



Ensemble Bayesian method for parameter distribution inference: application to reactor physics

Jia-Qin Zeng¹ · Hai-Xiang Zhang¹ · He-Lin Gong² · Ying-Ting Luo¹

Received: 22 August 2023 / Revised: 26 October 2023 / Accepted: 31 October 2023 / Published online: 11 December 2023

© The Author(s), under exclusive licence to China Science Publishing & Media Ltd. (Science Press), Shanghai Institute of Applied Physics, the Chinese Academy of Sciences, Chinese Nuclear Society 2023

Abstract

The estimation of model parameters is an important subject in engineering. In this area of work, the prevailing approach is to estimate or calculate these as deterministic parameters. In this study, we consider the model parameters from the perspective of random variables and describe the general form of the parameter distribution inference problem. Under this framework, we propose an ensemble Bayesian method by introducing Bayesian inference and the Markov chain Monte Carlo (MCMC) method. Experiments on a finite cylindrical reactor and a 2D IAEA benchmark problem show that the proposed method converges quickly and can estimate parameters effectively, even for several correlated parameters simultaneously. Our experiments include cases of engineering software calls, demonstrating that the method can be applied to engineering, such as nuclear reactor engineering.

Keywords Model parameters · Bayesian inference · Frequency distribution · Ensemble Bayesian method · KL divergence

1 Introduction

With the depth of its intellectual development, mathematical modeling and simulation (M & S) have emerged as a powerful tool that promises to revolutionize engineering and science. M & S are expected to describe a system, on the basis of which further research can be conducted. For example, the Consortium for the Advanced Simulation of Light Water Reactors (CASL) was established by the US Department of Energy (DOE) in 2010, with the goal of providing M & S capabilities that support and accelerate the improvement of nuclear energy's economic competitiveness and ensure nuclear safety [1].

Numerical simulations can differ significantly from experimental observations, and minimizing this difference has always been an important topic in engineering practice [2–4]. The difference between model and reality is particularly true in reactor physics, primarily for the following reasons:

- (1) The complex nature of nuclear phenomena. A nuclear system may involve neutron transport, thermal hydraulics, and fuel performance, among others, which makes it difficult to accurately model the neutronic behavior. Theoretically, this can be well described by the Boltzmann transport equation, incorporating coefficients derived from various experimentally evaluated neutron interaction cross sections. These cross sections have strong, discontinuous behavior in space due to material heterogeneities and extreme variations with energy due to resonance phenomena associated with compound nucleus formation [5–7].
- (2) Inaccuracy of input parameters. A model requires input design data or explanatory data, such as geometry, composition, and control parameters, and input model parameters, including physical constants such as interaction cross sections and material thermal and mechanical properties. Particularly, the *two-level* approach

This research was partially sponsored by the Natural Science Foundation of Shanghai (No.23ZR1429300) and the Innovation Fund of CNNC (Lingchuang Fund).

✉ He-Lin Gong
gonghelin@sjtu.edu.cn
Ying-Ting Luo
lyt83@163.com

¹ School of Mathematics, Sichuan University, Chengdu 610065, China

² Paris Elite Institute of Technology, Shanghai Jiao Tong University, Shanghai 200240, China

(i.e., a *micro-level* model for determining macro cross sections and a *macro-level* model for determining the final response of interests) for reactor physics calculation may lead inexact input parameters [5–7]. As a consequence, inexact parameters will clearly cause inaccurate outputs.

With the advancement of science and computational power, we can probably make fewer simplifications and approximations in our models, resulting to reduced differences, whereas the resources spent must still be considered for economic reasons. As we reduce this difference, it becomes increasingly dominated by the input parameters.

In particular, nuclear data, such as neutron cross sections and resonance parameters, are evaluated by combining approximated nuclear physics and experimental observations, which have epistemic and aleatoric uncertainties. These uncertainties then propagate through a neutronics model, creating differences and uncertainties [8]. It is easy to imagine that it would be the best option if we were able to re-evaluate the nuclear data. Unfortunately, the nuclear evaluation process is significantly long and costly, making it unrealistic. For example, [9] estimated the cost of a single new observation of a nuclear datum at 400,000 US dollars.

The uncertainty of the input parameters causes uncertain behavior in the model, resulting to the development of three fields in statistics: uncertainty quantification (UQ), sensitivity analysis (SA), and data assimilation (DA) [10]. They are beneficial in engineering design, with the potential to significantly reduce unnecessary costs. For example, knowledge of the sensitivity of attributes has been used to reduce costly design margins.

