

## Article

# ANOVA-GP Modeling for High-Dimensional Bayesian Inverse Problems

Xiaoyu Shi <sup>1,†</sup>, Hanyu Zhang <sup>1,†</sup> and Guanjie Wang <sup>2,\*</sup>

<sup>1</sup> School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China; shixy2022@shanghaitech.edu.cn (X.S.); zhanghy5@shanghaitech.edu.cn (H.Z.)

<sup>2</sup> School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

\* Correspondence: guanjie@lixin.edu.cn

† These authors contributed equally to this work.

**Abstract:** Markov chain Monte Carlo (MCMC) stands out as an effective method for tackling Bayesian inverse problems. However, when dealing with computationally expensive forward models and high-dimensional parameter spaces, the challenge of repeated sampling becomes pronounced. A common strategy to address this challenge is to construct an inexpensive surrogate of the forward model, which cuts the computational cost of individual samples. While the Gaussian process (GP) is widely used as a surrogate modeling strategy, its applicability can be limited when dealing with high-dimensional input or output spaces. This paper presents a novel approach that combines the analysis of variance (ANOVA) decomposition method with Gaussian process regression to handle high-dimensional Bayesian inverse problems. Initially, the ANOVA method is employed to reduce the dimension of the parameter space, which decomposes the original high-dimensional problem into several low-dimensional sub-problems. Subsequently, principal component analysis (PCA) is utilized to reduce the dimension of the output space on each sub-problem. Finally, a Gaussian process model with a low-dimensional input and output is constructed for each sub-problem. In addition to this methodology, an adaptive ANOVA-GP-MCMC algorithm is proposed, which further enhances the adaptability and efficiency of the method in the Bayesian inversion setting. The accuracy and computational efficiency of the proposed approach are validated through numerical experiments. This innovative integration of ANOVA and Gaussian processes provides a promising solution to address challenges associated with high-dimensional parameter spaces and computationally expensive forward models in Bayesian inference.

**Keywords:** Bayesian inverse problem; uncertainty quantification; ANOVA decomposition; principle component analysis; Gaussian process regression

**MSC:** 62F15; 65C20; 65D15



**Citation:** Shi, X.; Zhang, H.; Wang, G. ANOVA-GP Modeling for High-Dimensional Bayesian Inverse Problems. *Mathematics* **2024**, *12*, 301. <https://doi.org/10.3390/math12020301>

Academic Editor: Manuel Alberto M. Ferreira

Received: 17 December 2023

Revised: 10 January 2024

Accepted: 14 January 2024

Published: 17 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In numerous scientific and engineering domains, the inference of parameters of interest through observations, such as inverse problems [1], is a common requirement. For instance, in seismology, inverse problems arise in the attempt to determine the subsurface properties of the Earth, such as seismic velocity or material composition, based on recorded seismic waves from earthquakes or controlled sources [2]. Another example is heat conduction, which involves determining the initial temperature distribution or the thermal properties of a material from given temperature measurements at certain locations and times [3,4]. Recently, a statistical model for system reliability evaluation by jointly considering the correlated component lifetimes and the lifetime ordering constraints has been proposed in [5]. Based on the proposed model, the estimation of model parameters from the observed data is also discussed. In [6], two semiparametric additive mean models are proposed

for clustered panel count data, and some estimation equations are derived to estimate the regression parameters of interest for the proposed two models. In this study, we focus on inverse problems.

Classical inference methods typically yield a single-point estimate of the parameters by optimizing an objective function with a regularization term. However, Bayesian inference methods offer the advantage of not only providing parameter estimates but also quantifying the uncertainty associated with the obtained results. This characteristic proves more practical for addressing inverse problems [7]. Bayesian methods treat parameters as random variables. They establish prior distributions for these variables based on existing knowledge and then calculate the conditional probability, also known as the posterior distribution of the parameters, given the observations using Bayes' theorem. Consequently, statistical information such as the expectation and variance of the parameters can be derived.

While the concept of Bayesian inference methods is straightforward, practical implementation can encounter significant challenges. In most instances, the posterior probability distribution lacks an explicit expression and can only be approximated through numerical methods. Typically, the Markov chain Monte Carlo (MCMC) method [8] is employed to generate samples from the posterior distribution, which enables the estimation of posterior statistical information based on these samples. It is crucial to note that the MCMC method necessitates repeated evaluations of the forward model: entailing the complete simulation of the mapping from parameters to observations. However, the forward problem often involves the solution of computationally intensive models such as partial differential equations, especially for practical problems like groundwater inverse modeling [9] and seismic inversion [10]. Furthermore, achieving a robust posterior estimate typically requires a substantial number of samples. These two factors render MCMC simulations impractical for large-scale realistic problems.

To enhance the computational efficiency of MCMC simulations, two primary approaches can be pursued. The first involves reducing the number of required samples, which necessitates the development of more efficient sampling methods. The second aims to minimize the computational cost of a single sample. This can be achieved by pre-building a low-cost surrogate or simplified model for use in the MCMC process, which aligns with the focus of our study.

Existing research methods include polynomial chaos expansion [11], the Gaussian process [12,13], sparse grid interpolation [14], and reduced-basis models [15,16]. The performance of these methods for accelerating Bayesian computation is discussed in [17]. While successful for various inverse problems, their applicability is often constrained by the high dimensionality of unknown parameters in practical scenarios. For instance, hydraulic conductivity in groundwater models and wave speed in seismic inversion, which are often characterized by random fields, can result in dimensions ranging from tens of thousands to even more. Although dimension-reduction methods like truncated Karhunen–Loève (KL) expansion can be employed [18], practical parameters are typically nonsmooth and thus require a substantial number of KL modes. Directly building surrogates for such high-dimensional problems poses a significant challenge. The analysis of variance (ANOVA) methods [19,20], which are proposed for efficiently solving high-dimensional problems, aim to decompose a high-dimensional parameter space into a union of low-dimensional spaces such that standard surrogate modeling strategies can be applied. The adaptive reduced basis ANOVA method combines ANOVA decomposition with reduced basis methods [21,22] to significantly reduce the computational cost of Bayesian inverse problems without sacrificing accuracy. The combination of ANOVA decomposition with the Gaussian process (GP) is studied in various fields, such as designing the surrogate model for the forward problem [13] and recovering functional dependence from sparse data [23,24]. In this study, we primarily concentrate on ANOVA-GP modeling for high-dimensional Bayesian inverse problems.

In this study, we introduce a novel approach that integrates ANOVA decomposition with Gaussian processes to construct surrogate models. The resulting surrogate model

is seamlessly integrated into MCMC simulations, which provides a substantial acceleration in computation. This innovative approach addresses the challenges posed by high-dimensional input and output spaces and holds promise for efficiently handling complex problems in Bayesian inference. To further enhance the efficiency, we propose an adaptive ANOVA-GP-MCMC algorithm that is designed to dynamically update the ANOVA-GP model based on posterior samples obtained during the MCMC process. Numerical results demonstrate the efficiency and effectiveness of the algorithm for quantifying uncertainties of the parameters.

The remainder of the paper is structured as follows. Section 2 elaborates the problem formulation considered in this study, which encompasses Bayesian inverse problems and partial differential equations with random parameters. In Section 3, we delve into the methodology for constructing the ANOVA-GP model and introduce the adaptive ANOVA-GP-MCMC algorithm. Section 4 is dedicated to validating the accuracy and computational efficiency of the proposed approach for high-dimensional inverse problems through a series of numerical experiments. Finally, Section 5 encapsulates a summary of our research and outlines prospects for future work.

## 2. Problem Setup

### 2.1. Bayesian Inverse Problem

Let  $\mathcal{X}$  and  $\mathcal{Y}$  denote two separable Banach spaces and consider a forward model  $G$  mapping from  $\mathcal{X}$  to  $\mathcal{Y}$ . Assume that for an unknown parameter vector  $\xi \in \mathcal{X}$ , the only available observation is  $d = \epsilon(G(\xi))$ , where  $\epsilon : \mathcal{Y} \rightarrow \mathcal{Y}$  introduces measurement noise. The inverse problem involves inferring the parameter  $\xi$  from the observed data  $d$ .

In classical statistical inference, parameters are traditionally treated as deterministic values, and a common approach involves obtaining a point estimate by maximizing the likelihood function. However, in a Bayesian framework, parameters are treated as random variables and are assigned a prior distribution to encapsulate prior knowledge. Upon making an observation, belief in the parameter is updated using Bayes' rule: leading to the derivation of the posterior distribution. This Bayesian approach not only yields a single estimate but also provides a quantification of uncertainty associated with the parameter.

Let  $\pi(\xi) : \mathcal{X} \rightarrow \mathbb{R}_0^+$  denote the prior probability density function of the parameter and satisfy  $\int_{\mathcal{X}} \pi(\xi) d\xi = 1$ . This prior probability reflects our belief in the parameter before any observations are made and can be suitably chosen based on historical data, model structures, and expert knowledge.

