}\right)\right] \\
\mathcal{M}_i\left(x_i\right) &= \mathbb{E}_{-i}\left[\mathcal{M}\left(\boldsymbol{x}\right)|x_i\right] - \mathcal{M}_0 \\
\mathcal{M}_{ij}\left(\boldsymbol{x}_{ij}\right) &= \mathbb{E}_{-ij}\left[\mathcal{M}\left(\boldsymbol{x}\right)|\boldsymbol{x}_{ij}\right] - \mathcal{M}_i\left(x_i\right) - \mathcal{M}_j\left(x_j\right) - \mathcal{M}_0
\end{aligned} \tag{2}$$

with $\mathbb{E}_{1:n}\left[\cdot\right]$ indicating the expectation operator taken with respect to all the n input variables, $\mathbb{E}_{-i}\left[\cdot|x_i\right]$ referring to the conditional expectation taken with respect to $\boldsymbol{x}_{-i} = \boldsymbol{x} \setminus x_i$, and $\mathbb{E}_{-ij}\left[\cdot|\boldsymbol{x}_{ij}\right]$ denoting the conditional expectation taken with respect to $\boldsymbol{x}_{-ij} = \boldsymbol{x} \setminus \boldsymbol{x}_{ij}$.

On the premise that $\mathcal{M}\left(\boldsymbol{x}\right)$ is square-integrable and the input variables are independent with each other, all the HDMR components are orthogonal, and thus have zero covariance. Taking variance to both sides of Eq. (1) yields [37]:

$$V_y = \mathbb{V}_{1:n}\left[\mathcal{M}\left(\boldsymbol{x}\right)\right] = \sum_{i=1}^{n} V_i + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} V_{ij} + \cdots + V_{12\cdots n} \tag{3}$$

where $\mathbb{V}_{1:n}\left[\cdot\right]$ indicates the variance operator taken with respect to all the elements of $\boldsymbol{x}$, the first- and second- order partial variances are defined as:

$$\begin{aligned}
V_i &= \mathbb{V}_i\left[\mathcal{M}_i\left(x_i\right)\right] = \mathbb{V}_i\left[\mathbb{E}_{-i}\left[\mathcal{M}\left(\boldsymbol{x}\right)|x_i\right]\right] \\
V_{ij} &= \mathbb{V}_{ij}\left[\mathcal{M}_{ij}\left(\boldsymbol{x}_{ij}\right)\right] = \mathbb{V}_{ij}\left[\mathbb{E}_{-ij}\left[\mathcal{M}\left(\boldsymbol{x}\right)|\boldsymbol{x}_{ij}\right]\right] - V_i - V_j
\end{aligned} \tag{4}$$

and the higher-order partial variances are similarly defined.

Based on the variance decomposition in Eq. (3), the normalized main effect index S_i and total effect index S_{Ti} for x_i are defined by:

$$S_i = \frac{V_i}{V_y}, \quad \text{and} \quad S_{Ti} = \frac{V_{Ti}}{V_y} \tag{5}$$

respectively, where V_{Ti} equals to the summation of all the partial variance terms in Eq. (3) with subscript containing i, and it can be further derived as $V_{Ti} = \mathbb{E}_{-i}\left[\mathbb{V}_i\left[\mathcal{M}\left(\boldsymbol{x}\right)|\boldsymbol{x}_{-i}\right]\right]$ [7, 14]. Higher-order normalized sensitivity indices can be similarly defined, however, these two are usually of most concern to analysts. The main effect index S_i, also called the first-order effect index, measures the percentage of the model response variance contributed by x_i individually; the total effect index summarizes the overall contribution of x_i,

which includes the individual contribution of x_i and all of its interaction contributions with the other input variables. By subtracting the main effect index S_i from the total effect index S_{Ti}, we can generate the interaction sensitivity index $S_{Ii} = S_{Ti} - S_i$ which measures the total interaction contribution of x_i.

The aim this paper is to develop Bayesian inference methods for inferring the values of the sensitivity indices based on a set of arbitrary supervised training data, and specifically, the main and total effect indices are examplified. To achieve this target, we need to develop Bayesian inference formulas for the first-order (or main) partial variance V_i, the total partial variance V_{Ti}, and the total variance V_y.

2.2. Bayesian probabilistic integration

We review the BPI method by taking the inference of the constant HDMR component $\mathcal{M}_0$ in Eq. (1) as an example. Let $\mathcal{D} = \{\mathcal{X}, \mathcal{Y}\}$ denote a set of supervised training data, where $\mathcal{X}$ is a $(N \times n)$-dimensional sample matrix with each row being a sample of $\boldsymbol{x}$, and $\mathcal{Y}$ is the corresponding N-dimensional column vector of model response values. Then the inference task can be described as follows: given the training data $\mathcal{D}$, what can we infer about the value of $\mathcal{M}_0$?

In most BPI practices, the model response function $\mathcal{M}(\boldsymbol{x})$ is approximated by a GPR model $\hat{\mathcal{M}}(\boldsymbol{x}) \sim \mathcal{GP}(\mu_{\mathcal{M}}(\boldsymbol{x}), cov_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}'))$ trained based on the data set $\mathcal{D}$, where $\mu_{\mathcal{M}}(\boldsymbol{x})$ is the posterior mean prediction at point $\boldsymbol{x}$, and $cov_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}')$ is the posterior covariance at the two points $\boldsymbol{x}$ and $\boldsymbol{x}'$, which represents the spatial correlation of the GPR model. Making $\boldsymbol{x} = \boldsymbol{x}'$ yields the posterior variance $\sigma_{\mathcal{M}}^2(\boldsymbol{x}) = cov_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x})$, which measures the discretization error of the GPR prediction at the point $\boldsymbol{x}$. The training of the GPR model is also a Bayesian inference process. Let $b(\boldsymbol{x})$ denote the prior assumption of the mean of the GPR model, which can be assumed to be zero, constant, or polynomials. The prior information on the covariance of the GPR model is characterized by a kernel function $\kappa(\boldsymbol{x}, \boldsymbol{x}')$, which may be of various types [55, 56]. In this paper, we utilize the squared exponential kernel with distinct length scale parameter for each dimension, which is formulated as:

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \sigma_0^2 \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{x}')\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{x}')^\top}{2}\right) \tag{6}$$

where σ_0^2 is the variance parameter describing the variation of the GPR model, $\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2)$ is a diagonal matrix, and σ_i indicates the length scale of the i-th dimension, which describes the strength of correlation of the GPR model along this dimension. The hyper-parameters involved in the prior mean and prior covariance can be calculated by maximizing the logarithm of the likelihood function formulated based on $\mathcal{D}$. One can refer to Ref. [55] for more details. With these hyper-parameters having been computed, the posterior mean $\mu_{\mathcal{M}}(\boldsymbol{x})$ and posterior covariance $cov_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}')$ of the GPR model are inferred as:

$$\mu_{\mathcal{M}}(\boldsymbol{x}) = \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{M}}(\boldsymbol{x})\right] = b(\boldsymbol{x}) + \kappa(\boldsymbol{x}, \mathcal{X})^\top \mathcal{K}^{-1}(\mathcal{Y} - \boldsymbol{b}(\mathcal{X})) \tag{7}$$

and

$$cov_{\mathcal{M}}\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}\left(\boldsymbol{x}'\right)\right] = \kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right) - \kappa\left(\boldsymbol{x}, \mathcal{X}\right)^{\top}\mathcal{K}^{-1}\kappa\left(\boldsymbol{x}', \mathcal{X}\right) \tag{8}$$

respectively, where $\kappa\left(\boldsymbol{x}, \mathcal{X}\right)$ indicates the prior covariance between $\boldsymbol{x}$ and $\mathcal{X}$ (for a fixed $\boldsymbol{x}$, it is a column vector), $\mathcal{K}$ is a $(N \times N)$-dimensional positive definite matrix with its (i, j)-th element being the prior covariance of the i-th row $\mathcal{X}_i$ and j-th row $\mathcal{X}_j$ of $\mathcal{X}$. Given the GPR model $\hat{\mathcal{M}}\left(\boldsymbol{x}\right)$ for approximating $\mathcal{M}\left(\boldsymbol{x}\right)$, the induced random variable $\hat{\mathcal{M}}_0 = \mathbb{E}_{1:n}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right]$ for approximating $\mathcal{M}_0$ is a Gaussian random variable, and its posterior mean $\mu_{\mathcal{M}0}$ and posterior variance $\sigma_{\mathcal{M}0}^2$ can be explicitly inferred as [51, 50, 57]:

$$\mu_{\mathcal{M}0} = \mathbb{E}_{1:n}\left[\mu_{\mathcal{M}}\left(\boldsymbol{x}\right)\right] = \mathbb{E}_{1:n}\left[b\left(\boldsymbol{x}\right)\right] + \mathbb{E}_{1:n}\left[\kappa\left(\boldsymbol{x}, \mathcal{X}\right)\right]^{\top}\mathcal{K}^{-1}\left(\mathcal{Y} - \boldsymbol{b}\left(\mathcal{X}\right)\right) \tag{9}$$

and

$$\begin{aligned} \sigma_{\mathcal{M}0}^2 &= \mathbb{E}_{1:n}\mathbb{E}_{1:n}'\left[cov_{\mathcal{M}}\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] \\ &= \mathbb{E}_{1:n}\mathbb{E}_{1:n}'\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] + \mathbb{E}_{1:n}\left[\kappa\left(\boldsymbol{x}, \mathcal{X}\right)\right]^{\top}\mathcal{K}^{-1}\mathbb{E}_{1:n}'\left[\kappa\left(\boldsymbol{x}', \mathcal{X}\right)\right]. \end{aligned} \tag{10}$$

The closed-form expression for $\mathbb{E}_{1:n}\left[b\left(\boldsymbol{x}\right)\right]$ is trivial, and those for the kernel means $\mathbb{E}_{1:n}\left[\kappa\left(\boldsymbol{x}, \mathcal{X}\right)\right]$ and $\mathbb{E}_{1:n}\mathbb{E}_{1:n}'\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right]$ are formulated as [51, 57]:

$$\begin{aligned} \mathbb{E}_{1:n}\left[\kappa\left(\boldsymbol{x}, \mathcal{X}\right)\right] &= \sigma_0^2\left|\Sigma^{-1} + I_n\right|^{-1/2}\exp\left(-\frac{1}{2}\mathrm{vec}\left[\mathrm{diag}\left(\mathcal{X}\left(\Sigma + I_n\right)^{-1}\mathcal{X}^{\top}\right)\right]\right) \\ \mathbb{E}_{1:n}\mathbb{E}_{1:n}'\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] &= \sigma_0^2\left|2\Sigma^{-1} + I_n\right|^{-1/2} \end{aligned} \tag{11}$$

where I_n indicates the identity matrix of $n \times n$ dimensions, $\mathrm{vec}\left[\mathrm{diag}\left(\cdot\right)\right]$ indicates a column vector created with the diagonal elements of the argument. One notes that the formulations in Eqs. (9) and (10) hold for any kinds of kernels, whereas, the closed-form expressions in Eq. (11) are derived from the squared exponential kernel. For other kinds of kernel such as Matérn kernel 3/2 and 5/2, the closed-form expressions for $\mathbb{E}_{1:n}\left[\kappa\left(\boldsymbol{x}, \mathcal{X}\right)\right]$ and $\mathbb{E}_{1:n}\mathbb{E}_{1:n}'\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right]$ are also available in Ref. [50] for a summary.

Given the above results, we can now answer the question raised in the first paragraph of this subsection, i.e., the exact value of $\mathcal{M}_0$ may not be learned from the limited volume of data $\mathcal{D}$, but we can infer a Gaussian probability distribution $\mathcal{N}\left(u_{\mathcal{M}0}, \sigma_{\mathcal{M}0}^2\right)$ for describing its deterministic value, where the variation of this distribution summarizes the epistemic (or discretization) uncertainty on this deterministic value. Based on this, a $100\left(\Phi\left(\alpha\right) - \Phi\left(-\alpha\right)\right)\%$ posterior credibility interval can be generated as $\left[\mu_{\mathcal{M}0} - \alpha\sigma_{\mathcal{M}0}, \mu_{\mathcal{M}0} + \alpha\sigma_{\mathcal{M}0}\right]$, where $\alpha \in [0, 1]$. With the increase of the data volume, it is expected that this credibility interval shrinks to the true value of $\mathcal{M}_0$. In the next section, we will show how to infer the posterior features for the sensitivity indices based on the GPR model $\hat{\mathcal{M}}\left(\boldsymbol{x}\right)$.

3. Bayesian inference of sensitivity indices

In the previous section, the details of the BPI approach for estimating $\mathcal{M}_0$ have been reviewed, and it is concluded that, given the GPR representation of the model function $\mathcal{M}(\boldsymbol{x})$, the posterior distribution of $\mathcal{M}_0$ is also Gaussian. Indeed, the induced probabilistic models for any orders of HDMR components (e.g., $\mathcal{M}_i(x_i)$ and $\mathcal{M}_{ij}(\boldsymbol{x}_{ij})$) are Gaussian as well [46, 58]. This provides a basis to infer the posterior features of the first-order partial variances V_i, the total partial variance V_{Ti}, and the total V_y in this section. Before this, two lemmas are proposed as the cornerstone of all the subsequent inferences.