Data assimilation assimilates previous experience or experimental observations to reduce the differences and uncertainties in model predictions. It was first proposed in the field of meteorology and is widely used to predict meteorological phenomena, such as hurricanes [11]. Data assimilation methods can be roughly divided into two categories: variational data assimilation [12–14] and sequential data assimilation [15–17], whereas some hybrid data assimilation techniques have emerged [18–20]. Bayesian approaches to data assimilation were established after [21, 22] and [23] provided an overview of the Bayesian perspective, discussing some approaches from this perspective. In neutronics, data assimilation is primarily used to predict the bias of critical systems [24], thereby reducing the uncertainty of advanced reactor designs [25]. In recent years, research interest in data assimilation in neutrons has mainly focused on the adjustment of nuclear data, which could establish cross-correlations between nuclear data or nuclides that are traditionally unavailable [26, 27]. In addition, new methods based on stochastic sampling of input parameters have been proposed, such as MOCABA [28] and

BMC [29]. In addition, D. Siefman showed that these two methods yield similar results to each other, as well as the traditional method known as generalized linear least squares (GLLS) [30]. Readers can refer to [31–33] for the Bayesian applications in thermal hydraulics.

The method proposed in this study involves sampling the probability distribution, and the method we used is MCMC. The MCMC method began with the Metropolis method proposed by Metropolis et al. in 1953 and was extended by W.K. Hastings in 1970. In 1984, S. Geman et al. demonstrated how the method, known as the Gibbs sampler, can be adapted to high-dimensional problems that arise in Bayesian statistics [34]. In 1987, S. Duane et al. combined MCMC and molecular dynamics methods, known as the hybrid Monte Carlo method or Hamiltonian Monte Carlo (HMC) [35]. Later, scholars developed a series of improved HMC, such as the No-U-Turn Sampler (NUTS) [36] and stochastic gradient Langevin dynamics (SGLD) [37], based on HMC.

Although the Bayesian approach has been widely used in data assimilation, there is still a lack of research on the selection of a prior distribution, which is generally selected as the probability representation that exists before experimental observations, and is probably unsatisfactory. It is evident that the prior distribution affects the results of the method. This study aims to complete this task, namely, to provide the initial input of prior distributions for classical data assimilation methods. As an example, recent work [38] has applied the Bayesian neural network method to predict the beta decay lifetime of atomic nuclei and their uncertainty, which improves learning accuracy and uses prior distributions of parameters; therefore, it may be useful to consider using the method proposed in this work once before the algorithm to improve the performance of the method with more accurate prior distributions. Moreover, most data assimilation methods only provide point estimations for parameters; however, what if nuclear data are actually probabilistic? In fact, certain parameters in our understanding resemble distributions or random variables, such as the number of neutrons released per fission and the change in the concentration of nuclear fuel over a short period due to fuel depletion. We attempt to fill this research gap and take the first step in this distribution inference. Therefore, we propose an ensemble Bayesian method. It is also worth mentioning the following:

- (1) The proposed method is simulation-based. As strategies for the design and evaluation of complex nuclear systems have shifted from heavy reliance on expensive experimental validation to highly accurate numerical simulations, more attention has been paid to simulation-based approaches. This allows our method to play a role in guiding the overall system design as well as optimizing the setup of the validating benchmark experiments.

(2) The proposed method has good compatibility. Because there are numerous legacy codes used extensively in the design process, they are expected to persist as integral parts of nuclear design calculations. Our method can be incorporated in the legacy codes, making it tractable for nuclear systems simulation.

The remainder of this paper is organized as follows. In Sect. 2, we describe the parameter distribution inference problem in a general manner and describe the corresponding experiment. In Sect. 3, we first introduce the Bayesian inference and Metropolis–Hastings algorithms, followed by the core algorithm. In Sect. 4, we analyze the convergence of the method and illustrate a series of numerical tests on the proposed method, including a finite cylindrical reactor and the 2D IAEA benchmark problem. Finally, we provide a brief conclusion in the last section.

2 Problem description

When modeling a physical phenomenon, the first step is to introduce reliable theories, such as conservation laws and transport theory. In nuclear reactor physics, there are many classic models such as the finite cylindrical reactor and the 2D IAEA benchmark problem, which will also be presented in Sect. 4. However, it is worth emphasizing that the method proposed in this study is valid, regardless of the theories introduced. For clarity, we skip this step and consider the following general system:

$$\begin{cases} y(x, \mu) = f(x, \mu) \\ z(x, \mu) = y(x, \mu)(1 + \epsilon) \end{cases} \tag{1}$$

where $x \in D \subset \mathbf{R}^3$ is a spatial vector, $y \in \mathbf{R}$ is an observable physical quantity, $z \in \mathbf{R}$ is an observation of y possibly with noise ϵ , and $\mu = (\mu^{(1)}, \dots, \mu^{(n)})^T$ are input parameters of the model.

f represents the intrinsic functional relationship between these variables, which models the corresponding physical phenomenon. $\epsilon \sim N(0, \sigma^2)$ is the zero-mean Gaussian relative noise that models the observation noise.