Consider  $L(d, \xi) : \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}_0^+$  as the likelihood function, which represents the probability of the observed value  $d$  conditional on the parameter  $\xi$ . According to Bayes' rule, the posterior probability density function of the parameter is expressed as

$$\pi(\xi|d) = \frac{L(d, \xi)\pi(\xi)}{\int_{\mathcal{X}} L(d, \xi)\pi(\xi)d\xi}. \quad (1)$$

The denominator  $\int_{\mathcal{X}} L(d, \xi)\pi(\xi)d\xi$  is referred to as the evidence or normalization constant and equals the marginal probability of the observed data.

The likelihood function is constructed by assessing the disparity between the observed data vector  $d$  and the output of the forward model  $G(\xi)$ . In many practical physical processes, an additive error model is commonly employed; it is represented as  $d = G(\xi) + \epsilon$ , where  $\epsilon$  accounts for measurement errors. Moreover, it is often assumed that  $\epsilon$  is unbiased and uncorrelated across dimensions. Each dimension of  $\epsilon$  follows a normal distribution with zero mean and variance  $\sigma^2$ : denoted as  $\pi_{\epsilon} \sim \mathcal{N}(0, \sigma^2 I)$ . Under these assumptions, the likelihood function can be expressed as

$$L(d, \xi) = \pi_{\epsilon}(d - G(\xi)) = \exp\left\{-\frac{\|d - G(\xi)\|_2^2}{2\sigma^2}\right\}, \quad (2)$$

thus we have

$$\pi(\xi|d) \propto \exp\left\{\frac{\|d - G(\xi)\|_2^2}{2\sigma^2} + \log \pi(\xi)\right\}. \quad (3)$$

While for the aforementioned assumptions the posterior density function  $\pi(\xi|d)$  is known, there are two challenges that hinder the derivation of a straightforward explicit expression. The first challenge arises when the forward model functions as a black box or lacks an attainable explicit expression. The second challenge involves the computation of the normalizing constant, which may require handling intricate high-dimensional integrals that are either computationally demanding or, in some cases, infeasible. Hence, in this study, Markov chain Monte Carlo (MCMC) is employed to generate samples conforming to the posterior distribution. These samples are then used for estimating both the expected value and the uncertainty associated with the parameters.

MCMC was initially introduced by Metropolis [25] and later extended by Hastings [26]. This method constructs a Markov chain with the posterior distribution as the stationary distribution through a random walk. In our investigation, we employ the classic Metropolis–Hastings (MH) algorithm (refer to Algorithm 1) to generate a set of  $N$  samples conforming to the posterior distribution of the parameters. The proposal distribution is denoted as  $\pi_p(\cdot|\xi^{(j)})$  and is commonly chosen as a multivariate normal distribution with mean  $\xi^{(j)}$ .

---

**Algorithm 1** Classic Metropolis–Hastings (MH) algorithm

---

**Input:** Number of samples  $N$ , forward model  $G$ , noise distribution  $\pi_e$ , proposal distribution  $\pi_p$ , prior distribution  $\pi$ , and observation  $d$ .

**Output:** Posterior samples  $\{\xi^{(j)}\}_{j=1}^N$ .

```

1: Draw a sample  $\xi^{(0)}$  from the prior distribution  $\pi$  as the initial state.
2: for  $j := 1 : N$  do
3:   Generate the proposal  $\xi^* \sim \pi_p(\cdot|\xi^{(j-1)})$ .
4:   Compute the acceptance rate  $a := \min\left(1, \frac{\pi_e(d-G(\xi^*))\pi(\xi^*)}{\pi_e(d-G(\xi^{(j-1)}))\pi(\xi^{(j-1)})} \cdot \frac{\pi_p(\xi^{(j-1)}|\xi^*)}{\pi_p(\xi^*|\xi^{(j-1)})}\right)$ .
5:   Draw  $\rho \sim \text{Uniform}[0, 1]$ .
6:   if  $\rho < a$  then
7:     Let  $\xi^{(j)} = \xi^*$ .
8:   else
9:     Let  $\xi^{(j)} = \xi^{(j-1)}$ .
10:  end if
11: end for

```

---

As can be seen from the algorithm, each iteration of MCMC involves an evaluation of the forward model, rendering the MCMC process time-consuming. In the subsequent sections, we will introduce a strategy to expedite MCMC computations by employing surrogate models constructed using the ANOVA-GP method.

## 2.2. Partial Differential Equations with Random Parameters

In this study, we address Bayesian inverse problems for which the forward model is governed by partial differential equations with high-dimensional random parameters [13,21,22]. Specifically, details of the forward model considered in this paper are addressed as follows. Let  $D$  denote a physical domain (a subset of  $\mathbb{R}^2$  or  $\mathbb{R}^3$ ) that is bounded and connected and has a polygonal boundary  $\partial D$ . Let  $\xi = [\xi_1, \dots, \xi_M]^T$  be a random vector of dimension  $M$ , and denote its image by  $I^M$ . The associated problem of the forward model can be formulated as finding a function  $s(x, \xi) : D \times I^M \rightarrow \mathbb{R}$  that satisfies

$$\begin{cases} \mathcal{L}(x, \xi; s(x, \xi)) = f(x), & \forall (x, \xi) \in D \times I^M, \\ \mathfrak{b}(x, \xi; s(x, \xi)) = g(x), & \forall (x, \xi) \in \partial D \times I^M, \end{cases} \quad (4)$$

where  $\mathcal{L}$  is a partial differential operator,  $\mathfrak{b}$  is a boundary operator,  $f$  denotes the source function, and  $g$  determines the boundary conditions of the system. With an observation operator  $\mathfrak{c}$  (e.g., extracting values of the solution  $u$  at specific points in physical space), the forward model is expressed as

$$G(\xi) := u(\xi) = \mathfrak{c}(s(\xi)). \quad (5)$$

During the Markov chain Monte Carlo (MCMC) process, each forward model evaluation involves solving Equation (4). As mentioned earlier, our study will employ a computationally convenient surrogate model to avoid repeatedly solving the original equation, thereby enhancing the computational efficiency of MCMC.

### 3. Methodology

This section initiates with an elaborate description of the standard methodologies that form the foundation for our study. These methodologies include ANOVA decomposition, principal component analysis, and Gaussian process regression. Following this, we introduce the integration of these approaches to construct a surrogate model for the forward model discussed earlier: termed the ANOVA-GP model. Finally, we apply the ANOVA-GP model to MCMC sampling and propose an adaptive ANOVA-GP-MCMC approach. In this approach, the surrogate is dynamically updated based on the posterior knowledge acquired during the inversion process. This adaptive approach further enhances the efficiency and adaptability of the surrogate model in the context of Bayesian inference.

#### 3.1. ANOVA Decomposition

The general process of ANOVA decomposition, as outlined in [13,20,22], can be described as follows. Assume  $u(\xi) \in L^2(I^M)$ , where  $\xi = [\xi_1, \dots, \xi_M]^T \in I^M$ . Let  $\mathcal{P}$  be the collection of coordinate indices  $\{1, \dots, M\}$ , and any non-empty subset  $t \subseteq \mathcal{P}$  is referred to as an ANOVA index, with its cardinality denoted by  $|t|$ . For any given  $t$ ,  $\xi_t$  represents a  $|t|$ -dimensional vector composed of the values of  $\xi$  at corresponding indices. For instance, if  $t = 1, 2, 4$ , then  $|t| = 3$  and  $\xi_t = [\xi_1, \xi_2, \xi_4]^T \in I^3$ . Let  $d\mu$  be a probability measure on  $I^M$ . The ANOVA decomposition of  $u(\xi)$  can be expressed as the sum of  $2^M$  terms:

$$u(\xi) = u_0 + \sum_{t \neq \emptyset, t \subseteq \mathcal{P}} u_t(\xi_t), \quad (6)$$

where

$$u_0 = \int_{I^M} u(\xi) d\mu(\xi), \quad (7)$$

and  $u_t(\xi_t)$  is recursively defined as

$$u_t(\xi_t) = \int_{I^{M-|t|}} u(\xi) d\mu(\xi_{\mathcal{P} \setminus t}) - \sum_{s \neq \emptyset, s \subset t} u_s(\xi_s) - u_0, \quad \forall t \subseteq \mathcal{P}, \quad (8)$$

where  $u_t(\xi_t)$  is referred to as a  $|t|$ -order ANOVA term. Since we assume that  $\{\xi_i\}_{i=1}^M$  are independent of each other,  $d\mu(\xi_{\mathcal{P} \setminus t}) = \prod_{i \in \mathcal{P} \setminus t} d\mu(\xi_i)$ .