3.1. First principle for inference

Lemma 1. *Let $\hat{\mathcal{G}}(\boldsymbol{u}) \sim \mathcal{GP}\left(\mu_{\mathcal{G}}(\boldsymbol{u}), \sigma_{\mathcal{G}}^2(\boldsymbol{u})\right)$ indicate a GPR model for approximating the deterministic function $\mathcal{G}(\boldsymbol{u})$ with random arguments $\boldsymbol{u}$ of arbitrary dimension, and assume that the induced Gaussian variable $\mathbb{E}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right] \equiv 0$. The induced variance $\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]$ for approximating the variance $\mathbb{V}_{\boldsymbol{u}}[\mathcal{G}(\boldsymbol{u})]$ is a (non-Gaussian) random variable with the posterior mean $\mathbb{E}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]\right]$ and posterior variance $\mathbb{V}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]\right]$ being formulated as:*

$$\mathbb{E}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]\right] = \underbrace{\mathbb{V}_{\boldsymbol{u}}\left[\mu_{\mathcal{G}}(\boldsymbol{u})\right]}_{\vartheta_1} + \underbrace{\mathbb{E}_{\boldsymbol{u}}\left[\sigma_{\mathcal{G}}^2(\boldsymbol{u})\right]}_{\vartheta_2} \tag{12}$$

and

$$\mathbb{V}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]\right] = \mathbb{E}_{\boldsymbol{u}}\mathbb{E}_{\boldsymbol{u}}^{'}\left[\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2(\boldsymbol{u}), \hat{\mathcal{G}}^2(\boldsymbol{u}')\right]\right] \tag{13}$$

respectively, where $\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2(\boldsymbol{u}), \hat{\mathcal{G}}^2(\boldsymbol{u}')\right]$ is the covariance of the squared GPR model $\hat{\mathcal{G}}^2(\boldsymbol{u})$. The posterior variance in Eq. (13) is further derived as:

$$\mathbb{V}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]\right] = \overbrace{2\mathbb{E}_{\boldsymbol{u}}\mathbb{E}_{\boldsymbol{u}}^{'}\left[\mathrm{cov}_{\mathcal{D}}^2\left[\hat{\mathcal{G}}(\boldsymbol{u}), \hat{\mathcal{G}}(\boldsymbol{u}')\right]\right]}^{\gamma_1}$$
$$+ \underbrace{4\mathbb{E}_{\boldsymbol{u}}\mathbb{E}_{\boldsymbol{u}}^{'}\left[\mu_{\mathcal{G}}(\boldsymbol{u})\mu_{\mathcal{G}}(\boldsymbol{u}')\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}(\boldsymbol{u}), \hat{\mathcal{G}}(\boldsymbol{u}')\right]\right]}_{\gamma_2}. \tag{14}$$

The proof of Lemma 1 is presented in Appendix A. As can be seen from Eq. (12), the posterior mean of $\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]$ consists of two parts, i.e., ϑ_1 and ϑ_2, where ϑ_1 is the variance of the posterior mean $\mu_{\mathcal{G}}(\boldsymbol{u})$, and ϑ_2 refers to the expectation of the posterior variance $\sigma_{\mathcal{G}}^2(\boldsymbol{u})$, indicating that, as long as the closed-form expressions of the posterior mean and variance of the GPR model $\hat{\mathcal{G}}(\boldsymbol{u})$ are available, the analytical expression (not in closed form) of the posterior mean of $\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]$ can be generated. Eq. (13) indicates that the posterior variance of $\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]$ equals to the expectation of the posterior covariance of the induced stochastic process model $\hat{\mathcal{G}}^2(\boldsymbol{u})$ (not Gaussian). The posterior covariance of $\hat{\mathcal{G}}^2(\boldsymbol{u})$ is uniquely determined by the posterior mean and covariance of the GPR model $\hat{\mathcal{G}}(\boldsymbol{u})$, whose closed-form expression can be analogously formulated with Eq. (7) and (8). Thereof, once the closed-form expressions of the posterior

mean and variance of $\hat{\mathcal{G}}(\boldsymbol{u})$ is known, the analytical expression (not in closed form) of the posterior variance of $\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]$ can be obtained based on Eq. (14). Thus, although $\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]$ is no longer Gaussian, we can still compute its posterior mean and variance based on Lemma 1, where the posterior variance summarizes the discretization error for Bayesian inference of the deterministic variance $\mathbb{V}_{\boldsymbol{u}}[\mathcal{G}(\boldsymbol{u})]$. One notes that a similar result with Eq. (12) for deriving the closed-form expression of the posterior mean of V_i in a full Bayesian setting has been reported in Ref. [46].

Lemma 2. *For the posterior variance* $\mathbb{V}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]\right]$ *in Eq.* (14), *an upper bound can be further derived as:*

$$\mathbb{V}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}(\boldsymbol{u})\right]\right] \leqslant 2\mathbb{E}_{\boldsymbol{u}}^2\left[\sigma_{\mathcal{G}}^2(\boldsymbol{u})\right] + 4\mathbb{E}_{\boldsymbol{u}}^2\left[|\mu_{\mathcal{G}}(\boldsymbol{u})|\,\sigma_{\mathcal{G}}(\boldsymbol{u})\right]. \tag{15}$$

The proof of Lemma 2 is also presented in Appendix A. This lemma provides a simpler way to measure the discretization error for inferring the value of $\mathbb{V}_{\boldsymbol{u}}[\mathcal{G}(\boldsymbol{u})]$ as it only requires the expressions of the posterior mean and posterior variance of $\hat{\mathcal{G}}(\boldsymbol{u})$ without knowing that of the posterior covariance, and the dimension of the integral involved in Eq. (15) is half of that involved in Eq. (14).

It should be noted that Lemma 1 and Lemma 2 hold for any types of kernels. Inspired by these two lemmas, the remaining task is to formulate each of the three (partial) variance terms as the variance of a function (with zero expectation) approximated by a GPR model that can be derived from the properly trained one $\hat{\mathcal{M}}(\boldsymbol{x})$, and then derive the posterior mean, posterior variance and posterior covariance of this GPR model. The dimension of arguments of this function can be arbitrary. To do this, two basic rules needs to be clarified before inferring the posterior features of the (partial) variance terms, which are, i) any linear combination of a set of GPR models is still a Gaussian Process, and ii) the linear projection of a GPR model on any direction of the subset of its arguments (defined by the inner product of this GPR model with the density of these subset of arguments) is also a Gaussian process or a Gaussian random variable [50]. We try to avoid arcane mathematical concepts in this article, but these two rules are critical for understanding some of the results below. From the next subsection on, we start to derive the analytical expressions for all the three (partial) variance terms based on the two lemmas and the two rules given in this paragraph.

3.2. Inference of first-order partial variance

From the definition of the first-order HDMR components and the first-order partial variance, it is known that $\mathbb{V}_i = \mathbb{V}_i\left[\mathcal{M}_i(x_i)\right]$, where $\mathbb{E}_i\left[\mathcal{M}_i(x_i)\right] = 0$, indicating that the induced GPR derived from $\hat{\mathcal{M}}(\boldsymbol{x})$ can be formulated as $\hat{\mathcal{M}}_i(x_i) = \mathbb{E}_{-i}\left[\hat{\mathcal{M}}(\boldsymbol{x})\right] - \hat{\mathcal{M}}_0$, as $\mathbb{E}_{-i}\left[\hat{\mathcal{M}}(\boldsymbol{x})\right]$ is a linear projection of $\hat{\mathcal{M}}(\boldsymbol{x})$ along the direction of x_i, and $\hat{\mathcal{M}}_0 = \mathbb{E}_{1:n}\left[\hat{\mathcal{M}}(\boldsymbol{x})\right]$ is a linear projection of $\hat{\mathcal{M}}(\boldsymbol{x})$ along the direction of $\boldsymbol{x}$. Then based on Lemma 1 and Lemma 2, to make a Bayesian inference for $\mathbb{V}_i$, we need only to derive the posterior mean and posterior covariance of the one-dimensional GPR model $\hat{\mathcal{M}}_i(x_i)$, whose results are given by the following corollary.

9

Corollary 1. *Both the posterior mean and posterior covariance of $\hat{\mathcal{M}}_i(x_i)$ have closed-form expressions, which are formulated as:*

$$\mu_{\mathcal{M}i}(x_i) = \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{M}}_i(x_i)\right] = b_i(x_i) + \boldsymbol{\zeta}_i(x_i)^{\top}\mathcal{K}^{-1}\left(\mathcal{Y} - b(\mathcal{X})\right) - \mu_{\mathcal{M}0} \tag{16}$$

and

$$cov_{\mathcal{M}i}(x_i, x_i') = \text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}_i(x_i), \hat{\mathcal{M}}_i(x_i')\right] = c_{i,i'}(x_i, x_i') - c_i(x_i) - c_i(x_i') + c_0 \tag{17}$$

where

$$
\begin{aligned}
b_i(x_i) &= \mathbb{E}_{-i}[b(\boldsymbol{x})] \\
\boldsymbol{\zeta}_i(x_i) &= \mathbb{E}_{-i}[\kappa(\boldsymbol{x}, \mathcal{X})] \\
c_{i,i'}(x_i, x_i') &= \mathbb{E}_{-i}\mathbb{E}'_{-i}[cov_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}')] \\
c_i(x_i) &= \mathbb{E}_{-i}\mathbb{E}'_{1:n}[cov_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}')].
\end{aligned}
\tag{18}
$$

The proof of Corollary 1 as well as the closed-form expressions for the four terms in Eq. (18) are presented in Appendix B. One may note that the results given in Eqs. (16)-(17) hold for any form of kernels, but the closed-form expressions for the four terms in Eq. (18) differ from kernel to kernel. This is also true for the following two corollaries. The posterior variance of $\hat{\mathcal{M}}_i(x_i)$ can be simply generated by Eq. (17) as $\sigma^2_{\mathcal{M}i}(x_i) = cov_{\mathcal{M}i}(x_i, x_i)$. With the closed-form expressions of the posterior mean $\mu_{\mathcal{M}i}(x_i)$, the posterior variance $\sigma^2_{\mathcal{M}i}(x_i)$ and the posterior covariance $cov_{\mathcal{M}i}(x_i, x_i')$, it is then trivial to generate the analytical expressions for the posterior mean and posterior variance (also the upper bound) of the first-order partial variance V_i. For simplicity, we do not repeat it. One notes that the estimation of the posterior mean of V_i following the Bayesian scheme has also been investigated by Ref. [53], however, what exactly was estimated in that paper is actually the first term ϑ_1 of Eq. (12), thus is a biased estimate of the posterior mean of V_i. This bias unquestionably decreases with the increment of training data, but it hardly vanishes. The posterior variance of V_i as well as total effect indices were not investigated by Ref. [53], and are both treated in this work.

3.3. Inference of total partial variance

With simple mathematical derivation, the total partial variance can be formulated as:

$$V_{Ti} = \mathbb{V}_{1:n}\left[\mathcal{M}(\boldsymbol{x}) - \mathbb{E}_i[\mathcal{M}(\boldsymbol{x})]\right]. \tag{19}$$

One can refer to Appendix C for proof of Eq. (19). Let $\mathcal{M}_{Ti}(\boldsymbol{x}) = \mathcal{M}(\boldsymbol{x}) - \mathbb{E}_i[\mathcal{M}(\boldsymbol{x})]$, then it is apparently that $\mathbb{E}_{1:n}[\mathcal{M}_{Ti}(\boldsymbol{x})] = 0$. Thus the underlying GPR model for inferring the total partial variance can be

formulated as:

$$\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right) = \hat{\mathcal{M}}\left(\boldsymbol{x}\right) - \mathbb{E}_i\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right]. \tag{20}$$

Another feasible choice of the underlying GPR model for inferring the total partial variance is $\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}, x_i'\right) = \hat{\mathcal{M}}\left(\boldsymbol{x}\right) - \hat{\mathcal{M}}\left(\boldsymbol{x}_{-i}, x_i'\right)$ since $2V_{Ti} = \mathbb{V}_{1:n}\mathbb{V}_i'\left[\mathcal{M}\left(\boldsymbol{x}\right) - \mathcal{M}\left(\boldsymbol{x}_{-i}, x_i'\right)\right]$ [39]. In this paper, the formula given by Eq. (20) is utilized as the dimension of its arguments is lower. Based on Lemma 1 and Lemma 2, the following corollary for deriving the posterior mean and variance of the total partial variance can be generated.

Corollary 2. *The posterior mean and covariance of the underlying GPR model $\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right)$ are formulated as:*

$$\mu_{\mathcal{M}Ti}\left(\boldsymbol{x}\right) = \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right)\right] = \mu_{\mathcal{M}}\left(\boldsymbol{x}\right) - b_{-i}\left(\boldsymbol{x}_{-i}\right) - \zeta_{-i}\left(\boldsymbol{x}_{-i}\right)^{\top}\mathcal{K}^{-1}\left(\mathcal{Y} - b\left(\mathcal{X}\right)\right) \tag{21}$$

and

$$
\begin{aligned}
cov_{\mathcal{M}Ti}\left(\boldsymbol{x}, \boldsymbol{x}'\right) &= \mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}'\right)\right] \\
&= cov_{\mathcal{M}}\left(\boldsymbol{x}, \boldsymbol{x}'\right) - c_{-i}\left(\boldsymbol{x}, \boldsymbol{x}'_{-i}\right) - c_{-i}\left(\boldsymbol{x}', \boldsymbol{x}_{-i}\right) + c_{-ii}\left(\boldsymbol{x}_{-i}, \boldsymbol{x}'_{-i}\right)
\end{aligned}
\tag{22}
$$

respectively, where

$$
\begin{aligned}
b_{-i}\left(\boldsymbol{x}_{-i}\right) &= \mathbb{E}_i\left[b\left(\boldsymbol{x}\right)\right] \\
\zeta_{-i}\left(\boldsymbol{x}_{-i}\right) &= \mathbb{E}_i\left[\kappa\left(\boldsymbol{x}, \mathcal{X}\right)\right] \\
c_{-i}\left(\boldsymbol{x}, \boldsymbol{x}'_{-i}\right) &= \mathbb{E}_i'\left[cov_{\mathcal{M}}\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] \\
c_{-ii}\left(\boldsymbol{x}_{-i}, \boldsymbol{x}'_{-i}\right) &= \mathbb{E}_i\mathbb{E}_i'\left[cov_{\mathcal{M}}\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right]
\end{aligned}
\tag{23}
$$

The proof of Corollary 2 as well as the closed-form expressions for the four terms in Eq. (23) can be found in Appendix C. By substituting the expression of posterior mean in Eq. (21) and that of the posterior covariance in Eq. (22) into Lemma 1, one can generate the analytical formulations of the posterior mean and posterior variance of the total partial variance $\hat{V}_{Ti}$, and similarly by substituting the posterior mean and variance of $\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right)$ (see Eq. (C.10)) in Appendix C) into Lemma 2, a conservative estimator can be generated as an upper bound of the posterior variance of $\hat{V}_{Ti}$.

3.4. Inference of total variance

Based on the definition of the total variance, it is formulated as:

$$V_y = \mathbb{V}_{1:n}\left[\mathcal{M}\left(\boldsymbol{x}\right) - \mathcal{M}_0\right] \tag{24}$$

, indicating that the underlying GPR model is $\hat{\mathcal{M}}_y\left(\boldsymbol{x}\right) = \hat{\mathcal{M}}\left(\boldsymbol{x}\right) - \hat{\mathcal{M}}_0$, which obviously satisfies $\mathbb{E}_{1:n}\left[\hat{\mathcal{M}}_y\left(\boldsymbol{x}\right)\right] = 0$. Thus, the following corollary is presented for the total variance term V_y.