2.1 Perspective of random variables

In current research, most data assimilation methods only provide point estimations for the model parameters μ . However, certain parameters in our understanding resemble distributions or random variables, such as the number of neutrons released per fission and the change in the concentration of nuclear fuel over a short period due to fuel depletion. Here, we deal with the latter case. Thus, $\mu = \mu(\theta) = (\mu^{(1)}(\theta), \dots, \mu^{(n)}(\theta))^T$ is assumed to be an

n -dimensional parameter vector dependent on another parameter θ . The general parameter $\theta \in [a, b]$ might be time t or burnup Bu . We have the following theorem [39]: Let $\theta \sim U[a, b]$ and $\mu : \mathbf{R} \rightarrow \mathbf{R}^n$ be a Borel measurable map; then, $\mu(\theta)$ is an n -dimensional random vector.

In the following discussion, we are concerned with the overall statistical regularity of the parameter μ within the interval $[a, b]$ and will always treat μ as a random vector with an unknown real joint distribution. Our goal is to estimate this distribution using observational data.

2.2 Experimental setup

To achieve our goal, a special type of experimental setup was designed to obtain the observational data required by the method proposed in the next section. Specifically, we obtained data in the following manner:

- (i) Choose m number of spatial positions: $x_1, \dots, x_m \in D$ and set $S = \{x_1, \dots, x_m\}$ to be the set of sensor nodes.
- (ii) Repeatedly sample k number of $\theta : a \leq \theta_1 < \dots < \theta_k \leq b$ in the uniform distribution $U(a, b)$. As an example, if θ means time t , $\theta_1, \dots, \theta_k$ are the observation times.
- (iii) At each θ_j , m number of observations $z_{1,j}, \dots, z_{m,j}$ are obtained through the sensors.

Because we used m sensors, we obtained the following m observation equations:

$$z_i(x, \mu) = y(x, \mu)(1 + \epsilon_i), i = 1, \dots, m. \tag{2}$$

where ϵ_i is the zero-mean Gaussian relative noise of the i th sensor. It is natural to assume that they are independent and have a common distribution, i.e., $\epsilon_i \sim N(0, \sigma^2), \forall i = 1, \dots, m$.

After the previous steps, we obtain $m \times k$ observations $z_i(x_i, \mu(\theta_j)), i = 1, \dots, m; j = 1, \dots, k$. For simplicity, we use the notation

$$z_{i,j} = z_i(x_i, \mu(\theta_j)). \tag{3}$$

3 Ensemble Bayesian method

The Bayesian approach has been widely used in data assimilation and is the method proposed in this study. Subsequently, we provide an overview of the Bayesian inference.

3.1 Bayesian inference

Let X and Y be the quantities of interest that cannot and can be directly observed, respectively, of which the observations make up our data. From the perspective of Bayesian

inference, there exists a joint probability distribution of all unobserved and observable quantities denoted by $p(x, y)$. Bayesian inference determines the conditional distribution of the unobserved quantities of interest given the observation data. This is formally accomplished by applying the Bayesian formula:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \tag{4}$$

provided $0 < p(y) < +\infty$ is a normalized constant.

We obtain the formula for the posterior distribution, which we will sample later. Some commonly used sampling algorithms include adaptive rejection sampling. However, in many cases, there is no analytical solution for the posterior formula, which makes it difficult to use many sampling algorithms. To overcome this problem, we introduce the MCMC method, which is a technique for sampling probability distributions. Here, we consider the Metropolis–Hastings algorithm, which has unique advantages for posterior distribution sampling.

3.2 Metropolis–Hastings algorithm

The Classical Metropolis–Hastings algorithm [40] was first proposed in the early 1950s and extended by W.K. Hastings in 1970. As the posterior distribution that we derive later is continuous, the algorithm for the continuous state space is given below.