For any high-dimensional function, the ANOVA decomposition is a finite and exact expansion [19,27]. Additionally, every term in the expansion is orthogonal to the other terms [20], i.e.,

$$\int_{I^M} u_t(\xi_t) u_w(\xi_w) d\mu(\xi) = \delta_{tw}, \quad \forall t, w \subseteq \mathcal{P}, \quad (9)$$

and thus, every term except  $u_0$  has zero mean:

$$\int_{I^M} u_t(\xi_t) d\mu(\xi) = 0, \quad \forall t \subseteq \mathcal{P} \text{ and } t \neq \emptyset. \quad (10)$$

By (9), the variance of  $u(\xi)$  can be computed via

$$\mathbb{V}(u(\xi)) = \sum_{t \neq \emptyset, t \subset \mathcal{P}} \mathbb{V}(u_t(\xi_t)) \quad (11)$$

### 3.1.1. Anchored ANOVA Decomposition

When using the Lebesgue measure  $d\mu(\xi) := d\xi$ , Equation (6) is referred to as the classic ANOVA decomposition. However, this involves multiple high-dimensional integrals, necessitating complex computations. To alleviate this, one often utilizes the Dirac measure  $d\mu(\xi) := \delta(\xi - c)d\xi = \prod_{i=1}^M \delta(\xi_i - c_i)d\xi_i$ , where  $c = [c_1, \dots, c_M]^T \in I^M$  is a given anchor point. This approach is known as a Dirac ANOVA decomposition or anchored ANOVA decomposition [20,28]. Under this measure, Equations (7) and (8) can be expressed as

$$u_0 = u(c), \quad u_t(\xi_t) = u(\xi^{c,t}) - \sum_{s \neq \emptyset, s \subset t} u_s(\xi_s) - u_0, \quad \forall t \subseteq \mathcal{P}, \quad (12)$$

where  $\xi^{c,t} = [\xi_1^{c,t}, \dots, \xi_M^{c,t}]^T$  is defined as

$$\xi_i^{c,t} = \begin{cases} c_i, & \text{for } i \in \mathcal{P} \setminus t \\ \xi_i, & \text{for } i \in t \end{cases}. \quad (13)$$

This naturally raises a question: How should the anchor point  $c$  be chosen? Theoretically, any anchor point  $c$  can provide an accurate ANOVA decomposition. However, a judicious choice of the anchor point enables approximation of the original function with a small subset of expansion terms [20,28], which is a crucial factor for enhancing computational efficiency. In previous studies, selecting parameter points with function values close to the mean has proven to be effective. However, calculating the expected value of the function is usually challenging and can only be obtained through inexact numerical approximation. Another choice of the anchor point, as demonstrated in [20], is the expectation of the parameter, which is a strategy also adopted in our study.

### 3.1.2. Selection of the ANOVA Terms

In Equation (6), as the parameter dimension  $M$  increases, the number of terms in the ANOVA expansion grows exponentially, posing a significant challenge to computing the ANOVA decomposition for high-dimensional functions. In practice, it is often sufficient to retain only a small subset of the lower-order ANOVA terms to approximate the original function effectively [13,22]. Let  $\mathcal{J}$  be the set of active ANOVA indices; the original function can then be approximated as

$$u(\xi) \approx u_{\mathcal{J}}(\xi) = u_0 + \sum_{t \in \mathcal{J}} u_t(\xi_t), \quad (14)$$

which is also called the truncated ANOVA expansion. In this work, we adopt the method for selecting active ANOVA terms as explored in [22]. Define the set of active ANOVA indices at each order  $i$  as

$$\mathcal{J}_i := \{t \mid t \in \mathcal{J} \text{ and } |t| = i\}; \quad (15)$$

then  $\mathcal{J} = \cup_{i=1}^M \mathcal{J}_i$ . Initialize the set of first-order ANOVA indices as  $\mathcal{J}_1 = \{1, \dots, M\}$ . Assuming the prior distribution  $\pi_t(\xi_t)$  of parameter  $\xi_t$  is known, the prior expectation of  $u_t(\xi_t)$  is

$$\mathbb{E}(u_t) := \int_{I^{|t|}} u_t(\xi_t) \pi_t(\xi_t) d\xi_t; \quad (16)$$

the relative expectation of each ANOVA term is defined as

$$\gamma_t := \frac{\|\mathbb{E}(u_t)\|_2}{\|u_0 + \sum_{s \in \mathcal{J}_1 \cup \dots \cup \mathcal{J}_{|t|-1}} \mathbb{E}(u_s)\|_2}, \quad (17)$$



where  $\|\cdot\|_2$  represents the L2 norm of the function. The prior expectation of the ANOVA term  $\mathbb{E}(u_t)$  can be estimated using the Monte Carlo method:

$$\tilde{\mathbb{E}}(u_t) := \frac{1}{N_{model}} \sum_{j=1}^{N_{model}} u_t(\xi_t^{(j)}), \quad (18)$$

where  $\{\xi_t^{(j)}\}_{j=1}^{N_{model}}$  are  $N_{model}$  samples drawn from the prior distribution  $\pi_t(\xi_t)$ ; thus, the relative expectation can be approximated as

$$\tilde{\gamma}_t := \frac{\|\tilde{\mathbb{E}}(u_t)\|_2}{\|u_0 + \sum_{s \in \mathcal{J}_1 \cup \dots \cup \mathcal{J}_{|t|-1}} \tilde{\mathbb{E}}(u_s)\|_2}. \quad (19)$$

The relative expectation also serves as an indicator of the importance of the ANOVA terms. For each ANOVA term in  $\mathcal{J}_i$ , if its relative expectation exceeds a given threshold  $tol_{anova}$ , it is incorporated into the set  $\tilde{\mathcal{J}}_i$

$$\tilde{\mathcal{J}}_i := \{t \mid t \in \mathcal{J}_i \text{ and } \tilde{\gamma}_t \geq tol_{anova}\}. \quad (20)$$

Then, the set of active ANOVA indices at order  $(i+1)$  can be constructed through

$$\mathcal{J}_{i+1} := \{t \mid |t| = i+1 \text{ and } \forall s \subset t \text{ with } |s| = i \text{ satisfies } s \in \tilde{\mathcal{J}}_i\}. \quad (21)$$

This implies that for each selected ANOVA term in  $\mathcal{J}$ , both the term itself and any of its subsets exhibit relative expectation values greater than the specified threshold. The process automatically concludes when the set of active terms at some order becomes empty. It has been demonstrated that even with a very small threshold, only a small fraction of all expansion terms in  $\mathcal{J}$  are reserved, with most of them being low-order terms [13,28].

By performing ANOVA decomposition on  $\xi$  of  $G(\xi)$  in Equation (5) using the aforementioned approach, one can decompose the original problem into several low-dimensional sub-problems. Gaussian process regression is then employed for local modeling. Subsequently, Equation (6) is utilized to combine solutions from each sub-problem into the full solution. In the following section, we will delve into the details of Gaussian process regression.

### 3.2. Gaussian Process Regression

A Gaussian process is a collection of random variables defined over a continuous domain for which any finite combination of these random variables collectively follows a joint Gaussian distribution. It can be conceptualized as an extension of the multidimensional Gaussian distribution to infinite dimensions [29]. Following the representation in [30], a Gaussian process  $f(x)$  ( $x \in \mathcal{X}$ ) can be specified by a mean function  $m(x)$  and a covariance function  $k(x, x')$ : denoted as  $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$ . Here,  $x$  can be either one-dimensional or multidimensional. Given a series of observation points  $\{x_i\}_{i=1}^n \subset \mathcal{X}$ , according to the definition of a Gaussian process,  $[f(x_1), \dots, f(x_n)]^T$  is expected to adhere to a multidimensional Gaussian distribution:

$$[f(x_1), \dots, f(x_n)]^T \sim \mathcal{N}(m, K), \quad (22)$$

where  $m = [m(x_1), \dots, m(x_n)]^T$  and  $K_{ij} = k(x_i, x_j)$ . Next, we will delve into the details of using Gaussian processes for regression.