Corollary 3. *The posterior mean $\mu_{\mathcal{M}y}(\boldsymbol{x})$ and the posterior covariance $cov_y(\boldsymbol{x}, \boldsymbol{x}')$ of the induced GPR model $\hat{\mathcal{M}}_y(\boldsymbol{x})$ is formulated in closed form as:*

$$\mu_{\mathcal{M}y}(\boldsymbol{x}) = \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{M}}_y(\boldsymbol{x})\right] = \mu_{\mathcal{M}}(\boldsymbol{x}) - \mu_{\mathcal{M}0} \tag{25}$$

and

$$
\begin{aligned}
cov_{\mathcal{M}y}(\boldsymbol{x}, \boldsymbol{x}') &= \mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}_y(\boldsymbol{x}), \hat{\mathcal{M}}_y(\boldsymbol{x}')\right] \\
&= cov_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}') - \mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}(\boldsymbol{x}), \hat{\mathcal{M}}_0\right] - \mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}(\boldsymbol{x}'), \hat{\mathcal{M}}_0\right] + \sigma_{\mathcal{M}0}^2.
\end{aligned}
\tag{26}
$$

The mathematical derivations of Corollary 3 as well as the closed-form expression for the posterior covariance terms in Eq. (26) are presented in Appendix D. Similarly, by substituting the results in Corollary 3 into Lemma 1 and Lemma 2, we can generate the analytical expressions of the posterior mean $\mathbb{E}_{\mathcal{D}}\left[\hat{V}_y\right]$ and posterior variance $\mathbb{V}_{\mathcal{D}}\left[\hat{V}_y\right]$ for the total variance term V_y as well as an upper bound serves as a conservative estimate of the posterior variance $\mathbb{V}_{\mathcal{D}}\left[\hat{V}_y\right]$. In the next subsection, we summarize the above Bayesian inference and show the details for numerical implementation.

3.5. Summary of Inference and Numerical Implementation

Until now, we have generated the analytical formulations of the posterior mean and the posterior variance for the first-order partial variance $\hat{V}_i$, the total partial variance $\hat{V}_{Ti}$ and the total variance $\hat{V}_y$ via Subsections 3.1~3.4, which are sufficiently useful for estimating the main and total effect indices. However, if required, we can also infer the posterior features for the interaction partial variance terms as long as they can be formulated as the variance of an underlying GPR model induced from the properly trained one $\hat{\mathcal{M}}(\boldsymbol{x})$. For example, for the second-order partial variance V_{ij}, since $V_{ij} = \mathbb{V}_{ij}\left[\mathbb{E}_{-ij}\left[\mathcal{M}(\boldsymbol{x})\right] - \mathcal{M}_i(x_i) - \mathcal{M}_j(x_j) - \mathcal{M}_0\right]$, the underlying GPR model is given as:

$$\hat{\mathcal{M}}_{ij}(\boldsymbol{x}_{ij}) = \mathbb{E}_{-ij}\left[\hat{\mathcal{M}}(\boldsymbol{x})\right] - \hat{\mathcal{M}}_i(x_i) - \hat{\mathcal{M}}_j(x_j) - \hat{\mathcal{M}}_0. \tag{27}$$

Then, one needs only to derive the closed-form expressions for the posterior mean and covariance of $\hat{\mathcal{M}}_{ij}(\boldsymbol{x}_{ij})$ from the GPR model $\hat{\mathcal{M}}(\boldsymbol{x})$. For simplicity, we do not repeat them. The posterior mean provides a mean estimate of the corresponding (partial) variance terms, while the posterior variance, as a measure of epistemic uncertainty, summarizes the corresponding discretization error for this estimate resulting from the limited volume of training data. Thus, for all the (partial) variance terms, these two posterior characters are of special concern. Next, we show how to compute these two characters numerically for all the (partial) variance terms based on the theoretical results in Subsections 3.1~3.4.

From Corollary 1, it is known that the posterior mean $\mu_{\mathcal{M}i}(x_i)$ and the posterior variance $\sigma_{\mathcal{M}i}^2(x_i)$

are both univariate functions of x_i in closed form, thus by substituting them to Eq. (12) in Lemma 1, the posterior mean $\mathbb{E}_{\mathcal{D}}\left[\hat{V}_i\right]$ for the first-order partial variance V_i can be computed numerically by using any univariate numerical integration rule such as Gaussian-Hermite integration. Similarly, as the posterior covariance $cov_{\mathcal{M}i}\left(x_i, x_i'\right)$ is a bivariate function in closed form, thus by substituting it together with the posterior mean $\mu_{\mathcal{M}i}\left(x_i\right)$ to Eq. (14), the posterior variance $\mathbb{V}_{\mathcal{D}}\left[\hat{V}_i\right]$ can be computed by any two-dimensional numerical integration algorithm. Besides, by substituting the posterior variance $\sigma_{\mathcal{M}i}^2\left(x_i\right)$, which is a univariate function in close form, together with that of $\mu_{\mathcal{M}i}\left(x_i\right)$ into Eq. (15) in Lemma 2, an upper bound for the posterior variance $\mathbb{V}_{\mathcal{D}}\left[\hat{V}_i\right]$ can be computed by any univariate numerical integration algorithm.

For total partial variance V_{Ti}, univariate/bivariate numerical integration is not sufficient for estimating the posterior mean and variance. As shown by Eq. (21), the posterior mean $\mu_{\mathcal{M}Ti}\left(\boldsymbol{x}\right)$ of the underlying GPR model owns n arguments, thus the dimension of the involved integral for estimating the posterior mean $\mathbb{E}_{\mathcal{D}}\left[\hat{V}_{Ti}\right]$ is n. Further, from Eq. (22), the dimension of the integral for the posterior variance $\mathbb{V}_{\mathcal{D}}\left[\hat{V}_{Ti}\right]$ is $2n$. We suggest to use MCS and/or Sparse Grid Integration (SGI) [59] for estimating these integrals. To implement MCS, two sample matrix $\boldsymbol{A}$ and $\boldsymbol{A'}$ of dimension $(N \times n)$ needs first to be produced following standard normal distribution, where $\boldsymbol{A}$ can be regarded as the sample matrix of $\boldsymbol{x}$, and $\boldsymbol{A'}$ is the sample matrix of $\boldsymbol{x'}$. Then the corresponding sample values of $\mu_{\mathcal{M}y}\left(\boldsymbol{x}\right)$ and $cov_y\left(\boldsymbol{x}, \boldsymbol{x'}\right)$ can be computed by calling their closed-form expressions. These sample values can be used for estimating the posterior mean and variance of $\mathbb{E}_{\mathcal{D}}\left[\hat{V}_{Ti}\right]$ and $\mathbb{V}_{\mathcal{D}}\left[\hat{V}_{Ti}\right]$. The upper bound of $\mathbb{V}_{\mathcal{D}}\left[\hat{V}_{Ti}\right]$ can also be estimated by using only the sample matrix $\boldsymbol{A}$. While the SGI is utilized, the Gaussian-Hermite one-dimensional integration rule is recommended as the weight of the integrals are all standard Gaussian density, and then, based on these one-dimensional design points, the multi-dimensional collocation points as well as the corresponding weights are produced using the Smolyak algorithm [60]. One notes that both the implementation of MCS and SGI do not require any call of the model function, and the integrands are all infinitely smooth, thus these integrals can be accurately estimated with low computational cost.

From Corollary 3 it is known that the dimension of the induced GPR model for the total variance V_y is n, indicating that the dimensions of the integrals for computing the posterior mean and variance of V_y are n and $2n$ respectively. These two integrals can be similarly computed by using MCS based on the sample matrices $\boldsymbol{A}$ and $\boldsymbol{A'}$ or using SGI with Gaussian-Hermite one-dimensional integration rule.

One can also use the BPI to compute the integrals involved in Lemma 1 and Lemma 2 to any extent of accuracy, just like estimating $\mathcal{M}_0$ (see Eqs. (9) and (10)), as the integrands are all in closed form and infinitely smooth, whose values can be precisely computed without calling the model function. With this scheme, a posterior mean and a posterior variance will be generated for each of the posterior mean and variance of the (partial) variance terms. To avoid confusing the reader, we use MCS and/or SGI instead of BPI in this paper.

The numerical implementation introduced in this subsection is applicable when an arbitrary training

data set $\mathcal{D}$ is given, thus it is named as data-driven BPI. Note that in some publications (e.g., Ref. [20]), the terminology 'data-driven' refers to methods that require only measured data. However, in this work, the development is not directly applicable to this case as the probability distribution functions of model inputs are required for transforming them into the standard normal space. The terminology 'data-driven' used here is for making a difference with the adaptive design strategy introduced in the next section. If the users would like to apply the above data-driven BPI procedure to measured data, an additional step for statistically inferring the probability distribution functions of the input variables from the measured data is required before using the data-driven BPI approach.

It should be furthermore noted that, although the BPI approach is established based on GPR model, it is philosophically different with the standard GPR framework coupled with standard MCS, which is implemented in Refs. [47, 61]. The MCS estimators, based on model response samples, are computed by calling the GPR model. This procedure as such introduces statistical errors caused by the limited sample size, as well as numerical errors due to the general lack of fit of the GPR model to the true model function. The statistical errors can be measured by the COVs of the MCS estimates, but the second kind of errors cannot be properly measured. In the BPI framework, the second kind of errors is summarized by the posterior COVs of the HDMR components and those of the (partial) variance terms. Another superiority of the BPI procedure is that the spatial correlation information characterized by the posterior covariance of the GPR model is integrated to the cubature rules for improving the integration accuracy. This is also the philosophical difference between the Bayesian numerical analysis and the stochastic simulation for multi-dimensional integration. One can refer to the state-of-the-art developments of BPI in Refs. [48, 50, 51] for more detailed discussions and numerical demonstrations of the above superiority of BPI. It is also the above feature of BPI that enables us to develop an adaptive experiment design strategy for improving the convergence rate of the method, which is presented in the next section.

4. Adaptive Experiment Design

Until now, we have generated the analytical expressions of the posterior means and posterior variances for all the (partial) variance terms following the Bayesian inference scheme based on the training data $\mathcal{D}$. Based on these results, a data-driven method is established for estimating the variance-based sensitivity indices. However, in real-world applications, the sensitivity analysis may also be implemented for computer simulators such as finite element models, which makes it possible to design the training data $\mathcal{D}$. With this scheme, the algorithm is no longer data-driven, but it allows us to achieve the accurate estimation of the sensitivity indices with less data, and thus less model function calls, which can be of great significance if the simulators are expensive to evaluate. Inspired by this, an adaptive experiment design strategy originally developed by some of the authors in Ref. [51] is introduced here for active learning of the sensitivity indices.

The core of an adaptive design strategy is the so-called learning function or acquisition function, as it serves as the engine of the algorithm and determines the speed of convergence. The learning function utilized in this work is the weighted Posterior Variance Contribution (PVC) function [51], which is defined as:

$$
\begin{aligned}
\mathcal{L}^{\mathrm{PVC}}\left(\boldsymbol{x}\right) &= \boldsymbol{\phi}_n\left(\boldsymbol{x}\right)\mathbb{E}'_{1:\boldsymbol{n}}\left[cov_{\mathcal{M}}\left(\boldsymbol{x},\boldsymbol{x}'\right)\right] \\
&= \boldsymbol{\phi}_n\left(\boldsymbol{x}\right)\left(\mathbb{E}'_{1:\boldsymbol{n}}\left[\kappa\left(\boldsymbol{x},\boldsymbol{x}'\right)\right] - \kappa\left(\boldsymbol{x},\mathcal{X}\right)^\top\mathcal{K}^{-1}\mathbb{E}'_{1:\boldsymbol{n}}\left[\kappa\left(\boldsymbol{x}',\mathcal{X}\right)\right]\right).
\end{aligned}
\tag{28}
$$

where $\phi_n(\boldsymbol{x})$ denotes the standard Gaussian probability function in n dimensions. Based on the second row of Eq. (28), the closed-form expressions of the PVC function can be easily derived. In (28), the expectation $\mathbb{E}'_{1:\boldsymbol{n}}\left[\kappa\left(\boldsymbol{x}',\mathcal{X}\right)\right]$ has been explicitly formulated in Eq. (11), and the second expectation $\mathbb{E}'_{1:\boldsymbol{n}}\left[\kappa\left(\boldsymbol{x},\boldsymbol{x}'\right)\right]$ is expressed as:

$$
\mathbb{E}'_{1:\boldsymbol{n}}\left[\kappa\left(\boldsymbol{x},\boldsymbol{x}'\right)\right] = \sigma_0^2\left|\Sigma^{-1}+I\right|^{-1/2}\exp\left[-\frac{1}{2}\boldsymbol{x}\left(\Sigma+I\right)^{-1}\boldsymbol{x}^\top\right].
\tag{29}
$$

It is shown by the first row of Eq. (28) that the PVC function accumulates the correlation information of the GPR prediction at the arbitrary point $\boldsymbol{x}$ with those at all the other points across the full support of the input variables. It can also be found that the integral of the PVC function over the support of $\boldsymbol{x}$ equals to the posterior variance $\sigma_{\mathcal{M}0}^2$ of the constant HDMR component $\hat{\mathcal{M}}_0$. Therefore, the PVC function can be interpreted a measure of the contribution of the GPR predictor error at the point $\boldsymbol{x}$ to the posterior covariance $\sigma_{\mathcal{M}0}^2$ with the consideration of its correlation information with all the other points across the whole support of the input variables. Accordingly, by adding the point of $\boldsymbol{x}$ with the largest PVC value to the training data set $\mathcal{D}$, it is expected that the most reduction of the posterior variance $\sigma_{\mathcal{M}0}^2$ (thus the discretization error) can be achieved. One notes that the PVC function is in closed form and infinitely smooth, thus its global maximum is easy to compute. However, in most cases, the PVC function shows multimodal behavior (see the results of example 1 in Ref. [51] for detail), thus the global optimization algorithms, such as particle swarm [62], is recommended in order to avoid local convergence. Note that the behavior of the PVC function changes at each iteration, as shown in the results of example 1 in Ref. [51] for the evolution of a one-dimensional PVC function.

The other acquisition functions, e.g., those reviewed by Ref. [63], can also be used, but the comparison of their relative merits is not the focus of this work. It should also be noted that the definition of PVC in Eq. (28) applies for any form of kernels, but the closed-form expression given in Eq. (29) is derived based on the squared exponential kernel. For other type of kernels, the corresponding closed-form expressions can be obtained from Ref. [50].