Let π be a probability density function that must be sampled and π_s be the value of π in s . Let T be a transition function for any irreducible Markov chain with the same state space as π and $T_{s,t}$ be the value of a conditional density function in t given condition s . Furthermore, we assumed that we knew how to sample from T . Chain T is used as a proposal

chain, generating the elements of a sequence that the algorithm decides whether to accept. The steps of the Metropolis–Hastings algorithm are presented in Algorithm 1.

For the Metropolis–Hastings proposal distribution, if we choose a symmetric distribution with the current state as the symmetric point, such as a uniform distribution on an interval of length two centered at the current state, or a Gaussian distribution with the expectation of the current state, we can simplify Eq. (1) to:

$$\alpha = \min \left\{ 1, \frac{\pi_{\mu'}}{\pi_{\mu_{u-1}}} \right\}, \tag{5}$$

because $T_{\mu', \mu_{u-1}} = T_{\mu_{u-1}, \mu'}$. This is the main reason why this algorithm works well with a posterior distribution. Next, we use the Gaussian distribution with the expectation of the current state as the proposal distribution.

When π is the joint probability density function of multivariate random variables, the proposed chain T from which we know how to sample, is difficult to obtain. In this case, the single-component Metropolis–Hastings algorithm is used. We denote $T_s^{(v)}$ as a one-dimensional proposal distribution given condition s , satisfying

$$T_s^{(v)}(t^{(v)}) = \frac{T_{s, [s^{(1)}, \dots, s^{(v-1)}, t^{(v)}, s^{(v+1)}, \dots, s^{(n)}]} T}{\int_{t^{(v)}} T_{s, [s^{(1)}, \dots, s^{(v-1)}, t^{(v)}, s^{(v+1)}, \dots, s^{(n)}]} T dt^{(v)}} \tag{6}$$

where $\int_{t^{(v)}} T_{s, [s^{(1)}, \dots, s^{(v-1)}, t^{(v)}, s^{(v+1)}, \dots, s^{(n)}]} T dt^{(v)}$ is the normalized constant that makes $T_s^{(v)}$ a probability density, and the single-component Metropolis–Hastings algorithm is shown in Algorithm 2.

```

Input: initial values  $\mu_0$ .
Output:  $\mu_1, \mu_2, \dots$ .
for  $u \leftarrow 1$  to  $\infty$  do
    Choose a new candidate state  $\mu'$  according to  $T$ ; that is,
    choose  $\mu'$  with probability density  $T_{\mu_{u-1}, \mu'}$ ;
    Calculate the acceptance probability
        
$$\alpha = \min \left\{ 1, \frac{\pi_{\mu'} T_{\mu', \mu_{u-1}}}{\pi_{\mu_{u-1}} T_{\mu_{u-1}, \mu'}} \right\};$$

    Choose a number  $w$  from uniform distribution  $U(0, 1)$ ;
    if  $w \leq \alpha$  then
        |  $\mu_u \leftarrow \mu'$ ;
    else
        |  $\mu_u \leftarrow \mu_{u-1}$ .
    end
end

```

Algorithm 1 Metropolis–Hastings algorithm

```

Input: initial values  $\mu_0 = (\mu_0^{(1)}, \dots, \mu_0^{(n)})^T$ .
Output:  $\mu_1 = (\mu_1^{(1)}, \dots, \mu_1^{(n)})^T, \mu_2 = (\mu_2^{(1)}, \dots, \mu_2^{(n)})^T, \dots$ .
for  $u \leftarrow 1$  to  $\infty$  do
     $\mu' \leftarrow \mu_{u-1}$ ;
    for  $v \leftarrow 1$  to  $n$  do
        Choose a candidate state  $\beta^{(v)}$  of  $\mu^{(v)}$  according to
         $T_s^{(v)}$ ; that is, choose  $\beta^{(v)}$  with probability density
         $T_s^{(v)}(\beta^{(v)})$ ;  $\mu'^{(v)} \leftarrow \beta^{(v)}$ ;
        We calculate the acceptance probability
            
$$\alpha = \min \left\{ 1, \frac{\pi_{\mu'}}{\pi_{\mu_u}} \right\}.$$

        Choose a number  $w$  from a uniform distribution
         $U(0, 1)$ .
        if  $w \leq \alpha$  then
            |  $\mu_u^{(v)} \leftarrow \beta^{(v)}$ ;
        else
            |  $\mu_u^{(v)} \leftarrow \mu_{u-1}^{(v)}, \mu'^{(v)} \leftarrow \mu_{u-1}^{(v)}$ .
        end
    end
end

```

Algorithm 2 Single-component Metropolis–Hastings algorithm

In reality, we can directly choose a one-dimensional distribution from which we know how to sample, such as uniform or Gaussian distributions, as a one-dimensional proposal distribution $T_s^{(v)}$. This avoids the difficulty of sampling from a multidimensional distribution.