Given a set of observed points  $X = [x_1, \dots, x_n]^T$  and their corresponding observed values  $y = [y_1, \dots, y_n]^T$  as well as predicted points  $Z = [z_1, \dots, z_q]^T$ , one can utilize

the properties discussed earlier to derive the joint distribution of observed values  $y$  and predicted values  $f^* = f(Z) = [f(z_1), \dots, f(z_q)]^T$ :

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(Z) \end{bmatrix}, \begin{bmatrix} K(X, X) & K(X, Z) \\ K(X, Z)^T & K(Z, Z) \end{bmatrix}\right), \quad (23)$$

where  $m(X) = [m(x_1), \dots, m(x_n)]^T$ ;  $m(Z) = [m(z_1), \dots, m(z_q)]^T$ ; and the  $ij$ -th elements of  $K(X, Z)$ ,  $K(X, X)$ , and  $K(Z, Z)$  are  $k(x_i, z_j)$ ,  $k(x_i, x_j)$ , and  $k(z_i, z_j)$ , respectively. Thus, the conditional probability distribution of  $f^*$  is expressed as

$$p(f^*|X, y, Z) \sim \mathcal{N}(\hat{m}, \hat{\Sigma}), \quad (24)$$

where the predicted mean  $\hat{m}$  and the predicted covariance matrix  $\hat{\Sigma}$  are

$$\begin{aligned} \hat{m} &= K(X, Z)^T (K(X, X))^{-1} (y - m(X)) + m(Z), \\ \hat{\Sigma} &= K(Z, Z) - K(X, Z)^T (K(X, X))^{-1} K(X, Z). \end{aligned} \quad (25)$$

In practice, data are often standardized to be zero mean. Given that  $m(x) = 0$ , the expressions for the predicted mean and predicted covariance matrix become

$$\begin{aligned} \hat{m} &= K(X, Z)^T (K(X, X))^{-1} y, \\ \hat{\Sigma} &= K(Z, Z) - K(X, Z)^T (K(X, X))^{-1} K(X, Z). \end{aligned} \quad (26)$$

The covariance function, commonly referred to as the kernel of the Gaussian process, is evidently crucial in the outlined process. However, determining the specific form and parameters of the kernel poses a challenge. Currently, there is no unified solution for kernel selection across different problems. Researchers typically rely on the experience of their predecessors to manually choose a relatively suitable form. One of the most frequently used kernels in Gaussian processes is the squared exponential kernel, which is also known as the Gaussian kernel [31]:

$$k_{SE}(x, x') = \rho^2 \exp\left(-\frac{\|x - x'\|^2}{2l^2}\right), \quad (27)$$

where  $l$  represents the correlation length, and  $\rho^2$  is the variance. When extended to high-dimensional situations, where fluctuations in each dimension are not uniform, one can consider an automated relevance determination form for the kernel:

$$k_{SEard}(x, x') = \rho^2 \exp\left(-\frac{(x - x')^T L^{-1} (x - x')}{2}\right), \quad (28)$$

where  $x \in \mathbb{R}^M$ , and  $L = \text{diag}(l_1^2, \dots, l_M^2)$  is a diagonal matrix with the squared correlation length of each dimension on the diagonal.

With the specified kernel form, the hyperparameters still need to be determined through optimization. Let  $\beta = [l_1, \dots, l_M, \rho]^T$ , and our objective is to maximize the marginal probability of observed values and, equivalently, to minimize its negative logarithm [32]:

$$M(\beta) = -\log p(y|\beta) = \frac{1}{2} \log \det(K(\beta)) + \frac{1}{2} y^T K^{-1}(\beta) y + \frac{n}{2} \log(2\pi), \quad (29)$$

where  $K(\beta) = K(X, X; \beta)$  is the covariance matrix for a given parameter  $\beta$ .

As discussed earlier, our study involves Bayesian inversion, necessitating the repeated solution of a partial differential equation with random parameters. To streamline this process, we intend to utilize the finite element method to pre-calculate solutions for a subset of parameters. These parameters and their corresponding finite element solutions



will then serve as training samples in the Gaussian process to optimize kernel parameters and establish the Gaussian process model. This approach enables us to predict results directly when confronted with unknown parameters: eliminating the need to solve the equation anew.

However, classical Gaussian process regression faces challenges when handling multi-dimensional outputs. Modeling each dimension separately would imply assuming that each dimension is unrelated [33]. Hence, directly using the finite element solutions as the predicted output in Gaussian process regression may not be advisable. To address this, principal component analysis becomes necessary. It allows us to project the solution space of the equation into a lower-dimensional subspace, thereby improving the performance of Gaussian process regression.

### 3.3. Principal Component Analysis

This section provides a detailed description of how to employ principal component analysis (PCA) on each sub-problem to reduce the dimensionality of the output of the forward model.

Following the representation in [33], for every ANOVA term  $u_t(\xi_t)$ , assuming that we have  $N_{model}$  training samples:  $\{y_t^{(j)} | y_t^{(j)} = u_t(\xi_t^{(j)}) \in \mathbb{R}^d, j = 1, \dots, N_{model}\}$ . First, standardize these samples to zero mean:

$$\tilde{y}_t^{(j)} = y_t^{(j)} - \mu_t, \quad j = 1, \dots, N_{model}, \quad (30)$$

where

$$\mu_t = \frac{1}{N_{model}} \sum_{j=1}^{N_{model}} y_t^{(j)}. \quad (31)$$

Construct the matrix

$$S := [\tilde{y}_t^{(1)}, \tilde{y}_t^{(2)}, \dots, \tilde{y}_t^{(N_{model})}], \quad (32)$$

and perform singular value decomposition:

$$S = U\Sigma V^T, \quad (33)$$

where  $\Sigma$  is a diagonal matrix composed of the singular values

$$\Sigma = \text{diag}([\sigma_1, \sigma_2, \dots, \sigma_{N_{model}}]), \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{N_{model}}. \quad (34)$$

Given the threshold  $tol_{pca}$ , one can find  $0 \leq R \leq N_{model}$  such that

$$\frac{\sum_{j=1}^R \sigma_j^2}{\sum_{j=1}^{N_{model}} \sigma_j^2} > 1 - tol_{pca} \quad \text{and} \quad \frac{\sum_{j=1}^{R-1} \sigma_j^2}{\sum_{j=1}^{N_{model}} \sigma_j^2} \leq 1 - tol_{pca}. \quad (35)$$

Select the first  $R$  column vectors of the matrix  $U$  as the principal components matrix:

$$V_t := [v_1, \dots, v_R]. \quad (36)$$

Consequently,  $u_t(\xi_t)$  can be approximated as

$$u_t(\xi_t) \approx V_t \tilde{u}_t(\xi_t) + \mu_t, \quad (37)$$

where

$$\begin{aligned} \tilde{u}_t(\xi_t) &= [\tilde{u}_{t,1}(\xi_t), \dots, \tilde{u}_{t,R}(\xi_t)]^T \in \mathbb{R}^R, \\ \tilde{u}_{t,r}(\xi_t) &= v_r^T(u_t(\xi_t) - \mu_t), \quad \text{for } r = 1, \dots, R. \end{aligned} \quad (38)$$

Generally, singular values decay quite rapidly, making the value of  $R$  much smaller than the dimension of the model output. We then construct a Gaussian process model for each PCA mode  $\tilde{u}_{t,r}(\xi_t)$  ( $r = 1, \dots, R$ ) and ultimately obtain  $u_t(\xi_t)$  through Equation (37).

### 3.4. ANOVA-GP Modeling

This section outlines the overall ANOVA-GP modeling approach [33] used to construct a low-complexity surrogate model for the forward problem (5).

Given the input samples  $\Xi := \{\xi_t^{(j)}\}_{j=1}^{N_{model}}$  generated from a specific distribution, the first step is to conduct the truncated ANOVA expansion as described in Section 3.1. This decomposes the problem with high-dimensional inputs into several low-order sub-problems: yielding the active set  $\mathcal{J}$  and training sets  $\{\xi_t^{(j)}, u_t(\xi_t^{(j)})\}_{j=1}^{N_{model}}$  for  $t \in \mathcal{J}$ . The second step is to perform PCA analysis as presented in Section 3.3: projecting the high-dimensional outputs  $u_t(\xi_t^{(j)})$  of each sub-problem into a compact subspace denoted as  $\tilde{u}_t(\xi_t^{(j)}) = [\tilde{u}_{t,1}(\xi_t^{(j)}), \dots, \tilde{u}_{t,R}(\xi_t^{(j)})]^T$ . The final step is to perform Gaussian process regression for each local training set  $\{\xi_t^{(j)}, \tilde{u}_{t,r}(\xi_t^{(j)})\}_{j=1}^{N_{model}}$ , resulting in GP models

$$\hat{u}_{t,r} = \mathcal{GP}(m_{t,r}(\xi_t), k_{t,r}(\xi_t, \xi_t')) \quad (39)$$

for  $t \in \mathcal{J}$  and  $r = 1, \dots, R$ . The squared exponential kernel with automated relevance determination is selected as the covariance function here, and the hyperparameter  $\beta_{t,r}$  is determined through minimizing

$$\begin{aligned} M(\beta_{t,r}) &= -\log p(\alpha_{t,r} | \beta_{t,r}) \\ &= \frac{1}{2} \log \det(K(\beta_{t,r})) + \frac{1}{2} \alpha_{t,r}^T K^{-1}(\beta_{t,r}) \alpha_{t,r} + \frac{N_{model}}{2} \log(2\pi), \end{aligned} \quad (40)$$

where

$$\alpha_{t,r} = [\tilde{u}_{t,r}(\xi_t^{(1)}), \dots, \tilde{u}_{t,r}(\xi_t^{(N_{model})})]^T, \quad K(\beta_{t,r})_{ij} = k_{t,r}(\xi_t^{(i)}, \xi_t^{(j)}). \quad (41)$$

By substituting the predicted mean values of these models into Equation (37), we obtain the predictor of  $u_t(\xi_t)$ :

$$\hat{u}_t(\xi_t) = V_t U_t(\xi_t) + \mu_t, \quad U_t(\xi_t) = [\hat{m}_{t,1}(\xi_t), \dots, \hat{m}_{t,R}(\xi_t)]^T, \quad (42)$$

where

$$\hat{m}_{t,r}(\xi_t) = k_*^T K^{-1}(\beta_{t,r}) \alpha_{t,r}, \quad k_* = [k_{t,r}(\xi_t, \xi_t^{(1)}), \dots, k_{t,r}(\xi_t, \xi_t^{(N_{model})})]^T. \quad (43)$$

Finally, in accordance with Equation (14), we integrate all local predictors of sub-problems to obtain the overall ANOVA-GP model:

$$\hat{u}_{\mathcal{J}}(\xi) = u_0 + \sum_{t \in \mathcal{J}} \hat{u}_t(\xi_t). \quad (44)$$

Refer to Algorithm 2 for a detailed description of the entire process of obtaining the ANOVA-GP model.