Following the PVC acquisition function, an active learning algorithm is developed for adaptively inferring the sensitivity indices as follows:

Step 1. Create an initial training sample matrix of size $N_0 \times n$ for the input variables by using simple random sampling or Latin-hypercube Sampling (LHS) design. Compute the corresponding model function values at these initial design points, and initialize the training data set $\mathcal{D}$ with these sample points. Record the number of model function calls as $N = N_0$.

Step 2. Train or update the GPR model with $\mathcal{D}$.

Step 3. Evaluate the posterior mean $\mu_{\mathcal{M}0}$ and posterior variance $\sigma^2_{\mathcal{M}0}$ of the constant HDMR component $\hat{\mathcal{M}}_0$ using the closed-form expressions presented in Section 3.

Step 4. Judge whether the stopping criteria is satisfied or not by examining the posterior coefficient variation (COV) of $\hat{\mathcal{M}}_0$. If it is not satisfied, compute the next optimal design point by maximizing the PVC function, evaluate the corresponding model function value, add it to the training data set $\mathcal{D}$, let $N = N + 1$, and go to Step 2; otherwise go to Step 5.

Step 5. Compute the posterior means and variances for (partial) variance terms V_y, V_i and V_{Ti}. If necessary, check if the accuracy for estimating these (partial) variance terms is satisfied. If not, one can continue the active learning process by searching the global maxima of the PVC function; Otherwise, end the algorithm.

In Step 1, the initial samples can be generated following either uniform distribution within the support $[-3, 3]$ or standard Gaussian distribution. The posterior COV utilized in Step 4 is computed by $\sigma_{\mathcal{M}0}/\mu_{\mathcal{M}0}$. The stopping threshold can be identified by the analysts based on their requirement of accuracy. Following our previous work [51], it is suggested to use a delayed analysis scheme, which means ending the algorithm only when the stopping criteria is satisfied for several times (e.g., twice) in succession, in order to avoid pseudo convergence which may appear during the early stage of active learning when the rough behavior of model function is not well captured by the GPR model.

5. Test Examples and Applications

5.1. An Illustrative Example

Considering a two-dimensional model with g-function formulated as:

$$g(x_1, x_2) = \sum_{i=1}^{4} c_i \exp\left[-\alpha_{i1}(x_1 - \beta_{i1})^2 - \alpha_{i2}(x_2 - \beta_{i2})^2\right] \tag{30}$$

where $\boldsymbol{\alpha} = \left(\begin{smallmatrix} 2 & 3 & 1 & 4 \\ 3 & 2 & 4 & 1 \end{smallmatrix}\right)^\top$, $\boldsymbol{\beta} = \left(\begin{smallmatrix} -0.5 & 0.5 & -0.5 & 0.5 \\ -0.5 & -0.5 & 0.5 & 0.5 \end{smallmatrix}\right)^\top$, $\boldsymbol{c} = \left(\begin{smallmatrix} 1 & -1.5 & -1.5 & 2 \end{smallmatrix}\right)^\top$, x_1 and x_2 are independent standard normal random variables. This is a highly nonlinear model with large interaction effects, and the variance-based sensitivity indices can be analytically derived to provide comparison.

For implementing the adaptive BPI, the stopping criteria is set to be $\sigma_{\mathcal{M}0}/\mu_{\mathcal{M}0} \leq 0.2$, and the algorithm stops only when this criteria is satisfied for two times in succession. For illustrating the superiority of the

adaptive experiment design, we use $N_0 = 10$ initial samples to initialize both data-driven and adaptive BPI algorithms, where for data-driven BPI, the training data set is enriched by random sampling, while for adaptive BPI, it is enriched by searching the maximum point of the PVC function. The integrals (see Lemma 1 and Lemma 2) for estimating the posterior means and variances for each (partial) variance terms are numerically computed by MCS with 10^4 samples.
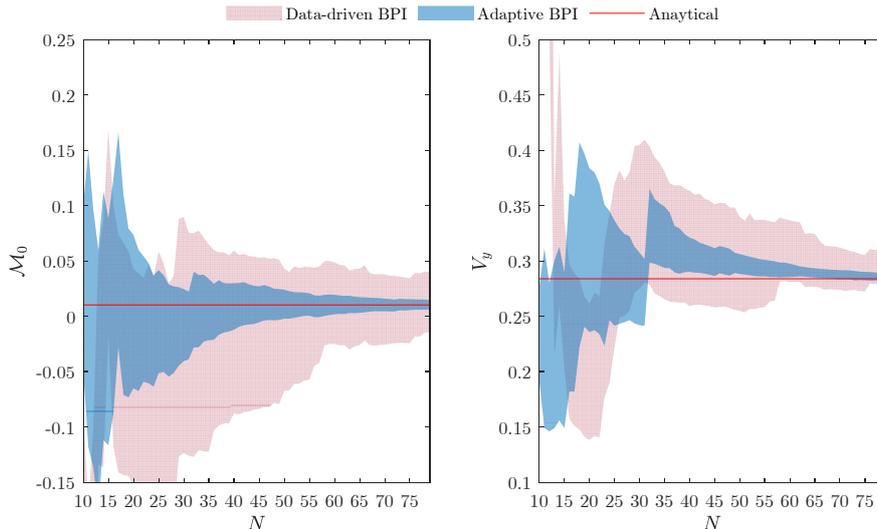
Figure 1: Point-wise evolution of the posterior features for the constant component $\mathcal{M}_0$ and the total variance V_y against the training data size for the illustrative example, where the vertical sections of the filled areas indicate the interval $[\mu_{\mathcal{M}0} - 2\sigma_{\mathcal{M}0}, \mu_{\mathcal{M}0} + 2\sigma_{\mathcal{M}0}]$ (left) and $\left[\mathbb{E}_{\mathcal{D}}\left[\hat{V}_y\right] - 2\sqrt{\mathbb{V}_{\mathcal{D}}\left[\hat{V}_y\right]}, \mathbb{E}_{\mathcal{D}}\left[\hat{V}_y\right] + 2\sqrt{\mathbb{V}_{\mathcal{D}}\left[\hat{V}_y\right]}\right]$ (right).

With the initial 10 training points, 69 more training points are produced adaptively by maximizing the PVC function before reaching the stopping criteria. Similarly, for data-driven BPI, 69 more training points are generated by random sampling without adaptive design. For each iteration step, the posterior characters of the model response expectation $\hat{\mathcal{M}}_0$ and the variance $\hat{V}_y$ are recorded, and compared in Figure 1. One notes that, given the GPR model $\hat{\mathcal{M}}(x)$, $\hat{\mathcal{M}}_0$ is a Gaussian variable, while $\hat{V}_y$ is also a random variable, but not Gaussian. Thus, the interval $[\mu_{\mathcal{M}0} - 2\sigma_{\mathcal{M}0}, \mu_{\mathcal{M}0} + 2\sigma_{\mathcal{M}0}]$ shown in Figure 1 is a 95.45% confidence interval, but $\left[\mathbb{E}_{\mathcal{D}}\left[\hat{V}_y\right] - 2\sqrt{\mathbb{V}_{\mathcal{D}}\left[\hat{V}_y\right]}, \mathbb{E}_{\mathcal{D}}\left[\hat{V}_y\right] + 2\sqrt{\mathbb{V}_{\mathcal{D}}\left[\hat{V}_y\right]}\right]$ is not. In any case, this posterior interval can be used to monitor the convergence of results. It is also noticed that the posterior confidence interval of the variance term V_y shows more bias than that of the constant HDMR component $\mathcal{M}_0$. This is also caused by the fact that, under the GPR approximation of $\mathcal{M}(\boldsymbol{x})$, $\hat{\mathcal{M}}_0$ follows a Gaussian distribution, whereas $\hat{V}_y$ follows a non-Gaussian distribution, which is also non-symmetric. In this paper, we don't investigate the exact posterior distribution of the variance and the partial variance terms as their posterior variances are sufficient for summarizing the numerical errors involved in their posterior mean estimates.

Figure 1 shows that the posterior distribution supports of both $\hat{\mathcal{M}}_0$ and $\hat{V}_y$ generated by data-driven

and adaptive BPI contain the corresponding true values with high probability, indicating the correctness of the Bayesian inference. It is also shown that the posterior distributions generated by both data-driven and adaptive BPI converge to the true values as the training data size increases, but the adaptive BPI shows much higher speed of convergence than the data-driven BPI, implying that the BPI is effective for inferring the model response expectation and variance, and with the adaptive experiment design driven by the PVC function, the convergence rate can be further improved. The above features make the BPI appealing for uncertainty quantification in both data-driven scenario and that of scientific computation involving expensive simulators. Although searching the global maximum point of the PVC function involves computational cost, it is much cheaper than that of evaluating the response value of the expensive simulator, thus is negligible. Take the first adaptively designed point as an example: using a personal laptop, the time consumed for computing the global maxima of the PVC function is only 0.103 second.
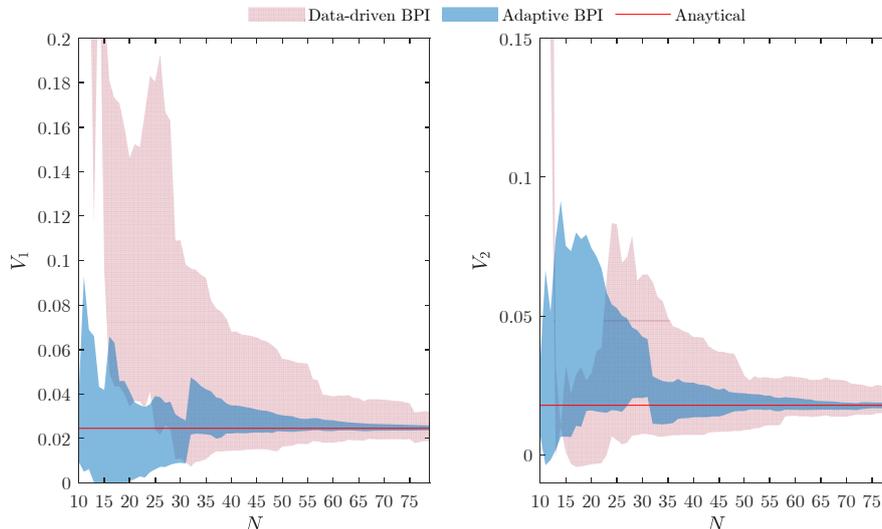
Figure 2: Evolution of the posterior features of the first-order partial variance V_i against the training data size for the illustrative example, where the vertical sections of the filled areas indicate the interval $\left[\mathbb{E}_{\mathcal{D}}\left[\hat{V}_i\right] - 2\sqrt{\mathbb{V}_{\mathcal{D}}\left[\hat{V}_i\right]}, \mathbb{E}_{\mathcal{D}}\left[\hat{V}_i\right] + 2\sqrt{\mathbb{V}_{\mathcal{D}}\left[\hat{V}_i\right]} \right]$.

We then discuss the results of the first-order partial variances, for which the evolution of the posterior features against the training data size is shown in Figure 2. As can be seen from the results generated by the data-driven BPI, when the training data size exceeds 50, the posterior distributions of both $\hat{V}_1$ and $\hat{V}_2$ shrinks to the true values. It is also shown that the adaptive BPI shows a much better convergence than the data-driven BPI, thus is more applicable to expensive simulators. The results generated by both methods with 77 training points show perfect agreements with the analytical solutions.

Next, let's focus on the posterior features of the first-order HDMR component functions, whose posterior intervals generated by both data-driven and adaptive BPI with all the 79 training points are shown in Figure

18
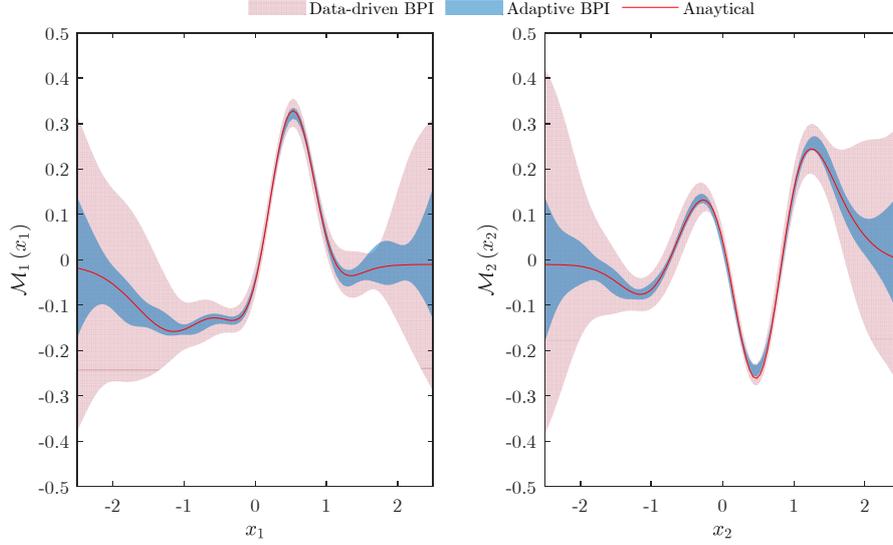
Figure 3: The 95.45% posterior confidence intervals of the first-order HDMR component functions of the illustrative example computed by data-driven and adaptive BPI algorithms, both of which with 79 training points.

3, together with the analytical results for comparison. Reminding that, given the GPR model $\hat{\mathcal{M}}(\boldsymbol{x})$, $\hat{\mathcal{M}}_i(x_i)$ is a univariate GPR model, the intervals given in Figure 3 are actually 95.45% confidence intervals. As can be seen, the confidence intervals generated by both data-driven and adaptive BPI contain the true results across the full support of x_i, indicating the correctness of the Bayesian inference. It is also shown that, with adaptive experiment design, the quality of inference for the HDMR component functions has also been largely improved.

The evolution of the posterior intervals for the total partial variances is schematically shown in Figure 4. As can be seen, although the posterior intervals in this figure are not as smooth as those in Figure 1 and Figure 2, but the posterior intervals generated by both methods approach the true value robustly and accurately, demonstrating the correctness and effectiveness of the Bayesian inference for total effect indices.