3.3 Ensemble Bayesian algorithm

In this section, we derive the core method combined with the experimental setup (observation data) in Sect. 2.2, which includes three steps: Bayesian inference, Metropolis–Hastings sampling, and performance evaluation.

3.3.1 Bayesian inference

Because $\theta_1, \dots, \theta_k$ are samples of $\theta \sim U[a, b]$, it is easy to verify that $\mu(\theta_1), \dots, \mu(\theta_k)$ are also samples of $\mu(\theta)$. In view of this, our basic idea is to estimate the distribution of μ by aggregating the individual estimates of $\mu(\theta_j), j = 1, \dots, k$.

For each θ_j , there are several observations $z_{1,j}, \dots, z_{m,j}$. Denote $\mu_j = \mu(\theta_j)$ be the value to be estimated. From the perspective of Bayesian inference and the Bayesian formula, we have

$$p(\mu_j | z_{1,j}, \dots, z_{m,j}) \propto \pi(\mu_j) \cdot p(z_{1,j}, \dots, z_{m,j} | \mu_j), \tag{7}$$

where $\pi(\mu_j)$ denotes the prior distribution. In an experiment, the prior distribution is usually given by assumptions or guesses or evaluated by traditional methods with uncertainties.

As $\epsilon_i \sim N(0, \sigma^2)$ and $z_{i,j} = y(x_i, \mu_j)(1 + \epsilon_i) = f(x_i, \mu_j)(1 + \epsilon_i)$, we have

$$z_{i,j} | \mu_j \sim N(f(x_i, \mu_j), (f(x_i, \mu_j)\sigma)^2). \tag{8}$$

By the independent assumption of the observation noise,

$$p(z_{1,j}, \dots, z_{m,j} | \mu_j) = \prod_{i=1}^m p(z_{i,j} | \mu_j), \tag{9}$$

is the product of m Gaussian density functions. Finally,

$$p(\mu_j | z_{1,j}, \dots, z_{m,j}) \propto \pi(\mu_j) \cdot \prod_{i=1}^m p(z_{i,j} | \mu_j), \tag{10}$$

where $\pi(\mu_j)$ is the prior distribution, and $p(z_{i,j} | \mu_j)$ is the density function of Gaussian distribution $N(f(x_i, \mu_j), (f(x_i, \mu_j)\sigma)^2)$.

From the previous process, we obtain posterior distributions of all μ_j . These are joint probability density functions of multivariate random variables and are continuous; therefore, they are suitable for the single-component Metropolis–Hastings algorithm presented in Algorithm 2.

3.3.2 Metropolis–Hastings sampling

As mentioned earlier, we intend to use a single-component Metropolis–Hastings algorithm to construct a Markov chain X_0, X_1, \dots whose stationary distribution is $p(\mu_j | z_{1,j}, \dots, z_{m,j})$. To describe the transition mechanism for X_0, X_1, \dots , assume that at time u the chain is at state μ_u , i.e., $X_u = \mu_u = (\mu_u^{(1)}, \dots, \mu_u^{(n)})$ and the first v components of X_{u+1} are $\mu_{u+1}^{(1)}, \dots, \mu_{u+1}^{(v)}$. Then, the $(v + 1)$ th component of X_{u+1} ; that is, $\mu_{u+1}^{(v+1)}$ is determined using a two-step procedure:

- (i) Choose a new state $\beta^{(v+1)}$ according to the Gaussian proposal distribution $N(\mu_u^{(v+1)}, \sigma^{(v+1)2})$. That is, choose $\beta^{(v+1)}$ with probability density

$$p(\beta^{(v+1)}) = \frac{1}{\sqrt{2\pi}\sigma^{(v+1)}} e^{-\frac{1}{2}\left(\frac{\beta^{(v+1)} - \mu_u^{(v+1)}}{\sigma^{(v+1)}}\right)^2}. \tag{11}$$

It should be mentioned here that the selection of the standard deviation of the proposal distribution $\sigma^{(v+1)}$ is considerable, and significantly large or small distribution will affect the results of the algorithm. We must carefully choose in combination with the dimension of parameter $\mu^{(v+1)}$.