**Algorithm 2** Construction of ANOVA-GP model**Input:** Sample set  $\Xi = \{\zeta^{(j)}\}_{j=1}^{N_{model}} \subset I^M$ .**Output:** ANOVA-GP model  $\hat{u}_{\mathcal{J}}$ .

- 1: Compute  $c = \frac{1}{N_{model}} \sum_{j=1}^{N_{model}} \zeta^{(j)}$ .
- 2: Compute the zero-order term  $u_0 = u(c)$ .
- 3: Initialize  $\mathcal{J} = \mathcal{J}_1 = \{1, \dots, M\}$  and  $i = 1$ .
- 4: **while**  $\mathcal{J}_i \neq \emptyset$  **do**
- 5:   Let  $\tilde{\mathcal{J}}_i = \emptyset$ .
- 6:   **for**  $t \in \mathcal{J}_i$  **do**
- 7:     Construct sample set  $\Xi_t = \{\zeta_t^{(j)}, \zeta^{(j)} \in \Xi \text{ for } j = 1, \dots, N_{model}\} \subset I^i$ .
- 8:     Compute  $\tilde{\gamma}_t = \frac{\|\tilde{\mathbb{E}}(u_t)\|}{\|u_0 + \sum_{s \in \mathcal{J}_1 \cup \dots \cup \mathcal{J}_{|t|-1}} \tilde{\mathbb{E}}(u_s)\|}$  through Equations (8) and (18).
- 9:     **if**  $\tilde{\gamma}_t > tol_{anova}$  **then**
- 10:       Update  $\tilde{\mathcal{J}}_i = \tilde{\mathcal{J}}_i \cup \{t\}$ .
- 11:     **end if**
- 12:   **end for**
- 13:   Construct  $\mathcal{J}_{i+1} = \{t \mid |t| = i + 1 \text{ and } \forall s \subset t \text{ with } |s| = i \text{ satisfies } s \in \tilde{\mathcal{J}}_i\}$ .
- 14:   Update  $\mathcal{J} = \mathcal{J} \cup \mathcal{J}_{i+1}$  and  $i = i + 1$ .
- 15: **end while**
- 16: **for**  $t \in \mathcal{J}$  **do**
- 17:   For sample set  $\{u_t(\zeta_t^{(j)})\}_{j=1}^{N_{model}}$ , obtain the mean  $\mu_t$ , principal components matrix  $V_t$ , and PCA coefficients  $\{\tilde{u}_{t,r}(\zeta_t^{(j)})\}_{j=1}^{N_{model}}$  through Equations (30)–(38).
- 18:   **for**  $r = 1 : R$  **do**
- 19:     Train a Gaussian process regression model  $\hat{u}_{t,r} = \mathcal{GP}(m_{t,r}(\zeta_t), k_{t,r}(\zeta_t, \zeta'_t))$  with training set  $\{\zeta_t^{(j)}, \tilde{u}_{t,r}^{(j)}\}_{j=1}^{N_{model}}$ , for which the hyperparameter is determined by minimizing (40).
- 20:   **end for**
- 21:   Construct the local predictor  $\hat{u}_t(\zeta_t)$  through Equations (41)–(43).
- 22: **end for**
- 23: Integrate the overall ANOVA-GP model  $\hat{u}_{\mathcal{J}}(\zeta) = u_0 + \sum_{t \in \mathcal{J}} \hat{u}_t(\zeta_t)$ .

**3.5. Adaptive ANOVA-GP-MCMC**

Section 3.4 introduces a surrogate modeling strategy with the aim of replacing the computationally intensive forward model with fast ANOVA-GP predictions in MCMC iterations. This is done to reduce computational costs and enhance sampling efficiency. Ideally, the sample set  $\Xi$  should be selected from the posterior distribution. However, this is impractical since we lack knowledge of the posterior information of the parameters before conducting MCMC. Thus, prior to obtaining posterior information, an initial model must be built using samples that align with the prior probability distribution. Nevertheless, using a model based solely on priors may be inefficient, especially when the prior distribution significantly differs from the posterior distribution [34]. The impact of employing a sample set that does not conform to the actual distribution on model efficiency is primarily observed in two aspects. Firstly, in the anchored ANOVA decomposition, the selection of the anchor point and active terms involves statistical information about parameters. Research by Gal et al. [15] indicates that appropriate anchor point selection—specifically, choosing the expectation of the parameters (i.e., the posterior mean in the Bayesian inverse problem setting)—enables achieving a given approximation accuracy with very few expansion terms. Additionally, the selection of active terms relies on their relative expectations. Therefore, utilizing a sample set that adheres to the posterior distribution leads to a more efficient ANOVA representation. Secondly, during the Gaussian process regression, if the training sample set deviates significantly from the actual distribution of parameters, it affects the accuracy of prediction. To address this issue, we design an adaptive algorithm that updates our ANOVA-GP model based on the obtained posterior samples during the

MCMC process to enhance computational efficiency. We refer to this approach as the adaptive ANOVA-GP-MCMC algorithm.

This process is detailed in Algorithm 3. Initially, we obtain  $N_{model}$  samples from the prior distribution  $\pi(\xi)$  and construct an initial ANOVA-GP model based on these samples. Then we employ the resulting model as the initial surrogate of the forward model in MCMC.

---

**Algorithm 3** Adaptive ANOVA-GP-MCMC algorithm

---

**Input:** Number of MCMC samples  $N$ , number of samples to construct model  $N_{model}$ , number of samples for updating model  $N_{update}$ , noise distribution  $\pi_\epsilon$ , proposal distribution  $\pi_p$ , prior distribution  $\pi$ , and observation  $d$ .

**Output:** Posterior samples  $\{\xi^{(j)}\}_{j=1}^N$ .

- 1: Generate  $N_{model}$  samples from  $\pi(\xi)$  and construct the prior ANOVA-GP model  $\hat{u}_{\mathcal{J}}$  according to Algorithm 2.
  - 2: Draw a sample  $\xi^{(0)}$  from the prior distribution  $\pi$  as the initial state.
  - 3: Let  $Update\_label = 1$ .
  - 4: **for**  $j = 1 : N$  **do**
  - 5:   Generate the proposal  $\xi^* \sim \pi_p(\cdot | \xi^{(j-1)})$ .
  - 6:   Compute the acceptance rate  $a := \min\left(1, \frac{\pi_\epsilon(d - \hat{u}_{\mathcal{J}}(\xi^*))\pi(\xi^*)}{\pi_\epsilon(d - \hat{u}_{\mathcal{J}}(\xi^{(j-1)}))\pi(\xi^{(j-1)})} \cdot \frac{\pi_p(\xi^{(j-1)} | \xi^*)}{\pi_p(\xi^* | \xi^{(j-1)})}\right)$ .
  - 7:   Draw  $\rho \sim \text{Uniform}[0, 1]$ .
  - 8:   **if**  $\rho < a$  **then**
  - 9:     Let  $\xi^{(j)} = \xi^*$ .
  - 10:   **else**
  - 11:     Let  $\xi^{(j)} = \xi^{(j-1)}$ .
  - 12:   **end if**
  - 13:   **if**  $j \bmod N_{update} = 0$  and  $Update\_label = 1$  **then**
  - 14:     Let  $\mathcal{J}' = \mathcal{J}$ .
  - 15:     Randomly select  $N_{model}$  samples from  $\{\xi^{(k)}\}_{k=j-N_{update}+1}^j$ , reconstruct the ANOVA-GP model  $\hat{u}_{\mathcal{J}'}$ , and update  $\mathcal{J}$ .
  - 16:     **if**  $\mathcal{J} = \mathcal{J}'$  **then**
  - 17:       Let  $Update\_label = 0$ .
  - 18:     **end if**
  - 19:   **end if**
  - 20: **end for**
- 

Suppose we have observation  $d$ . Initialize the Markov chain  $\Xi^* := \{\xi^{(1)}\}$ , where  $\xi^{(1)}$  is obtained from  $\pi(\xi)$ . Subsequently, for  $j \geq 1$ , obtain a candidate sample  $\xi^*$  from the proposal distribution  $\pi_p(\cdot | \xi^{(j)})$  based on the current sample  $\xi^{(j)}$ . Then, adopt the prediction  $\hat{u}_{\mathcal{J}}(\xi)$  of the ANOVA-GP model to approximate the output  $u(\xi)$  of the forward model. Compute the acceptance rate of the candidate sample  $\xi^*$ :

$$a := \min\left(1, \frac{\pi_\epsilon(d - \hat{u}_{\mathcal{J}}(\xi^*))\pi(\xi^*)}{\pi_\epsilon(d - \hat{u}_{\mathcal{J}}(\xi^{(j)}))\pi(\xi^{(j)})} \cdot \frac{\pi_p(\xi^{(j)} | \xi^*)}{\pi_p(\xi^* | \xi^{(j)})}\right), \quad (45)$$

which is the probability that the candidate sample  $\xi^*$  is accepted. If accepted, the state of the Markov chain transits to  $\xi^*$ , i.e.,  $\xi^{(j+1)} = \xi^*$ ; otherwise, it remains the same, i.e.,  $\xi^{(j+1)} = \xi^{(j)}$ . Each time  $N_{update}$  new samples are generated, we randomly select  $N_{model}$  samples from them to reconstruct the ANOVA-GP model. This updating procedure stops when the active set  $\mathcal{J}$  of the newly updated model remains unchanged.