The converged results generated by both the data-driven and adaptive BPI are summarized in Table 2, together with the reference solutions. As can be seen, the upper bounds generated in Lemma 1 provide conservative estimations of the posterior variances (thus the discretization errors) for those (partial) variance terms. With the same number of training points, the adaptive BPI provides much accurate estimations for all the variance terms, and thus the sensitivity indices. Therefore, the data-driven is recommended when only supervised data is available or when the computer simulators are extremely cheap to compute; otherwise, if the simulators are computationally expensive, the adaptive BPI is recommended.
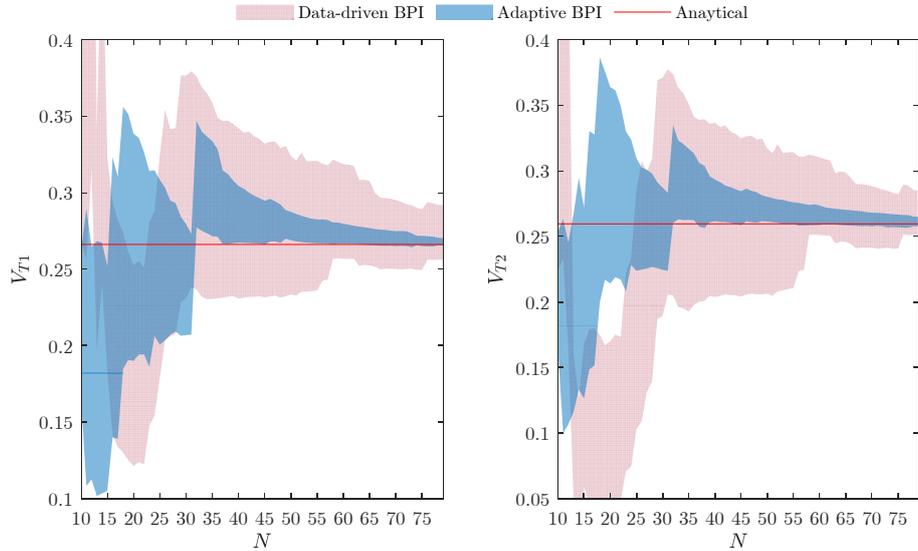
Figure 4: Evolution of the posterior features of the total partial variance V_{Ti} against the training data size for the illustrative example, where the vertical sections of the filled areas indicate the interval $\left[\mathbb{E}_{\mathcal{D}} \left[\hat{V}_{Ti} \right] - 2 \sqrt{ \mathbb{V}_{\mathcal{D}} \left[\hat{V}_{Ti} \right] }, \mathbb{E}_{\mathcal{D}} \left[\hat{V}_{Ti} \right] + 2 \sqrt{ \mathbb{V}_{\mathcal{D}} \left[\hat{V}_{Ti} \right] } \right]$.

5.2. A Dam Seepage Model

Consider a dam seepage model which has been utilized by some of the authors for demonstrating reliability sensitivity analysis [9] and the effectiveness of the PVC function for integration [51]. The layout of the structure is shown in Figure 5. The model output of interest is the confined seepage under a dam, which rests over a soil foundation consisting of two permeable layers and one impermeable layer. Denote the permeability of the silty sand layer along the vertical and horizontal directions as $k_{yy,1}$ and $k_{xx,1}$ respectively and those f the silty gravel layers as $k_{yy,2}$ and $k_{xx,2}$ respectively. The water depth in denoted by h_D. All these five parameters are assumed to be input random variables, and their distribution information is reported in

Table 2: Results of the illustrative example.

Methods	Indices	V_y	V_1	V_2	V_{T1}	V_{T2}	N
Data-driven BPI	Posterior means	0.2936	0.0254	0.0193	0.2740	0.2682	$10 + 69 = 79$
	Posterior COVs	0.0245	0.1311	0.1087	0.0324	0.0310	
	Upper bounds of Posterior COVs	0.0765	0.1874	0.2371	0.0850	0.0931	
Adaptive BPI	Posterior means	0.2858	0.0247	0.0178	0.2678	0.2612	$10 + 69 = 79$
	Posterior COVs	0.0051	0.0227	0.0272	0.0045	0.0070	
	Upper bounds of Posterior COVs	0.0401	0.0706	0.0830	0.0470	0.0473	
Analytical	–	0.2841	0.0245	0.0179	0.2662	0.2596	–

Table 3. The model is governed by a PDE formulated as:

$$k_{xx,1}\frac{\partial^2 h_W}{\partial x^2} + k_{yy,i}\frac{\partial^2 h_W}{\partial^2 y} = 0, \qquad i = 1, 2 \tag{31}$$

with h_W indicating the hydraulic head segment. A detailed description of the model as well as the boundary conditions can be found in Ref. [9]. A FE model is built with 1628 quadratic triangular elements to solve the PDE, and predict the value of h_W. The seepage p is then computed from the value of h_W. For example, the seepage among the CD segment is computed as:

$$q = -\int_{\mathrm{CD}} k_{yy,2}\frac{\partial h_W}{\partial y}\mathrm{d}x. \tag{32}$$

The response concerned here is exactly q multiplied by 10^6.

Figure 5: Sketch of a dam seepage model.

Table 3: Probability distribution information of the five random input variables of the dam seepage model.

Input variables	$x_1 = k_{xx,1}$	$x_2 = k_{yy,1}$	$x_3 = k_{xx,2}$	$x_4 = k_{yy,2}$	$x_5 = h_D$
Distribution type	Lognormal	Lognormal	Lognormal	Lognormal	Uniform
Distribution Parameters	$\mu = \sigma = 5 \times 10^{-7}$	$\mu = \sigma = 2 \times 10^{-7}$	$\mu = \sigma = 5 \times 10^{-6}$	$\mu = \sigma = 2 \times 10^{-6}$	[7, 10]
Units	[m/s]	[m/s]	[m/s]	[m/s]	[m]

The stopping threshold is set to be 0.15%, and 10 samples generated by LHS design are used to initialize the adaptive BPI. The algorithm adaptively produces 38 more training points before reaching the stopping criteria, thus the total number of model function calls is 48. The data-driven BPI is also implemented with 48 training data generated by LHS design. For comparison, we also compute the (partial) variance terms

21

with the MCS estimators reported in Ref. [39]. The results computed by the three algorithms are then compared in Table 4. It is shown that the posterior means of all terms produced by both data-driven BPI and adaptive BPI match well with the reference solutions generated by MCS, but the results of adaptive BPI show smaller posterior COVs, indicating that, with the same number of model function calls, the adaptive BPI produces more robust inference for the sensitivity indices. It can also be seen that, for this model, the first input parameter, i.e. $k_{xx,1}$ accounts for the most contribution to model response variance, followed by $k_{yy,1}$, and then $k_{xx,2}$. It is also seen that the interaction contributions among these five variables are much smaller than their respective individual contribution.

Table 4: Results for the dam seepage model.

Indices	Data-driven BPI		Adaptive BPI		MCS [39]	
	Posterior means	Posterior COVs	Posterior means	Posterior COVs	Mean estimates	COVs
V_y	1.380	0.0304	1.335	0.0150	1.346	0.0144
V_1	0.8988	0.0356	0.8623	0.0148	0.8863	0.0359
V_2	0.1716	0.0581	0.1932	0.0387	0.1969	0.0638
V_3	0.0718	0.0644	0.0833	0.0726	0.0705	0.1139
V_4	0.0410	0.0992	0.0417	0.0946	0.0535	0.1300
V_5	0.0635	0.0630	0.0539	0.0871	0.0480	0.1284
V_{T1}	1.003	0.0412	0.9407	0.0171	0.9618	0.0156
V_{T2}	0.2270	0.0593	0.2523	0.0319	0.2549	0.0189
V_{T3}	0.1005	0.0584	0.1144	0.0573	0.1044	0.0179
V_{T4}	0.0623	0.0887	0.0639	0.0961	0.0776	0.0334
V_{T5}	0.0903	0.0703	0.0705	0.0895	0.0630	0.0150
N	10+38=48		10+38=48		1.4×10^5	

5.3. A turbine blade model

Consider a jet turbine blade model adapted from the Matlab PDE toolbox. The FE model is shown in Figure 6(a). The blades are of vital importance for a jet engine as the failure (usually caused by fatigue) often results in fatal accidents. The computation of the thermal stress and the deformation of the blade is thus of vital importance. However, due to the uncertainties presented in the material parameters as well as loads and working temperature, these two important quantities often show large dispersion. Then the aim of this application is to investigate the effects of these input uncertainties on the maximal stress. The blades are usually made of nickel-based alloy (e.g., NIMONIC 90) as they must resist the extremely high temperature of the gasses. Working in such high temperature, the material will expand significantly, and thus large mechanical stress and deformation are inevitable. The FE model in Figure 6(a), with tetrahedral mesh, is then developed for simulating these two responses of the structures. There are three steps of implementation

for the deterministic numerical simulation. First, perform structural analysis with the consideration of only the pressure loads; second, perform thermal stress analysis accounting only the thermal effects; third, combine the structural analysis and thermal analysis to compute the maximum stress. The results of the these three steps performed at the mean values of the input parameters are shown in Figure 6(b)-(d). The model output of concern here is the maximum stress, as can be seen in Figure 6(d). The Young's modulus (denoted as x_1) is assumed to be Gaussian random variable with mean values 227 [GPa]; the coefficient of thermal expansion (denoted as x_2) follows Gaussian distribution with mean 12.7×10^{-6}[1/K]; the Poisson's ratio (denoted as x_3) also follows Gaussian distribution with mean 0.27; the thermal conductivity (denoted as x_4) follows a Gaussian distribution with mean 11.5 [W/m/K]; the pressure loads on the pressure and suction sides of the blade (denoted as x_5 and x_6 respectively) follow Gaussian distribution with means 800 [KPa] and 600 [KPa] respectively. The COVs of all the above six random variables are assumed to be 0.1. One notes that above assumption of probability distribution is subjective, which means that the uncertainty is epistemic, and comes from the lack of data on these input variables. Then, the target is to find which variables are responsible for the uncertainty of model output. With a personal laptop, each call of the FE model consumes about 11 seconds, which is computationally more expensive than the numerical example reported previously. Therefore, we use the data-driven BPI and adaptive BPI to estimate the sensitivity indices.

The stopping threshold for implementing the adaptive BPI is set to be 0.1%, and the prior mean of the GPR model is set to be constant. With six samples generated by LHS design as initial training points, the adaptive BPI produces sixteen more design points before fulfilling the stopping criteria twice in succession. Thus the total number of model function calls for implementing the adaptive BPI procedure is twenty-two. For comparison, the data-driven BPI is implemented with 40 random training points generated by LHS design. The results are then reported in Table 5. As can be seen, the results produced by the two methods match well, but those computed with adaptive BPI procedure show smaller posterior variance, thus are more accurate.

It can be seen from Table 5 that, the summation of the main partial variance almost equal to the total variance, and the total partial variance of each input variable almost equal to its main partial variance. The above two phenomenons both indicate that the model function is additive, i.e., no interaction effect exists among input variables. It can also be concluded that, among the six input variables, the Young's modulus x_1 and the coefficient x_2 of thermal expansion make the largest contribution to the model output variance, followed by the pressure load x_5. All the other three variables almost have no contribution to the model output uncertainty. Therefore, for reducing the epistemic uncertainty of model output, one needs to collect information on x_1 and x_2.
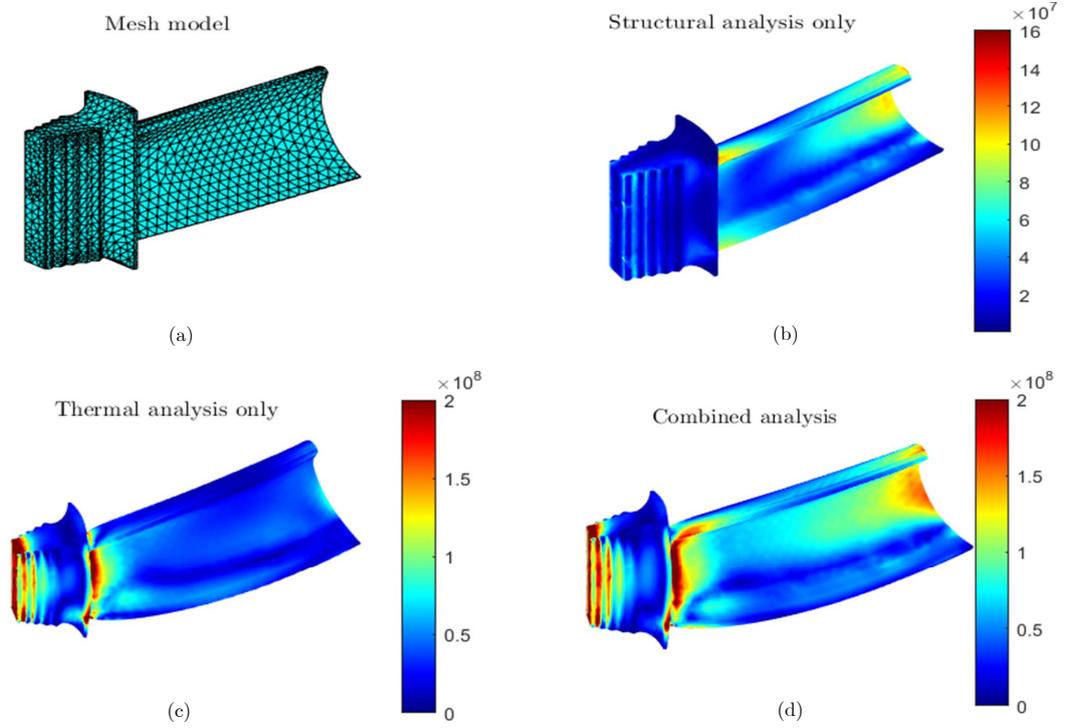
Figure 6: The FE model and the deterministic analysis results of the turbine blade.

5.4. Dynamic analysis of a satellite

Consider the FE model of a satellite as shown in Figure 7, which has been taken from Refs. [64, 65]. The satellite is composed of a cubic core, which is reinforced by internal stiffening beams. Reflectors and photovoltaic panels are linked to this cubic core by means of connecting beams. The finite element model comprises beam and shell elements, leading to a total of 1262 elements and 7146 degrees-of-freedom. It is considered that the Young's moduli associated with the different parts of the satellite are uncertain and modeled as log-normal random variables. The mean value of the Young's moduli for the different parts of the satellite are as follows: cubic core (μ_{E_1}) and photovoltaic panels (μ_{E_2}) $\mu_{E_1} = \mu_{E_2} = 6.89 \times 10^9$[Pa]; reflectors $\mu_{E_3} = 8 \times 10^{10}$[Pa]; connecting beams $\mu_{E_4} = 8 \times 10^{10}$[Pa]; and stiffening beams $\mu_{E_5} = 8 \times 10^{11}$[Pa]. The coefficient of variation associated with each random variable is equal to 20%. The computation of the first natural frequency of the satellite is of vital importance as its value should be kept away from the frequency of ambient vibration. However, due to the randomness of the above-mentioned material properties, the

Table 5: Results for the turbine blade model.