- (ii) Decide whether to accept $\beta^{(v+1)}$ or not. Let

$$\begin{aligned} \alpha(\mu_u^{(v+1)}, \beta^{(v+1)}) &= \frac{p(\mu_{u+1}^{(1)}, \dots, \mu_{u+1}^{(v)}, \beta^{(v+1)}, \mu_u^{(v+2)}, \dots, \mu_u^{(n)} | z_{1,j}, \dots, z_{m,j})}{p(\mu_{u+1}^{(1)}, \dots, \mu_{u+1}^{(v)}, \mu_u^{(v+1)}, \mu_u^{(v+2)}, \dots, \mu_u^{(n)} | z_{1,j}, \dots, z_{m,j})} \\ &= \frac{\pi(\mu_{u+1}^{(1)}, \dots, \beta^{(v+1)}, \dots, \mu_u^{(n)})}{\pi(\mu_{u+1}^{(1)}, \dots, \mu_u^{(v+1)}, \dots, \mu_u^{(n)})} \\ &= \prod_{i=1}^n \frac{f(x_i, \mu_{u+1}^{(1)}, \dots, \mu_u^{(v+1)}, \dots, \mu_u^{(n)})}{f(x_i, \mu_{u+1}^{(1)}, \dots, \beta^{(v+1)}, \dots, \mu_u^{(n)})} \\ &= e^{-\frac{1}{2} \sum_{i=1}^n \left[\frac{[z_{i,j} - f(x_i, \mu_{u+1}^{(1)}, \dots, \beta^{(v+1)}, \dots, \mu_u^{(n)})]^2}{(f(x_i, \mu_{u+1}^{(1)}, \dots, \beta^{(v+1)}, \dots, \mu_u^{(n)})\sigma)^2} - \frac{[z_{i,j} - f(x_i, \mu_{u+1}^{(1)}, \dots, \mu_u^{(v+1)}, \dots, \mu_u^{(n)})]^2}{(f(x_i, \mu_{u+1}^{(1)}, \dots, \mu_u^{(v+1)}, \dots, \mu_u^{(n)})\sigma)^2} \right]} \end{aligned} \tag{12}$$

If $\alpha(\mu_u^{(v+1)}, \beta^{(v+1)}) \geq 1$, then $\beta^{(v+1)}$ is accepted as the $(v + 1)$ th component of X_{u+1} , i.e., $\mu_{u+1}^{(v+1)} = \beta^{(v+1)}$. If $\alpha(\mu_u^{(v+1)}, \beta^{(v+1)}) < 1$, then $\beta^{(v+1)}$ is accepted with probability $\alpha(\mu_u^{(v+1)}, \beta^{(v+1)})$. If $\beta^{(v+1)}$ is not accepted, then $\mu_{u+1}^{(v+1)}$ is kept as the $v + 1$ th component of X_{u+1} , i.e., $\mu_{u+1}^{(v+1)} = \mu_u^{(v+1)}$. In other words, let U be uniformly distributed on $(0, 1)$, then

$$\mu_{u+1}^{(v+1)} = \begin{cases} \beta^{(v+1)}, & U < \alpha(\mu_u^{(v+1)}, \beta^{(v+1)}) \\ \mu_u^{(v+1)}, & \text{else} \end{cases}. \tag{13}$$

Although the Markov chain X_0, X_1, \dots has a stationary distribution π , not all outputs μ_i in Algorithm 2 are samples of

$p(\mu_j|z_{1,j}, \dots, z_{m,j})$. According to the MCMC theory, for any sample function of the Markov chain X_0, X_1, \dots , only when the number of iterations is sufficiently large, the process is approximately random sampling from the distribution $p(\mu_j|z_{1,j}, \dots, z_{m,j})$.

Let N_{con} be the number of convergence steps, N_{tot} be the total number of iteration steps. Denote $N = N_{tot} - N_{con}$. For any $j = 1, \dots, k$, we collect N number of samples $\hat{\mu}_{j,N_{con}+1}, \dots, \hat{\mu}_{j,N_{tot}}$ from $p(\mu_j|z_{1,j}, \dots, z_{m,j})$ and aggregate all of them to obtain a frequency distribution as estimation of the real distribution of μ .

MCMC is known to suffer from the problem of dependence between samples because the generated samples are implemented through a Markov chain. However, our technical path is to examine these samples as a whole and then achieve an overall estimation of the parameters; thus, the impact of these correlations on the method is minimal.

3.3.3 Performance evaluation

To quantify the quality of the estimation, we introduced the Kullback–Leibler divergence [41], which has been widely used in information theory. Let $p(x)$ and $q(x)$ be two probability density functions over R^n . The Kullback–Leibler divergence (KL divergence) between the distribution $p(x)$ and $q(x)$ is defined as

$$KL(p||q) = - \int p(x) \cdot \log \frac{q(x)}{p(x)} dx. \tag{14}$$

$KL(p||q) \geq 0$ with equality if and only if $p(x) = q(x)$. Kullback–Leibler divergence quantifies how close a probability distribution is to another one; that is, a larger KL divergence indicates a larger difference between the two distributions.