#### 4. Numerical Study

In this section, we consider a Bayesian inverse problem in which the forward model is governed by partial differential equations with random parameters. To demonstrate the effi-

ciency and effectiveness of the ANOVA-GP modeling approach and the proposed adaptive ANOVA-GP-MCMC algorithm, both the forward and the inverse problems are investigated. In this section, the finite element solutions are obtained by IFISS [35,36]. All the results presented here are obtained on a desktop with a 2.20 GHz Intel Xeon E52630 v4 CPU.

#### 4.1. Problem Setup

Consider a parametric two-dimensional elliptic problem on a unit square  $D = (0, 1) \times (0, 1)$ :

$$\nabla \cdot (-\exp(a(x, \xi)) \nabla s(x, \xi)) = f(x), \quad x \in D, \quad \xi \in \mathbb{R}^M \quad (46)$$

with boundary conditions

$$\begin{cases} \exp(a(x, \xi)) \nabla s(x, \xi) \cdot \vec{n} = 0, & x \in (0, 1) \times \{0, 1\} \\ s(x, \xi) = 0, & x \in \{0, 1\} \times [0, 1], \end{cases} \quad (47)$$

where  $\vec{n}$  denotes the unit normal vector. In (47), the notation  $(0, 1)$  represents the open interval of real numbers from 0 to 1, while  $[0, 1]$  denotes the closed interval. The set  $\{0, 1\}$  consists of the real numbers 0 and 1. This equation characterizes steady flows in porous media, where  $a(x, \xi)$  is an unknown permeability field,  $s(x, \xi)$  represents the pressure head, and  $f(x)$  serves as the source term, which we set to

$$f(x) = 3 \exp\left(-(x_1 - 0.5)^2 - (x_2 - 0.5)^2\right). \quad (48)$$

The permeability field  $a(x, \xi)$  is parameterized through a truncated Karhunen–Loève (KL) expansion [18] of a Gaussian random field featuring a constant mean function  $a_0(x)$  and an exponential kernel covariance function.

$$C(x, x') = \sigma_f^2 \exp\left(-\frac{\|x - x'\|_2^2}{\tau}\right), \quad (49)$$

where  $\sigma_f$  represents the standard deviation and  $\tau$  is the correlation length. Denoting the eigenpairs of the covariance function (49) as  $\{\lambda_k, a_k(x)\}_{k=1}^\infty$ , the truncated Karhunen–Loève expansion can be expressed as

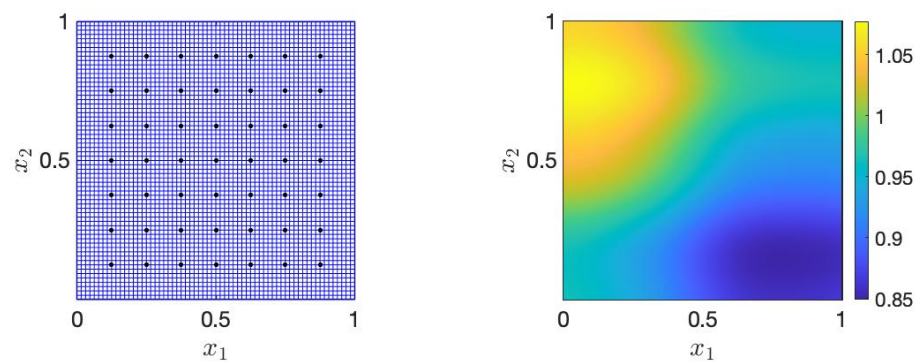
$$a(x, \xi) = a_0(x) + \sum_{k=1}^M \sqrt{\lambda_k} a_k(x) \xi_k, \quad (50)$$

where  $\{\xi_k\}_{k=1}^M$  are independent and identically distributed (i.i.d.) random variables following a standard normal distribution  $\mathcal{N}(0, 1)$ . In this numerical example, we set  $a_0(x) = 1$ ,  $\sigma_f^2 = 0.5$ , and  $\tau = 2$ . To capture 95% of the total variance of the random field, the KL expansion is truncated at  $M = 12$ . The prior distribution of the parameters  $\{\xi_k\}_{k=1}^M$  is specified as  $\mathcal{N}(0, 1)$ .

Once  $\xi$  is specified, we employ the finite element method [35,36] with bilinear basis functions to numerically solve this equation. The forward model is defined as

$$G(\xi) := u(\xi) = \mathbf{c}(s(\xi)) = \{s(x, \xi), x \in S\} \in \mathbb{R}^d, \quad (51)$$

where  $S$  is the collection of sensor points in the physical space and is defined as  $S := \{(x_1, x_2) \in D \mid x_1 = 0.125i, x_2 = 0.125j, \forall i, j = 1, 2, \dots, 7\}$ . Consequently, the dimension of the model output is 49. To generate the ground truth of the permeability field  $a^*(x)$  utilized in the inverse problem, we randomly generate  $\{\xi_k^*\}_{k=1}^M$  from the prior distribution  $\mathcal{N}(0, 1)$  and assemble them through (50). Figure 1 illustrates the discretization of the physical domain and the locations of sensors as well as the actual permeability field  $a^*(x)$ .



**Figure 1.** The FEM grids and sensor locations (left) and the ground truth of the permeability field (right).

#### 4.2. Performance in Forward Problem

This section assesses the accuracy and efficiency of the ANOVA-GP models. As outlined in Section 3.4, the construction of an ANOVA-GP model necessitates samples of the parameters. In the absence of posterior information, the prior distribution of the parameters is considered. The anchor point  $c$  is chosen as the prior mean of the parameters, and the sample set  $\{\zeta^{(j)}\}_{j=1}^{N_{model}}$  consists of  $N_{model}$  samples randomly drawn from the prior distribution. The number of samples is set to  $N_{model} = 100$ . The threshold for retaining PCA modes is defined as  $tol_{pca} = 10^{-3}$ . We construct three ANOVA-GP models with different thresholds  $tol_{anova} = 10^{-2}, 10^{-3}, 10^{-4}$  for selecting active ANOVA terms and employ the following methods to evaluate their performance.

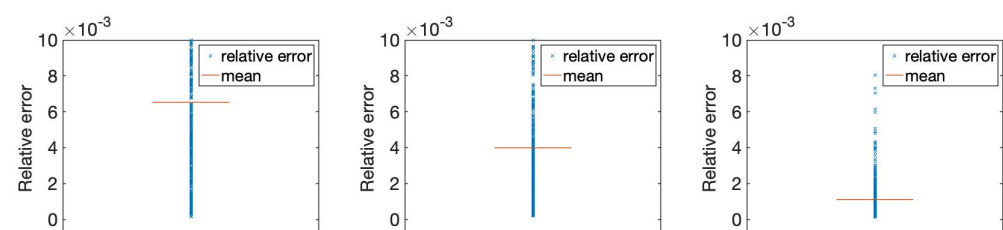
The test samples  $\{\zeta^{(j)}\}_{j=1}^{N_{test}}$  are randomly drawn from the prior distribution, with  $N_{test} = 500$ . The reference solutions  $u(\zeta^{(j)})$  are obtained using the finite element method. The relative error between the predicted mean  $\bar{u}_{\mathcal{J}}(\zeta^{(j)})$  given by the ANOVA-GP model and the reference solution is defined as

$$\text{Relative error} = \frac{\|u(\zeta^{(j)}) - \bar{u}_{\mathcal{J}}(\zeta^{(j)})\|_2}{\|u(\zeta^{(j)})\|_2}. \quad (52)$$

Table 1 and Figure 2 demonstrate that as  $tol_{anova}$  decreases, more ANOVA terms are included in  $\mathcal{J}$ : leading to improved accuracy but at the expense of increased prediction cost. To strike a balance between accuracy and efficiency,  $tol_{anova} = 10^{-3}$  is adopted in subsequent experiments.