Indices	Data-driven BPI		Adaptive BPI	
	Posterior means	Posterior COVs	Posterior means	Posterior COVs
V_y	1.4037	0.0130	1.4114	0.0097
V_1	0.6352	0.0212	0.6229	0.0121
V_2	0.6820	0.0172	0.7009	0.0176
V_3	0.0103	0.1205	0.0122	0.0568
V_4	0.0053	0.1055	0.0068	0.0893
V_5	0.0528	0.1465	0.0357	0.0436
V_6	0.0139	0.1305	0.0188	0.0652
V_{T1}	0.6443	0.0199	0.6282	0.0103
V_{T2}	0.6948	0.0174	0.7138	0.0176
V_{T3}	0.0110	0.2367	0.0128	0.0540
V_{T4}	0.0056	0.4951	0.0078	0.0966
V_{T5}	0.0614	0.1568	0.0354	0.0325
V_{T6}	0.0163	0.1626	0.0192	0.0556
N	40		6+16=22	

first natural frequency of the structure also involves uncertainty. It is therefore necessary to quantify this uncertainty, and measure the contribution of each input uncertainty on the output uncertainty. For this purpose, we estimate the variance of the first natural frequency as well as Sobol' main and total effect indices with respect to each uncertain input variable. The computational cost of each FE model call is definitely more expensive than the first numerical example, but less expensive than the blade model, thus the MCS is also implemented as reference solutions.
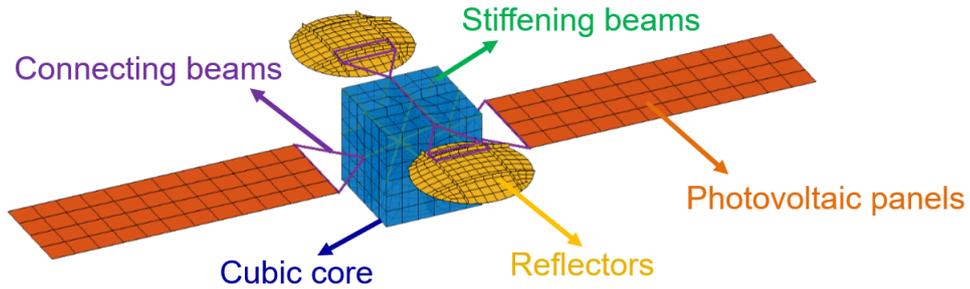
Figure 7: The FE model of a satellite.

For this example, it is found that the data-driven BPI can accurately estimate all the (partial) variance terms with small number of training points, thus there is no need to implement the adaptive BPI. The results

by data-driven BPI with twenty training points are reported in Table 6, together with the MCS estimates for comparison. It should be noted that, as the MCS estimators are unbiased, the small values of COVs reported in the last column of Table 6 indicates that the MCS estimates are all accurate and can be served as reference solutions. As can be seen, the posterior mean estimates generated by BPI show good agreements with the reference solutions, and the posterior COVs reported in the third column of Table 6 are all small enough, proving the accuracy of the BPI approach for this example. Although for the very small partial variances, such as V_3, the posterior COV is large, considering its posterior mean estimate (3.4376×10^{-9}), the errors can be omitted.

It is also shown that, the Young's modulus of the connecting beams (E_4) is one affecting the most for the model output variance, followed by that of photovoltaic panel (E_2), and those of the other three components are not influential. This conclusion matches well with the physical intuition when examining the first mode shape, as it comprises translation of the photovoltaic panel.

Table 6: Results for the satellite model.

Indices	Data-driven BPI		MCS [39]	
	Posterior means	Posterior COVs	Mean estimates	COVs
V_y	0.0254	0.0009	0.0249	0.0200
V_1	0.0000	0.7943	0.0000	14.495
V_2	0.0017	0.0024	0.0029	0.5651
V_3	0.0000	31.509	0.0000	9.6961
V_4	0.0237	0.0010	0.0227	0.2928
V_5	0.0000	0.8758	0.0000	17.801
V_{T1}	0.0000	1.5425	0.0000	0.0256
V_{T2}	0.0018	0.0032	0.0017	0.0229
V_{T3}	0.0000	69.174	0.0000	0.0230
V_{T4}	0.0238	0.0009	0.0234	0.0201
V_{T5}	0.0000	1.6575	0.0000	0.0250
N	20		3.5×10^4	

6. Conclusions and discussions

The estimation of the variance-based sensitivity indices is regarded as an statistical inference problem in this work, and based on a set of supervised training data, the posterior features (including means and variances) for all the (partial) variance terms involved in the sensitivity indices are analytically derived following two newly developed first principles and the rationale of BPI. Although the posterior distributions of these (partial) variance terms are no longer Gaussian, these inferred posterior features provide sufficient

information on the values of the sensitivity indices as well as the related discretization errors. Based on those posterior features, two BPI strategies are then developed for estimating the sensitivity indices. The first strategy, i.e., the data-driven BPI, is applicable to data, and does not require any design of experiments and computer simulators. This strategy is also recommended if the computer simulators are very cheap to evaluate. The second strategy, i.e., the adaptive BPI, is especially developed for computer simulators, where the training data used for inference can be designed, and this strategy is strongly recommended if the simulators are expensive to evaluate as it can apparently reduce the required number of model function calls to achieve estimates of same accuracy.

The results of the numerical and engineering examples, on the one hand, demonstrated the correctness of the analytical expressions of all the posterior features, and on the other hand, proved the effectiveness of the two strategies. It is also shown that the quality of the estimates of both methods relies on that of the trained GPR model. With the adaptive experiment design, the adaptive BPI strategy can not only improve the convergence rates of the sensitivity indices, but also that of the GPR model, as revealed by the results for first-order HDMR component functions. It is also noted that, for different examples, different values of stopping criteria are specified. For real-world applications, this threshold value can be determined based on the users' error tolerance on the estimated sensitivity indices.

The main and total effect indices have been treated in this work, however, if required, other sensitivity indices such as the second-order partial variance V_{ij} can be similarly inferred following the two principles and two operating rules of the GPR model reported in subsection 3.1. It needs to be emphasized that the two BPI approaches cannot be directly applied to high-dimensional cases (e.g., with $n > 30$). This is mainly because the Euclidean distance used in the kernel function of the GPR model is not informative in high-dimensional space [66]. This is left for future research. Besides, the source codes of the two BPI algorithms are available upon reasonable request to the corresponding author.

Acknowledgment

Appendix A. Proof of Lemma 1 and Lemma 2

Based on the total variance law, the posterior mean $\mathbb{E}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right]$ can be decomposed as:

$$
\begin{aligned}
\mathbb{E}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right] &= \mathbb{V}_{\mathcal{D},\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right] - \mathbb{V}_{\mathcal{D}}\left[\mathbb{E}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right] \\
&= \mathbb{V}_{\boldsymbol{u}}\left[\mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right] + \mathbb{E}_{\boldsymbol{u}}\left[\mathbb{V}_{\mathcal{D}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right] - \mathbb{V}_{\mathcal{D}}\left[\mathbb{E}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right] \; .
\end{aligned}
\tag{A.1}
$$

Noting that in the last term $\mathbb{E}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right] = 0$, Eq. (A.1) can be further derived as:

$$
\mathbb{E}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right] = \mathbb{V}_{\boldsymbol{u}}\left[\mu_{\mathcal{G}}\left(\boldsymbol{u}\right)\right] + \mathbb{E}_{\boldsymbol{u}}\left[\sigma_{\mathcal{G}}^2\left(\boldsymbol{u}\right)\right]
\tag{A.2}
$$

which is exactly Eq. (12).

The posterior variance $\mathbb{V}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right]$ is then derived as:

$$
\begin{aligned}
&\mathbb{V}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right] \\
&= \int \left[\int \hat{\mathcal{G}}^2\left(\boldsymbol{u}\right)\phi\left(\boldsymbol{u}\right) du - \int \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right)\right]\phi\left(\boldsymbol{u}\right)\mathrm{d}u\right]^2 f\left(\hat{\mathcal{G}}\,|\mathcal{D}\right)\mathrm{d}\hat{\mathcal{G}} \\
&= \int \left[\int \left(\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right) - \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right)\right]\right)\phi\left(\boldsymbol{u}\right)\mathrm{d}\boldsymbol{u}\right]\left[\int \left(\hat{\mathcal{G}}^2\left(\boldsymbol{u}'\right) - \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}'\right)\right]\right)\phi\left(\boldsymbol{u}'\right)\mathrm{d}\boldsymbol{u}'\right] f\left(\hat{\mathcal{G}}\,|\mathcal{D}\right)\mathrm{d}\hat{\mathcal{G}} \\
&= \int \int \int \left(\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right) - \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right)\right]\right)\left(\hat{\mathcal{G}}^2\left(\boldsymbol{u}'\right) - \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}'\right)\right]\right) f\left(\hat{\mathcal{G}}\,|\mathcal{D}\right)\mathrm{d}\hat{\mathcal{G}}\phi\left(\boldsymbol{u}\right)\mathrm{d}\boldsymbol{u}\phi\left(\boldsymbol{u}'\right)\mathrm{d}\boldsymbol{u}' \\
&= \int \int \mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right), \hat{\mathcal{G}}^2\left(\boldsymbol{u}'\right)\right]\phi\left(\boldsymbol{u}\right)\mathrm{d}\boldsymbol{u}\phi\left(\boldsymbol{u}'\right)\mathrm{d}\boldsymbol{u}' \\
&= \mathbb{E}_{\boldsymbol{u}}\mathbb{E}'_{\boldsymbol{u}}\left[\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right), \hat{\mathcal{G}}^2\left(\boldsymbol{u}'\right)\right]\right]
\end{aligned}
\tag{A.3}
$$

the last row of which is exactly Eq. (13). The posterior covariance $\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right), \hat{\mathcal{G}}^2\left(\boldsymbol{u}'\right)\right]$ in Eq. (13) can be formulated explicitly with the posterior mean and posterior covariance of the GPR model $\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right)$. Consider two correlated normal variables $v_1 \sim \mathcal{N}\left(u_1, \sigma_1^2\right)$ and $v_2 \sim \mathcal{N}\left(u_2, \sigma_2^2\right)$ with covariance σ_{12}^2, then the covariance of x_1^2 and x_2^2 is derived as $\mathrm{cov}\left[x_1^2, x_2^2\right] = 2\sigma_{12}^4 + 4\mu_1\mu_2\sigma_{12}^2$. Thus, by regarding $\hat{\mathcal{G}}\left(\boldsymbol{u}\right)$ and $\hat{\mathcal{G}}\left(\boldsymbol{u}'\right)$ as two correlated normal variables, the posterior covariance $\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right), \hat{\mathcal{G}}^2\left(\boldsymbol{u}'\right)\right]$ can be expressed as:

$$
\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}^2\left(\boldsymbol{u}\right), \hat{\mathcal{G}}^2\left(\boldsymbol{u}'\right)\right] = 2\mathrm{cov}_{\mathcal{D}}^2\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right), \hat{\mathcal{G}}\left(\boldsymbol{u}'\right)\right] + 4\mu_{\mathcal{G}}\left(\boldsymbol{u}\right)\mu_{\mathcal{G}}\left(\boldsymbol{u}'\right)\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right), \hat{\mathcal{G}}\left(\boldsymbol{u}'\right)\right]
\tag{A.4}
$$

substituting which into Eq. (13) yields Eq. (14).

We start from Eq. (14) to prove Lemma 2. By using the Cauchy-Schwarz inequality, the following

inequality holds for the posterior covariance $\text{cov}_{\mathcal{D}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right),\hat{\mathcal{G}}\left(\boldsymbol{u}'\right)\right]$:

$$\text{cov}_{\mathcal{D}}^2\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right),\hat{\mathcal{G}}\left(\boldsymbol{u}'\right)\right] \leqslant \mathbb{V}_{\mathcal{D}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\mathbb{V}_{\mathcal{D}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}'\right)\right] = \sigma_{\mathcal{G}}^2\left(\boldsymbol{u}\right)\sigma_{\mathcal{G}}^2\left(\boldsymbol{u}'\right). \tag{A.5}$$

Substituting Eq. (A.5) to Eq. (14) yields:

$$\begin{aligned}
\mathbb{V}_{\mathcal{D}}\left[\mathbb{V}_{\boldsymbol{u}}\left[\hat{\mathcal{G}}\left(\boldsymbol{u}\right)\right]\right] &\leqslant 2\mathbb{E}_{\boldsymbol{u}}\mathbb{E}_{\boldsymbol{u}}'\left[\sigma_{\mathcal{G}}^2\left(\boldsymbol{u}\right)\sigma_{\mathcal{G}}^2\left(\boldsymbol{u}'\right)\right] + 4\mathbb{E}_{\boldsymbol{u}}\mathbb{E}_{\boldsymbol{u}}'\left[|\mu_{\mathcal{G}}\left(\boldsymbol{u}\right)||\mu_{\mathcal{G}}\left(\boldsymbol{u}'\right)|\sigma_{\mathcal{G}}\left(\boldsymbol{u}\right)\sigma_{\mathcal{G}}\left(\boldsymbol{u}'\right)\right] \\
&= 2\mathbb{E}_{\boldsymbol{u}}^2\left[\sigma_{\mathcal{G}}^2\left(\boldsymbol{u}\right)\right] + 4\mathbb{E}_{\boldsymbol{u}}^2\left[|\mu_{\mathcal{G}}\left(\boldsymbol{u}\right)|\sigma_{\mathcal{G}}\left(\boldsymbol{u}\right)\right]
\end{aligned} \tag{A.6}$$

which is exactly (15) and concludes the proof.