Because the Gaussian distribution is the most common type of distribution and is uniquely determined by its expectation and variance, we propose estimating the parameter using the Gaussian distribution with the expectation and variance being the mean and variance of the frequency distribution, respectively. We then evaluated the performance of the estimation by comparing the KL divergence of the prior distribution with the real distribution and that of the Gaussian distribution used for the estimation with the real distribution.

Suppose $p(x)$ and $q(x)$ are the density functions of $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, respectively.

$$KL(p||q) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}. \tag{15}$$

4 Numerical examples

In this section, we examine the effectiveness of the proposed method. Two models in nuclear reactor physics, namely the finite cylindrical reactor and the 2D IAEA benchmark problem, are tested below. The first test case, adapted from [42], has an analytical solution; the 2D IAEA benchmark problem is one of a classical benchmark problem in nuclear physics [42] and has only a numerical solution. To verify the effectiveness of the distribution estimation method, some parameters in these models were assumed to be uncertain in the following tests:

4.1 The finite cylindrical reactor

A finite cylindrical reactor is a multiplying system of a uniform reactor in the shape of a cylinder of physical radius R and height H . Its behavior is modeled using the following two-dimensional monoenergetic diffusion equation:

$$\begin{cases} D\Delta\phi(r, z) - \Sigma_a\phi(r, z) = \frac{1}{k_{eff}}\nu\Sigma_f\phi(r, z) \\ \phi(R_e, z) = 0, \phi(r, H_e) = 0, r \in (0, R_e), z \in (0, H_e) \end{cases}, \tag{16}$$

where $R_e = R + d, H_e = H + d$, and d is the extrapolated length.

By replacing the Laplacian with its cylindrical form, as shown in Fig. 1, the analytical solution of (16) is as follows when $k_{eff} = 1$:

$$\phi(r) = A \cdot J_0\left(\frac{2.405r}{R_e}\right) \cos\left(\frac{\pi z}{H_e}\right), \tag{17}$$

where $J_0(r)$ is a zeroth-order first-order Bessel function.

Our goal is to estimate the distribution of certain uncertain parameters from the observations. Later, we compare two alternative schemes and examine the convergence of the method by applying them to single-parameter estimates, followed by multiparameter estimates. We set $A = 20, H_e = 18$

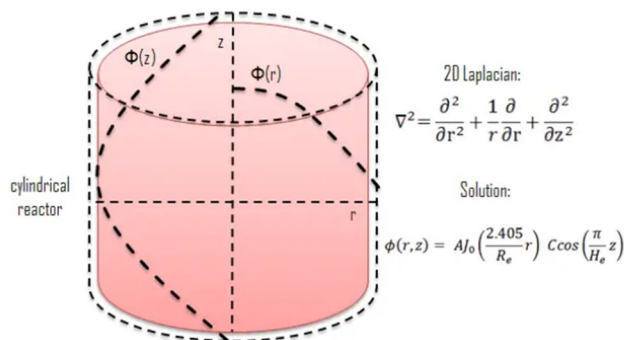


Fig. 1 (Color online) Geometry of the finite cylindrical reactor [43]

and use the ensemble Bayesian method proposed in the previous section to estimate the distribution of the parameter R_c . In other words, we consider the following system:

$$\begin{cases} y(x, \mu) = 20 \cdot \left(\frac{2.405x}{\mu} \right) \cos\left(\frac{\pi x}{18}\right), \\ z(x, \mu) = y(x, \mu)(1 + \epsilon) \end{cases} \quad (18)$$

where $x \in [0.5, 7.5] \subset \mathbf{R}$ is one-dimensional spatial position, μ is the input parameter we are interested in and is assumed to have a real distribution, and J_0 is the zeroth-order first-order Bessel function. $y \in \mathbf{R}$ is an observable physical quantity, and $z \in \mathbf{R}$ is the observation of y with noise ϵ . In this study, observations were prepared using synthetic data from simulations rather than real observations.