**Table 1.** The mean of relative errors and the mean of prediction costs of ANOVA-GP models when  $tol_{anova} = 10^{-2}$  (left),  $10^{-3}$  (middle), and  $10^{-4}$  (right).

	$tol_{anova} = 10^{-2}$	$tol_{anova} = 10^{-3}$	$tol_{anova} = 10^{-4}$
<b>Mean cost (s)</b>	$1.279 \times 10^{-2}$	$1.684 \times 10^{-2}$	$2.184 \times 10^{-2}$
<b>Mean relative error</b>	$6.531 \times 10^{-3}$	$3.992 \times 10^{-3}$	$1.114 \times 10^{-3}$



**Figure 2.** The relative errors of ANOVA-GP models when  $tol_{anova} = 10^{-2}$  (left),  $10^{-3}$  (middle), and  $10^{-4}$  (right) for 500 test samples.



As mentioned in Section 3.1.2,  $\mathcal{J}$  comprises only a few low-order terms. Table 2 provides the counts of active terms in  $\mathcal{J}_i$  and  $\tilde{\mathcal{J}}_i$  at each order ( $i = 1, 2, 3$ ) as well as for the overall set  $\mathcal{J}$ . In all settings, only terms up to second-order are selected into  $\mathcal{J}$ , which is consistent with findings from prior research [13].

**Table 2.** The number of active ANOVA terms at each order when  $tol_{anova} = 10^{-2}$  (left),  $10^{-3}$  (middle), and  $10^{-4}$  (right).

$tol_{anova}$	$ \mathcal{J}_1 $	$ \tilde{\mathcal{J}}_1 $	$ \mathcal{J}_2 $	$ \tilde{\mathcal{J}}_2 $	$ \mathcal{J}_3 $	$ \mathcal{J} $
$10^{-2}$	12	1	0	0	0	12
$10^{-3}$	12	3	3	0	0	15
$10^{-4}$	12	4	6	3	0	18

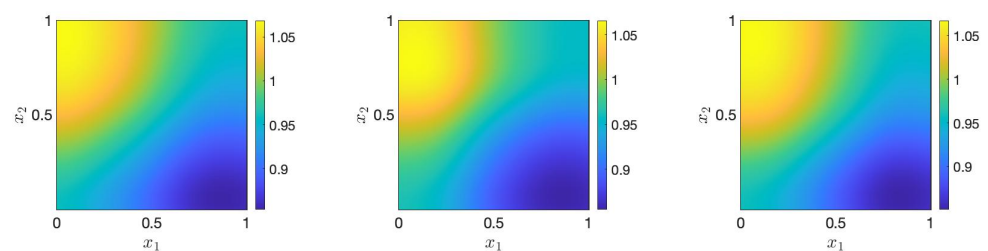
#### 4.3. Performance in Inverse Problem

This section assesses the performance of the ANOVA-GP model in MCMC and verifies the validity of the adaptive ANOVA-GP-MCMC algorithm (see Algorithm 3).

Three sets of MCMC inversions are carried out in the numerical study. The first set utilizes the finite element method (referred to as the full model in the subsequent text) in the forward evaluation. The second set employs the ANOVA-GP model constructed with prior samples (refer to Section 4.2) as a surrogate for the forward model. The third set adopts the adaptive ANOVA-GP-MCMC algorithm (see Algorithm 3), where the ANOVA-GP model is adaptively updated during sampling. These three sets are denoted as full MCMC, prior ANOVA-GP-MCMC, and adaptive ANOVA-GP-MCMC, respectively, in the subsequent text.

In all cases,  $10^6$  samples are generated through MCMC. The proposal distribution  $\pi_p(\cdot|\zeta^{(j)})$  is set to a multivariate Gaussian distribution with a mean value  $\zeta^{(j)}$  and a covariance matrix of  $0.015^2 I$ . The measurement noise of sensors is modeled as independent and identically distributed Gaussian distributions  $\epsilon \sim \mathcal{N}(0, 0.001^2)$ . For constructing the ANOVA-GP surrogate, we set  $N_{model} = 100$ . In adaptive ANOVA-GP-MCMC, the model is updated after  $N_{update} = 10^3$  new samples are generated until the set of active terms  $\mathcal{J}$  remains unchanged. The acceptance rates for the three groups are 31.13%, 31.25%, and 30.98%, respectively.

Figure 3 displays the mean of posterior samples generated by full MCMC, prior ANOVA-GP-MCMC, and adaptive ANOVA-GP-MCMC, from which we can observe that the results obtained from the three sets are similar to the ground truth (see Figure 1), indicating the effectiveness of the proposed approach.



**Figure 3.** The means of posterior samples generated by full MCMC (left), prior ANOVA-GP-MCMC (middle), and adaptive ANOVA-GP-MCMC (right).

Table 3 illustrates the average cost for a single sample in the three sets. The model based on posterior samples accelerates the sampling speed by 82.73%, evidencing the substantial improvement in efficiency achieved by the proposed method.

**Table 3.** The mean costs per sample of full MCMC (left), prior ANOVA-GP-MCMC (middle), and adaptive ANOVA-GP-MCMC (right).

	Full	Prior	Adaptive
Cost per sample (s)	$4.663 \times 10^{-2}$	$2.414 \times 10^{-2}$	$8.051 \times 10^{-3}$
Speedup	\	48.23%	82.73%

To access the accuracy of the estimated posterior mean and variance, we introduce the relative mean error  $\epsilon_{\text{mean}}$  and the relative variance error  $\epsilon_{\text{var}}$  as

$$\epsilon_{\text{mean}} := \frac{\|\mathbf{E}_{\Xi^*}(a(x, \xi)) - \mathbf{E}_{\text{ref}}\|_2}{\|\mathbf{E}_{\text{ref}}\|_2}, \quad \epsilon_{\text{var}} := \frac{\|\mathbf{V}_{\Xi^*}(a(x, \xi)) - \mathbf{V}_{\text{ref}}\|_2}{\|\mathbf{V}_{\text{ref}}\|_2}, \quad (53)$$

where the mean and variance of the posterior samples are computed through

$$\begin{aligned} \mathbf{E}_{\Xi^*}(a(x, \xi)) &:= \frac{1}{|\Xi^*|} \sum_{j=1}^{|\Xi^*|} a(x, \xi^{(j)}), \\ \mathbf{V}_{\Xi^*}(a(x, \xi)) &:= \frac{1}{|\Xi^*|} \sum_{j=1}^{|\Xi^*|} (a(x, \xi^{(j)}) - \mathbf{E}_{\Xi^*}(a(x, \xi)))^2. \end{aligned} \quad (54)$$

The reference mean estimate  $\mathbf{E}_{\text{ref}}$  and the reference variance estimate  $\mathbf{V}_{\text{ref}}$  in (53) are obtained by full MCMC with  $10^6$  samples.

Figure 4 illustrates the trend of  $\epsilon_{\text{mean}}$  and  $\epsilon_{\text{var}}$  during the sampling process of the three sets. The figure reveals that, owing to the acceleration of individual sampling, the relative error decreases more rapidly using the ANOVA-GP model than with the full model. Particularly in the adaptive ANOVA-GP model, as the anchor point and the samples used for construction are closer to the posterior distribution, fewer ANOVA terms are selected in  $\mathcal{J}$ , and a smaller number of PCA modes remain in each local problem.

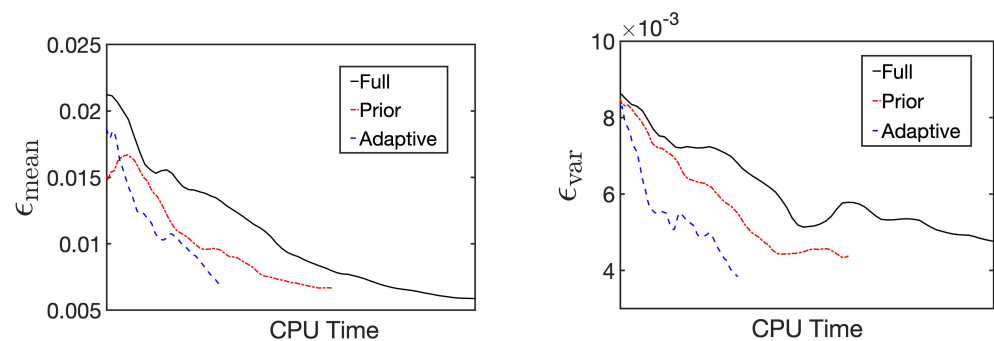
**Figure 4.** Errors in mean and variance estimates ( $\epsilon_{\text{mean}}$  and  $\epsilon_{\text{var}}$ ) for full MCMC, prior ANOVA-GP-MCMC, and adaptive ANOVA-GP-MCMC.

Table 4 shows the total number of PCA modes in the model used in prior ANOVA-GP-MCMC (referred to as the prior ANOVA-GP model) and the updating model in adaptive ANOVA-GP-MCMC (referred to as the posterior ANOVA-GP model). Because of the substantial reduction in the cost of a single prediction compared to the prior-based model, adaptive ANOVA-GP-MCMC achieves a given accuracy even more rapidly.