Appendix B. Mathematical proofs for main partial variance

The definition of the induced univariate GPR model $\hat{\mathcal{M}}_i\left(x_i\right)$ can be reformulated as:

$$\hat{\mathcal{M}}_i\left(x_i\right) = \mathbb{E}_{-i}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right] - \mathbb{E}_{1:n}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right]. \tag{B.1}$$

Then the posterior mean of $\hat{\mathcal{M}}_i\left(x_i\right)$ can be expressed as:

$$\begin{aligned}
\mu_{\mathcal{M}i}\left(x_i\right) &= \mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{M}}_i\left(x_i\right)\right] \\
&= \mathbb{E}_{-i}\left[\mathbb{E}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right]\right] - \mu_{\mathcal{M}0} \\
&= \mathbb{E}_{-i}\left[b\left(\boldsymbol{x}\right)\right] + \mathbb{E}_{-i}\left[\kappa\left(\boldsymbol{x},\mathcal{U}\right)\right]^{\top}\mathcal{K}^{-1}\left(\mathcal{Y} - b\left(\mathcal{X}\right)\right) - \mu_{\mathcal{M}0} \\
&= b_i\left(x_i\right) + \boldsymbol{\zeta}_i\left(x_i\right)^{\top}\mathcal{K}^{-1}\left(\mathcal{Y} - b\left(\mathcal{X}\right)\right) - \mu_{\mathcal{M}0}
\end{aligned} \tag{B.2}$$

which is exactly Eq. (16). For the commonly used prior mean functions, the closed-form expression for $b_i\left(x_i\right)$ is easy to derive , and we don't give more details. The closed-form expression of $\boldsymbol{\zeta}_i\left(x_i\right)$ formulated as:

$$\boldsymbol{\zeta}_i\left(x_i\right) = \sigma_0^2\left|\Sigma_{-i}^{-1} + I\right|^{-1/2}\exp\left(-\frac{1}{2}\text{vec}\left[\text{diag}\left[\mathcal{X}_{,-i}\left(\Sigma_{-i} + I_{n-1}\right)^{-1}\mathcal{X}_{,-i}^{\top}\right]\right] - \frac{1}{2\sigma_i^2}\left(x_i - \mathcal{X}_{,i}\right)^2\right). \tag{B.3}$$

Next, we derive the closed-form expressions for the posterior variance of $\hat{\mathcal{M}}_i(x_i)$. By definition,

$$
\begin{aligned}
cov_{\mathcal{M}i}(x_i, x_i') &= \mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}_i(x_i), \hat{\mathcal{M}}_i(x_i')\right] \\
&= \mathbb{E}_{\mathcal{D}}\left[\left(\hat{\mathcal{M}}_i(x_i) - \mu_{\mathcal{M}i}(x_i)\right)\left(\hat{\mathcal{M}}_i(x_i') - \mu_{\mathcal{M}i}(x_i')\right)\right] \\
&= \mathbb{E}_{\mathcal{D}}\left[\begin{array}{c}\left(\mathbb{E}_{1:n}''\left[\hat{\mathcal{M}}(\boldsymbol{x}_{-i}'', x_i) - \hat{\mathcal{M}}(\boldsymbol{x}'')\right] - \mathbb{E}_{1:n}''\left[\mu_{\mathcal{M}}(\boldsymbol{x}_{-i}'', x_i) - \mu_{\mathcal{M}}(\boldsymbol{x}'')\right]\right) \\ \cdot \left(\mathbb{E}_{1:n}'''\left[\hat{\mathcal{M}}(\boldsymbol{x}_{-i}''', x_i') - \hat{\mathcal{M}}(\boldsymbol{x}''')\right] - \mathbb{E}_{1:n}'''\left[\mu_{\mathcal{M}}(\boldsymbol{x}_{-i}''', x_i') - \mu_{\mathcal{M}}(\boldsymbol{x}''')\right]\right)\end{array}\right] \\
&= \mathbb{E}_{1:n}''\mathbb{E}_{1:n}'''\left[\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}(\boldsymbol{x}_{-i}'', x_i) - \hat{\mathcal{M}}(\boldsymbol{x}''), \hat{\mathcal{M}}(\boldsymbol{x}_{-i}''', x_i') - \hat{\mathcal{M}}(\boldsymbol{x}''')\right]\right] \\
&= c_{i,i'}(x_i, x_i') - c_i(x_i) - c_i(x_i') + c_0
\end{aligned}
\tag{B.4}
$$

which is exactly Eq. (17). In the last row of Eq. (B.4), only the first two terms need to be further derived to closed-form expressions. The first term is formulated as:

$$
c_{i,i'}(x_i, x_i') = \mathbb{E}_{-i}\mathbb{E}_{-i}'\left[\kappa(\boldsymbol{x}, \boldsymbol{x}')\right] - \boldsymbol{\zeta}_i(x_i)^\top \mathcal{K}^{-1}\boldsymbol{\zeta}_i(x_i')
\tag{B.5}
$$

where $\boldsymbol{\zeta}_i(x_i)$ is formulated in Eq. (B.3), and

$$
\mathbb{E}_{-i}\mathbb{E}_{-i}'\left[\kappa(\boldsymbol{x}, \boldsymbol{x}')\right] = \sigma_0^2 \left|2\Sigma_{-i}^{-1} + I_{n-1}\right|^{-1/2}\exp\left(-\frac{(x_i - x_i')^2}{\sigma_i^2}\right).
\tag{B.6}
$$

The term $c_i(x_i)$ in Eq. (B.4) is formulated as:

$$
c_i(x_i) = \mathbb{E}_{-i}\mathbb{E}_{1:n}'\left[\kappa(\boldsymbol{x}, \boldsymbol{x}')\right] - \boldsymbol{\zeta}_i(x_i)^\top \mathcal{K}^{-1}\mathbb{E}_{1:n}\left[\kappa(\boldsymbol{x}, \mathcal{X})\right]
\tag{B.7}
$$

where the closed-form expression for $\mathbb{E}_{1:n}\left[\kappa(\boldsymbol{x}, \mathcal{X})\right]$ is in Eq. (11), and

$$
\mathbb{E}_{-i}\mathbb{E}_{1:n}'\left[\kappa(\boldsymbol{x}, \boldsymbol{x}')\right] = \sigma_0^2 \left|2\Sigma_{-i}^{-1} + I\right|^{-1/2}\left(\sigma_i^{-2} + 1\right)^{-1/2}\exp\left(-\frac{1}{2(\sigma_i^2 + 1)}x_i^2\right).
\tag{B.8}
$$

Appendix C. Mathematical proofs for total partial variance

We first prove Eq. (19). The variance of $\mathcal{M}_{Ti}(\boldsymbol{x}) = \mathcal{M}(\boldsymbol{x}) - \mathbb{E}_i\left[\mathcal{M}(\boldsymbol{x})\right]$ can be derived as:

$$
\begin{aligned}
\mathbb{V}_{1:n}\left[\mathcal{M}(\boldsymbol{x}) - \mathbb{E}_i\left[\mathcal{M}(\boldsymbol{x})\right]\right] &= \int_{\mathbb{R}^n}\left(\mathcal{M}(\boldsymbol{x}) - \mathcal{M}_0 + \mathcal{M}_0 - \mathbb{E}_i\left[\mathcal{M}(\boldsymbol{x})\right]\right)^2 \phi_n(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\
&= V_y - 2\int_{\mathbb{R}^n}\left(\mathcal{M}(\boldsymbol{x}) - \mathcal{M}_0\right)\left(\mathbb{E}_i\left[\mathcal{M}(\boldsymbol{x})\right] - \mathcal{M}_0\right)\phi_n(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + \mathbb{V}_{-i}\left[\mathbb{E}_i\left[\mathcal{M}(\boldsymbol{x})\right]\right].
\end{aligned}
\tag{C.1}
$$

By integrating x_i out, the n-dimensional integral in the second term of the last line of Eq. (C.1) can be further derived as:

$$\int_{\mathbb{R}^n} \left(\mathcal{M}\left(\boldsymbol{x}\right) - \mathcal{M}_0\right)\left(\mathbb{E}_i\left[\mathcal{M}\left(\boldsymbol{x}\right)\right] - \mathcal{M}_0\right)\phi_n\left(\boldsymbol{x}\right)\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^{n-1}} \left(\mathbb{E}_i\left[\mathcal{M}\left(\boldsymbol{x}\right)\right] - \mathcal{M}_0\right)^2 \phi_{n-1}\left(\boldsymbol{x}_{-i}\right)\mathrm{d}\boldsymbol{x}_{-i}$$
$$= \mathbb{V}_{-i}\left[\mathbb{E}_i\left[\mathcal{M}\left(\boldsymbol{x}\right)\right]\right] \tag{C.2}$$

by substituting which into Eq. (C.1) yields Eq. (19).

Next we derive the closed-form expression for the posterior mean of $\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right)$. Obviously

$$\mu_{\mathcal{M}Ti}\left(\boldsymbol{x}\right) = \mu_{\mathcal{M}}\left(\boldsymbol{x}\right) - \mathbb{E}_i\left[\mu_{\mathcal{M}}\left(\boldsymbol{x}\right)\right] = \mu_{\mathcal{M}}\left(\boldsymbol{x}\right) - b_{-i}\left(\boldsymbol{x}_{-i}\right) - \zeta_{-i}\left(\boldsymbol{x}_{-i}\right)^\top \mathcal{K}^{-1}\left(\mathcal{Y} - b\left(\mathcal{X}\right)\right) \tag{C.3}$$

where $b_{-i}\left(\boldsymbol{x}_{-i}\right) = \mathbb{E}_i\left[b\left(\boldsymbol{x}\right)\right]$ is easy to derive, and we skip the details. The term $\zeta_{-i}\left(\boldsymbol{x}_{-i}\right)$ in Eq. (C.3) can be expressed in closed form as:

$$\zeta_{-i}\left(\boldsymbol{x}_{-i}\right) = \mathbb{E}_i\left[\kappa\left(\boldsymbol{x}, \mathcal{X}\right)\right]$$
$$= \sigma_0^2\left(\sigma_i^{-2} + 1\right)^{-1/2}\exp\left(-\frac{1}{2\left(\sigma_i^2 + 1\right)}\mathcal{X}_{,i}^2 - \frac{1}{2}\left(\boldsymbol{x}_{-i} - \mathcal{X}_{,-i}\right)\Sigma_{-i}^{-1}\left(\boldsymbol{x}_{-i} - \mathcal{X}_{,-i}\right)^\top\right). \tag{C.4}$$

The posterior covariance of $\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right)$ is formulated as:

$$\mathrm{cov}_{\mathcal{M}Ti}\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}'\right)\right]$$
$$= \mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}\left(\boldsymbol{x}'\right)\right] - \underbrace{\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \mathbb{E}_i'\left[\hat{\mathcal{M}}\left(\boldsymbol{x}'\right)\right]\right]}_{c_{-i}\left(\boldsymbol{x}, \boldsymbol{x}'_{-i}\right)}$$
$$- \underbrace{\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}'\right), \mathbb{E}_i\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right]\right]}_{c_{-i}\left(\boldsymbol{x}', \boldsymbol{x}_{-i}\right)} + \underbrace{\mathrm{cov}_{\mathcal{D}}\left[\mathbb{E}_i\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right], \mathbb{E}_i'\left[\hat{\mathcal{M}}\left(\boldsymbol{x}'\right)\right]\right]}_{c_{-ii}\left(\boldsymbol{x}_{-i}, \boldsymbol{x}'_{-i}\right)} \tag{C.5}$$

where the first term is exactly the posterior covariance $\mathrm{cov}_{\mathcal{M}}\left(\boldsymbol{x}\right)$ of the GPR model $\hat{\mathcal{M}}\left(\boldsymbol{x}\right)$. To achieve the closed-form expression for $\mathrm{cov}_{\mathcal{M}Ti}\left(\boldsymbol{x}\right)$, we need to derive the closed-form expressions for the two posterior covariance terms $c_{-i}\left(\boldsymbol{x}, \boldsymbol{x}'_{-i}\right)$ and $c_{-ii}\left(\boldsymbol{x}_{-i}, \boldsymbol{x}'_{-i}\right)$. The term $c_{-i}\left(\boldsymbol{x}, \boldsymbol{x}'_{-i}\right)$ is derived as:

$$c_{-i}\left(\boldsymbol{x}, \boldsymbol{x}'_{-i}\right) = \mathbb{E}_i'\left[\mathrm{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}\left(\boldsymbol{x}'\right)\right]\right] = \mathbb{E}_i'\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] - \kappa\left(\boldsymbol{x}, \mathcal{X}\right)^\top \mathcal{K}^{-1}\zeta_{-i}\left(\boldsymbol{x}'_{-i}\right) \tag{C.6}$$

where $\zeta_{-i}\left(\boldsymbol{x}'_{-i}\right)$ is given in Eq. (C.4), and $\mathbb{E}_i'\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right]$ is formulated as:

$$\mathbb{E}_i'\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] = \sigma_0^2\left(\sigma_i^{-2} + 1\right)^{-1/2}\exp\left(-\frac{1}{2\left(\sigma_i^2 + 1\right)}x_i^2 - \frac{1}{2}\left(\boldsymbol{x}_{-i} - \boldsymbol{x}'_{-i}\right)\Sigma_{-i}^{-1}\left(\boldsymbol{x}_{-i} - \boldsymbol{x}'_{-i}\right)^\top\right). \tag{C.7}$$

The term $c_{-ii}\left(\boldsymbol{x}_{-i}, \boldsymbol{x}'_{-i}\right)$ is derived as:

$$c_{-ii}\left(\boldsymbol{x}_{-i}, \boldsymbol{x}'_{-i}\right) = \mathbb{E}_i \mathbb{E}'_i\left[\text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}\left(\boldsymbol{x}'\right)\right]\right] = \mathbb{E}_i \mathbb{E}'_i\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] - \zeta_{-i}\left(\boldsymbol{x}_{-i}\right)^\top \mathcal{K}^{-1} \zeta_{-i}\left(\boldsymbol{x}'_{-i}\right) \quad \text{(C.8)}$$

where

$$\mathbb{E}_i \mathbb{E}'_i\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] = \sigma_0^2 \left(2\sigma_i^{-2} + 1\right)^{-1/2} \exp\left(-\frac{1}{2}\left(\boldsymbol{x}_{-i} - \boldsymbol{x}'_{-i}\right)\Sigma_{-i}^{-1}\left(\boldsymbol{x}_{-i} - \boldsymbol{x}'_{-i}\right)^\top\right). \quad \text{(C.9)}$$

Although the closed-form expression of the posterior variance of $\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right)$ can be easily derived from that of $cov_{\mathcal{M}Ti}\left(\boldsymbol{x}, \boldsymbol{x}'\right)$, we still give the details here for ease of understanding. From Eq. (C.5), the posterior variance of $\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right)$ degrades into:

$$\begin{aligned} \sigma_{\mathcal{M}Ti}^2\left(\boldsymbol{x}\right) &= \text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}_{Ti}\left(\boldsymbol{x}'\right)\right] \\ &= \sigma_{\mathcal{M}}^2\left(\boldsymbol{x}\right) - 2\text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \mathbb{E}_i\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right]\right] + \mathbb{V}_{\mathcal{D}}\left[\mathbb{E}_i\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right]\right] \end{aligned} \quad \text{(C.10)}$$

where

$$\begin{aligned} \text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \mathbb{E}_i\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right]\right] &= \mathbb{E}'_i\left[\text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}\left(\boldsymbol{x}_{-i}, x'_i\right)\right]\right] \\ &= \sigma_0^2 \left(\sigma_i^{-2} + 1\right)^{-1/2} \exp\left(-\frac{1}{2\left(\sigma_i^2 + 1\right)} x_i^2\right) - \kappa\left(\boldsymbol{x}, \mathcal{X}\right)^\top \mathcal{K}^{-1} \zeta_{-i}\left(\boldsymbol{x}_{-i}\right) \end{aligned} \quad \text{(C.11)}$$

and

$$\begin{aligned} \mathbb{V}_{\mathcal{D}}\left[\mathbb{E}_i\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right)\right]\right] &= \mathbb{E}_i \mathbb{E}'_i\left[\text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}\left(\boldsymbol{x}_{-i}, x'_i\right)\right]\right] \\ &= \sigma_0^2 \left(2\sigma_i^{-2} + 1\right)^{-1/2} - \zeta_{-i}\left(\boldsymbol{x}_{-i}\right)^\top \mathcal{K}^{-1} \zeta_{-i}\left(\boldsymbol{x}_{-i}\right). \end{aligned} \quad \text{(C.12)}$$

Appendix D. Mathematical proofs for total variance

The underlying GPR model for infering the total variance is $\hat{\mathcal{M}}_y\left(\boldsymbol{x}\right) = \hat{\mathcal{M}}\left(\boldsymbol{x}\right) - \hat{\mathcal{M}}_0$, thus the posterior mean is

$$\mu_{\mathcal{M}y}\left(\boldsymbol{x}\right) = \mu_{\mathcal{M}}\left(\boldsymbol{x}\right) - \mu_{\mathcal{M}0} \quad \text{(D.1)}$$

and the posterior covariance is:

$$cov_y\left(\boldsymbol{x}, \boldsymbol{x}'\right) = cov_{\mathcal{M}}\left(\boldsymbol{x}, \boldsymbol{x}'\right) - \text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}_0\right] - \text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}'\right), \hat{\mathcal{M}}_0\right] + \sigma_{\mathcal{M}0}^2 \quad \text{(D.2)}$$

where

$$\begin{aligned} \text{cov}_{\mathcal{D}}\left[\hat{\mathcal{M}}\left(\boldsymbol{x}\right), \hat{\mathcal{M}}_0\right] &= \mathbb{E}'_{1:n}\left[cov_{\mathcal{M}}\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] \\ &= \mathbb{E}'_{1:n}\left[\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right] - \kappa\left(\boldsymbol{x}, \mathcal{X}\right)^\top \mathcal{K}^{-1} \mathbb{E}'_{1:n}\left[\kappa\left(\boldsymbol{x}', \mathcal{X}\right)\right]. \end{aligned} \quad \text{(D.3)}$$

References

[1] H. G. Matthies, Quantifying uncertainty: modern computational representation of probability and applications, in: Extreme Man-made and Natural Hazards in Dynamics of Structures, Springer, 2007, pp. 105–135.

[2] M. Faes, D. Moens, Recent trends in the modeling and quantification of non-probabilistic uncertainty, Archives of Computational Methods in Engineering (2019) 1–39.

[3] B. Iooss, P. Lemaître, A review on global sensitivity analysis methods, in: Uncertainty Management in Simulation-optimization of Complex Systems, Springer, 2015, pp. 101–122.

[4] P. Wei, Z. Lu, J. Song, Variable importance analysis: a comprehensive review, Reliability Engineering & System Safety 142 (2015) 399–432.

[5] E. Borgonovo, E. Plischke, Sensitivity analysis: a review of recent advances, European Journal of Operational Research 248 (3) (2016) 869–887.

[6] W. Li, S. Chen, Z. Jiang, D. W. Apley, Z. Lu, W. Chen, Integrating bayesian calibration, bias correction, and machine learning for the 2014 SANDIA verification and validation challenge problem, Journal of Verification, Validation and Uncertainty Quantification 1 (1) (2016).

[7] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global sensitivity analysis: the primer, John Wiley & Sons, 2008.

[8] V. Chabridon, M. Balesdent, J.-M. Bourinet, J. Morio, N. Gayton, Reliability-based sensitivity estimators of rare event probability in the presence of distribution parameter uncertainty, Reliability Engineering & System Safety 178 (2018) 164–178.

[9] M. A. Valdebenito, H. A. Jensen, H. Hernández, L. Mehrez, Sensitivity estimation of failure probability applying line sampling, Reliability Engineering & System Safety 171 (2018) 99–111.

[10] P. Wei, Z. Lu, J. Song, Regional and parametric sensitivity analysis of Sobol' indices, Reliability Engineering & System Safety 137 (2015) 87–100.

[11] M. D. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (2) (1991) 161–174.

[12] F. Campolongo, J. Cariboni, A. Saltelli, An effective screening design for sensitivity analysis of large models, Environmental Modelling & Software 22 (10) (2007) 1509–1518.

[13] A. Saltelli, J. Marivoet, Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques, Reliability Engineering & System Safety 28 (2) (1990) 229–253.

[14] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, Reliability Engineering & System Safety 52 (1) (1996) 1–17.

[15] G. Archer, A. Saltelli, I. Sobol, Sensitivity measures, ANOVA-like techniques and the use of bootstrap, Journal of Statistical Computation and Simulation 58 (2) (1997) 99–120.

[16] G. Li, J. Hu, S.-W. Wang, P. G. Georgopoulos, J. Schoendorf, H. Rabitz, Random sampling-high dimensional model representation (rs-hdmr) and orthogonality of its different order component functions, The Journal of Physical Chemistry A 110 (7) (2006) 2474–2485.

[17] P. Wei, Z. Lu, J. Song, A new variance-based global sensitivity analysis technique, Computer Physics Communications 184 (11) (2013) 2540–2551.

[18] E. Borgonovo, A new uncertainty importance measure, Reliability Engineering & System Safety 92 (6) (2007) 771–784.

[19] P. Wei, Z. Lu, J. Song, Moment-independent sensitivity analysis using copula, Risk Analysis 34 (2) (2014) 210–222.

[20] E. Plischke, E. Borgonovo, C. L. Smith, Global sensitivity measures from given data, European Journal of Operational Research 226 (3) (2013) 536–550.

[21] I. Sobol, S. Kucherenko, Derivative based global sensitivity measures, Procedia-Social and Behavioral Sciences 2 (6) (2010) 7745–7746.

[22] S. Kucherenko, et al., Derivative based global sensitivity measures and their link with global sensitivity indices, Mathematics and Computers in Simulation 79 (10) (2009) 3009–3017.

[23] M. Lamboni, B. Iooss, A.-L. Popelin, F. Gamboa, Derivative-based global sensitivity measures: general links with sobol' indices and numerical tests, Mathematics and Computers in Simulation 87 (2013) 45–54.

[24] P. Wei, Z. Lu, W. Hao, J. Feng, B. Wang, Efficient sampling methods for global reliability sensitivity analysis, Computer Physics Communications 183 (8) (2012) 1728–1743.

[25] L. Li, I. Papaioannou, D. Straub, Global reliability sensitivity estimation based on failure samples, Structural Safety 81 (2019) 101871.

[26] Z. Kala, Global sensitivity analysis of reliability of structural bridge system, Engineering Structures 194 (2019) 36–45.

[27] W. Hao, Z. Lu, L. Li, A new interpretation and validation of variance based importance measures for models with correlated inputs, Computer Physics Communications 184 (5) (2013) 1401–1413.

[28] T. A. Mara, S. Tarantola, Variance-based sensitivity indices for models with dependent inputs, Reliability Engineering & System Safety 107 (2012) 115–121.

[29] C. Prieur, S. Tarantola, Variance-based sensitivity analysis: Theory and estimation algorithms, In R. Ghanem, D. Higdon, and H. Owhadi, editors, Handbook of Uncertainty Quantification, Springer International Publishing, Cham (2017) 1217–1239.

[30] R. Cukier, C. Fortuin, K. E. Shuler, A. Petschek, J. Schaibly, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory, The Journal of Chemical Physics 59 (8) (1973) 3873–3878.

[31] A. Saltelli, R. Bolado, An alternative way to compute fourier amplitude sensitivity test (FAST), Computational Statistics & Data Analysis 26 (4) (1998) 445–460.

[32] C. Xu, G. Gertner, Understanding and comparisons of different sampling approaches for the fourier amplitudes sensitivity test (FAST), Computational Statistics & Data Analysis 55 (1) (2011) 184–198.

[33] S. Tarantola, D. Gatelli, T. A. Mara, Random balance designs for the estimation of first order global sensitivity indices, Reliability Engineering & System Safety 91 (6) (2006) 717–727.

[34] T. A. Mara, Extension of the RBD-FAST method to the computation of global sensitivity indices, Reliability Engineering & System Safety 94 (8) (2009) 1274–1281.

[35] J. C. Helton, F. Davis, J. D. Johnson, A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling, Reliability Engineering & System Safety 89 (3) (2005) 305–330.

[36] I. M. Sobol', D. Asotsky, A. Kreinin, S. Kucherenko, Construction and comparison of high-dimensional Sobol' generators, Wilmott 2011 (56) (2011) 64–79.

[37] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Mathematics and Computers in Simulation 55 (1-3) (2001) 271–280.

[38] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, Computer Physics Communications 145 (2) (2002) 280–297.

[39] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index, Computer physics communications 181 (2) (2010) 259–270.

[40] A. B. Owen, Better estimation of small Sobol' sensitivity indices, ACM Transactions on Modeling and Computer Simulation (TOMACS) 23 (2) (2013) 1–17.

[41] M. Ratto, A. Pagano, P. Young, State dependent parameter metamodelling and sensitivity analysis, Computer Physics Communications 177 (11) (2007) 863–876.

[42] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering & System Safety 93 (7) (2008) 964–979.

[43] M. Steiner, J.-M. Bourinet, T. Lahmer, An adaptive sampling method for global sensitivity analysis based on least-squares support vector regression, Reliability Engineering & System Safety 183 (2019) 323–340.

[44] W. Chen, R. Jin, A. Sudjianto, Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty, Journal of Mechanical Design 127 (5) (2005) 875–446.

[45] A. Marrel, B. Iooss, B. Laurent, O. Roustant, Calculations of Sobol' indices for the Gaussian process metamodel, Reliability Engineering & System Safety 94 (3) (2009) 742–751.

[46] J. E. Oakley, A. O'Hagan, Probabilistic sensitivity analysis of complex models: a Bayesian approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66 (3) (2004) 751–769.

[47] L. L. Gratiet, C. Cannamela, B. Iooss, A bayesian approach for global sensitivity analysis of (multifidelity) computer codes, SIAM/ASA Journal on Uncertainty Quantification 2 (1) (2014) 336–363.

[48] J. Cockayne, C. J. Oates, T. J. Sullivan, M. Girolami, Bayesian probabilistic numerical methods, SIAM Review 61 (4) (2019) 756–789.

[49] P. Hennig, C. J. Schuler, Entropy search for information-efficient global optimization, The Journal of Machine Learning Research 13 (1) (2012) 1809–1837.

[50] 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.

[51] P. Wei, X. Zhang, M. Beer, Adaptive experiment design for probabilistic integration, Computer Methods in Applied Mechanics and Engineering 365 (2020) 113035.

[52] O. A. Chkrebtii, D. A. Campbell, B. Calderhead, M. A. Girolami, et al., Bayesian solution uncertainty quantification for differential equations, Bayesian Analysis 11 (4) (2016) 1239–1267.

[53] Y. Zhou, Z. Lu, K. Cheng, W. Yun, A Bayesian Monte Carlo-based method for efficient computation of global sensitivity indices, Mechanical Systems and Signal Processing 117 (2019) 498–516.

[54] R. Lebrun, A. Dutfoy, Do Rosenblatt and Nataf isoprobabilistic transformations really differ?, Probabilistic Engineering Mechanics 24 (4) (2009) 577–584.

[55] C. E. Rasmussen, C. Williams, Gaussian processes for machine learning, vol. 1, MIT press 39 (2006) 40–43.

[56] B. Schölkopf, A. J. Smola, F. Bach, et al., Learning with kernels: support vector machines, regularization, optimization, and beyond, MIT press, 2002.

[57] C. E. Rasmussen, Z. Ghahramani, Bayesian Monte Carlo, Advances in neural information processing systems (2003) 505–512.

[58] P. Wei, F. Liu, M. Valdebenito, M. Beer, Bayesian probabilistic propagation of imprecise probabilities with large epistemic uncertainty, Mechanical Systems and Signal Processing 149 (2021) 107219.

[59] T. Gerstner, M. Griebel, Numerical integration using sparse grids, Numerical algorithms 18 (3-4) (1998) 209.

[60] S. A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, in: Doklady Akademii Nauk, Vol. 148, Russian Academy of Sciences, 1963, pp. 1042–1045.

[61] L. L. Gratiet, S. Marelli, B. Sudret, Metamodel-based sensitivity analysis: polynomial chaos expansions and gaussian processes, In R. Ghanem, D. Higdon, and H. Owhadi, editors, Handbook of Uncertainty Quantification, Springer International Publishing, Cham (2017) 1289–1325.

[62] Y. Shi, et al., Particle swarm optimization: developments, applications and resources, in: Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546), Vol. 1, IEEE, 2001, pp. 81–86.

[63] J. N. Fuhg, A. lie Fau, U. Nackenhorst, State-of-the-art and comparative review of adaptive sampling methods for kriging, Archives of Computational Methods in Engineering (2020) 1–59.

[64] T. Abrahamsson, FEMcali, https://www.mathworks.com/matlabcentral/fileexchange/44317-femcali, MATLAB Central File Exchange (Retrieved August 25, 2020).

[65] T. Abrahamsson, D. Kammer, Finite element model calibration using frequency responses with damping equalization, Mechanical systems and signal processing 62 (2015) 218–234.

[66] I. Bilionis, R. Tripathy, M. Gonzalez, Gaussian processes with built-in dimensionality reduction: Applications in high-dimensional uncertainty propagation, arXiv preprint arXiv:1602.04550 (2016).