**Table 4.** The number of total PCA modes in the prior ANOVA-GP model (left) and the posterior ANOVA-GP model (right).

	Prior ANOVA-GP Model	Posterior ANOVA-GP Model
Number of PCA modes	47	29

## 5. Conclusions

In high-dimensional Bayesian inverse problems, establishing a cost-effective surrogate model for the forward model is pivotal for accelerating the MCMC process. Our study leverages ANOVA decomposition and PCA to break down the high-dimensional original problem into several low-dimensional input–output sub-problems. Gaussian process regression is then employed to build local predictors. To enhance the overall efficiency of the model in the Bayesian inversion setting, we propose an adaptive ANOVA-GP-MCMC algorithm that updates the model using posterior samples obtained during the sampling process. Through numerical experiments, we first validate the computational accuracy and efficiency of the ANOVA-GP model. Subsequently, we apply this model to the MCMC process to evaluate the effectiveness of our adaptive algorithm. It is important to note that our study currently focuses on ANOVA decomposition based on a single anchor point: limiting its applicability to Bayesian inverse problems with single-peak posterior distributions. For multi-peak distributions, a potential solution is to employ multiple anchor points for ANOVA decomposition. Additionally, although the number of selected ANOVA terms does not increase exponentially with the growth of the parameter dimension, this method may still lose its acceleration effect when dealing with extremely high-dimensional parameters. Developing a high-dimensional-model reduction method with broader applicability than ANOVA decomposition will be a key focus of our future work.

**Author Contributions:** Conceptualization, X.S., H.Z. and G.W.; methodology, X.S., H.Z. and G.W.; software, X.S. and H.Z.; validation, X.S., H.Z. and G.W.; investigation, X.S., H.Z. and G.W.; writing—original draft preparation, X.S. and H.Z.; writing—review and editing, X.S., H.Z. and G.W.; visualization, X.S. and H.Z.; project administration, G.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Science and Technology Commission of Shanghai Municipality (STCSM), grant number 22DZ1101200.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of variance
GP	Gaussian process
PCA	Principle component analysis
MCMC	Markov chain Monte Carlo

## References

1. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*; SIAM: Philadelphia, PA, USA, 2005.
2. Keilis-Borok, V.; Yanovskaja, T. Inverse problems of seismology (structural review). *Geophys. J. Int.* **1967**, *13*, 223–234. [\[CrossRef\]](#)
3. Beck, J.V. Nonlinear estimation applied to the nonlinear inverse heat conduction problem. *Int. J. Heat Mass Transf.* **1970**, *13*, 703–716. [\[CrossRef\]](#)
4. Wang, J.; Zabaras, N. A Bayesian inference approach to the inverse heat conduction problem. *Int. J. Heat Mass Transf.* **2004**, *47*, 3927–3941. [\[CrossRef\]](#)
5. Luo, C.; Shen, L.; Xu, A. Modelling and estimation of system reliability under dynamic operating environments and lifetime ordering constraints. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108136. [\[CrossRef\]](#)
6. Wang, W.; Cui, Z.; Chen, R.; Wang, Y.; Zhao, X. Regression analysis of clustered panel count data with additive mean models. *Stat. Pap.* **2023**, 1–22. [\[CrossRef\]](#)
7. Tarantola, A. Popper, Bayes and the inverse problem. *Nat. Phys.* **2006**, *2*, 492–494. [\[CrossRef\]](#)
8. Robert, C.P.; Casella, G.; Casella, G. *Monte Carlo Statistical Methods*; Springer: New York, NY, USA, 1999; Volume 2.
9. Yeh, W.W.G. Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resour. Res.* **1986**, *22*, 95–108. [\[CrossRef\]](#)
10. Virieux, J.; Operto, S. An overview of full-waveform inversion in exploration geophysics. *Geophysics* **2009**, *74*, WCC1–WCC26. [\[CrossRef\]](#)

11. Marzouk, Y.M.; Najm, H.N.; Rahn, L.A. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comput. Phys.* **2007**, *224*, 560–586. [\[CrossRef\]](#)
12. Wang, H.; Li, J. Adaptive Gaussian process approximation for Bayesian inference with expensive likelihood functions. *Neural Comput.* **2018**, *30*, 3072–3094. [\[CrossRef\]](#)
13. Chen, C.; Liao, Q. ANOVA Gaussian process modeling for high-dimensional stochastic computational models. *J. Comput. Phys.* **2020**, *416*, 109519. [\[CrossRef\]](#)
14. Ma, X.; Zabaras, N. An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method. *Inverse Probl.* **2009**, *25*, 035013. [\[CrossRef\]](#)
15. Galbally, D.; Fidkowski, K.; Willcox, K.; Ghattas, O. Non-linear model reduction for uncertainty quantification in large-scale inverse problems. *Int. J. Numer. Methods Eng.* **2010**, *81*, 1581–1608. [\[CrossRef\]](#)
16. Lieberman, C.; Willcox, K.; Ghattas, O. Parameter and state model reduction for large-scale statistical inverse problems. *SIAM J. Sci. Comput.* **2010**, *32*, 2523–2542. [\[CrossRef\]](#)
17. Frangos, M.; Marzouk, Y.; Willcox, K.; van Bloemen Waanders, B. Surrogate and reduced-order modeling: A comparison of approaches for large-scale statistical inverse problems. In *Large-Scale Inverse Problems and Quantification of Uncertainty*; John Wiley & Sons: Chichester, UK, 2010; pp. 123–149.
18. Li, J. A note on the Karhunen–Loève expansions for infinite-dimensional Bayesian inverse problems. *Stat. Probab. Lett.* **2015**, *106*, 1–4. [\[CrossRef\]](#)
19. Sobol', I.M. Theorems and examples on high dimensional model representation. *Reliab. Eng. Syst. Saf.* **2003**, *79*, 187–193. [\[CrossRef\]](#)
20. Gao, Z.; Hesthaven, J.S. On ANOVA expansions and strategies for choosing the anchor point. *Appl. Math. Comput.* **2010**, *217*, 3274–3285. [\[CrossRef\]](#)
21. Elman, H.C.; Liao, Q. Reduced basis collocation methods for partial differential equations with random coefficients. *SIAM/ASA J. Uncertain. Quantif.* **2013**, *1*, 192–217. [\[CrossRef\]](#)
22. Liao, Q.; Li, J. An adaptive reduced basis ANOVA method for high-dimensional Bayesian inverse problems. *J. Comput. Phys.* **2019**, *396*, 364–380. [\[CrossRef\]](#)
23. Ren, O.; Boussaidi, M.A.; Voytsekhovskiy, D.; Ihara, M.; Manzhos, S. Random Sampling High Dimensional Model Representation Gaussian Process Regression (RS-HDMR-GPR) for representing multidimensional functions with machine-learned lower-dimensional terms allowing insight with a general method. *Comput. Phys. Commun.* **2022**, *271*, 108220. [\[CrossRef\]](#)
24. Boussaidi, M.A.; Ren, O.; Voytsekhovskiy, D.; Manzhos, S. Random sampling high dimensional model representation Gaussian process regression (RS-HDMR-GPR) for multivariate function representation: Application to molecular potential energy surfaces. *J. Phys. Chem. A* **2020**, *124*, 7598–7607. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [\[CrossRef\]](#)
26. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [\[CrossRef\]](#)
27. Rabitz, H.; Aliş, Ö.F. General foundations of high-dimensional model representations. *J. Math. Chem.* **1999**, *25*, 197–233. [\[CrossRef\]](#)
28. Ma, X.; Zabaras, N. An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. *J. Comput. Phys.* **2010**, *229*, 3884–3915. [\[CrossRef\]](#)
29. Ki, W.C.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 14.
30. Rasmussen, C.E. Gaussian processes in machine learning. In *Summer School on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 63–71.
31. Schulz, E.; Speekenbrink, M.; Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **2018**, *85*, 1–16. [\[CrossRef\]](#)
32. Williams, C.; Rasmussen, C. Gaussian processes for regression. *Adv. Neural Inf. Process. Syst.* **1995**, *8*, 514–520.
33. Chen, Z.; Fan, J.; Wang, K. Remarks on multivariate Gaussian process. *arXiv* **2020**, arXiv:2010.09830.
34. Li, J.; Marzouk, Y.M. Adaptive construction of surrogates for the Bayesian solution of inverse problems. *SIAM J. Sci. Comput.* **2014**, *36*, A1163–A1186. [\[CrossRef\]](#)
35. Elman, H.C.; Silvester, D.J.; Wathen, A.J. *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*; Oxford University Press: Oxford, UK, 2014.
36. Elman, H.C.; Ramage, A.; Silvester, D.J. IFISS: A computational laboratory for investigating incompressible flow problems. *SIAM Rev.* **2014**, *56*, 261–273. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.