To prepare the observation data for our experiments, we first select m positions at equal intervals in $[0.5, 7.5]$ and repeatedly obtain the observations for the corresponding positions under k pairs of μ sampled from the real distribution. Suppose $\epsilon \sim N(0, \sqrt{0.2})$, the real distribution of μ is $N(11.0, 0.36)$ and the prior distribution is $N(12.0, 0.25)$, which could be an empirical estimate. In this case, the number of sensors m , number of samples k , number of convergence steps, and total number of steps of the Metropolis–Hastings algorithms N_{con} and N_{tot} , which are introduced in Sect. 2.2, affect the results. Subsequently, we fix the value of $N = N_{\text{tot}} - N_{\text{con}} = 100$ and discuss the effect of m, k, N_{con} on the result. Before doing so, we compare two alternative schemes after sampling from the posterior distribution $p(\mu_j | z_{1,j}, \dots, z_{m,j})$.

4.1.1 Comparison of two alternatives schemes

In this subsection, we compare these two schemes. Because μ_1, \dots, μ_k are k samples of μ , a straightforward idea is to

restore μ_1, \dots, μ_k . After introducing the Metropolis–Hastings algorithm, we can sample from the posterior distribution $p(\mu_j | z_{1,j}, \dots, z_{m,j})$ and traditional practice is to estimate μ_j using samples from $p(\mu_j | z_{1,j}, \dots, z_{m,j})$. Another scheme is to aggregate all samples from the posterior distributions $p(\mu_j | z_{1,j}, \dots, z_{m,j}), j = 1, \dots, k$, as proposed earlier, rather than using their means.

Figure 2a, b presents the results of the first and second schemes, respectively, with $m = 200, k = 200, N_{\text{con}} = 100$. We can see that although the frequency distributions obtained by these two schemes have almost the same expectation, their variances are quite different; that is, the second scheme will obtain a variance much closer to the real case. We believe this was caused by the loss of observation information during the averaging step.

More importantly, the estimation of the second scheme is intuitively much better than that of the first. This is because more samples are obtained in the second scheme, resulting in a more robust estimation.

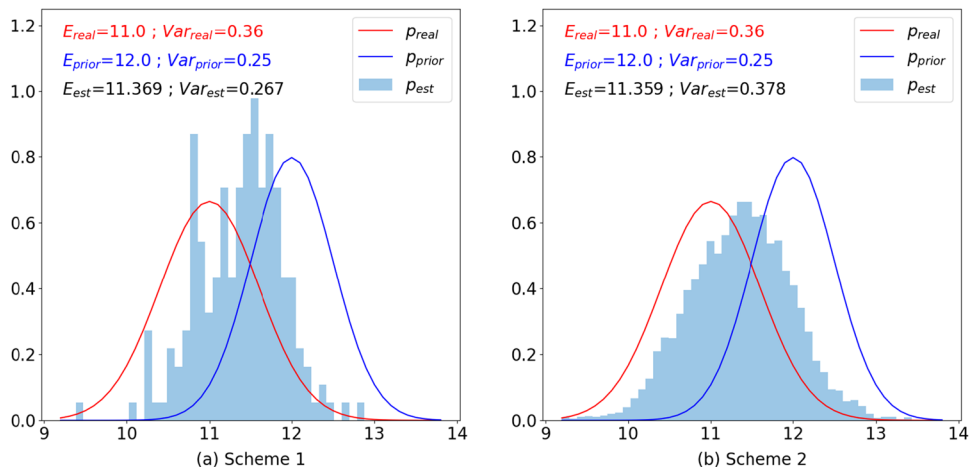
4.1.2 Convergence analysis

In this subsection, we discuss the effect of the parameters of the method on the result (i.e., the number of sensors m , number of samples k , and number of convergence steps of the Metropolis–Hastings algorithm N_{con}) and analyze the convergence of the method by setting $N = N_{\text{tot}} - N_{\text{con}} = 100$.

Figure 3 shows the effect of N_{con} on the performance of the proposed method under different settings of (m, k) . This shows that the Metropolis–Hastings sampling algorithm used in the ensemble Bayesian method converges rapidly. When N_{con} is somewhat large, such as $N_{\text{con}} = 100$, the increase hardly affects the performance of the method.

Figure 4 shows the effect of k on the performance of the method under different settings of m when $N_{\text{con}} = 100$ is fixed. This shows that when k is small, there is a fluctuation in performance, and as k increases, the performance tends

Fig. 2 (Color online) Finite cylindrical reactor problem: comparison of schemes 1 and 2, in which scheme 1 aggregates k number of means whereas scheme 2 aggregates $k \times N$ number of samples. Not only is the result of scheme 2 intuitively better, but scheme 2 also obtains expectation and variance, particularly the variance, closer to the real values



to stabilize; that is, the expectation and variance tend to be closer to the real values, which are much closer than the prior values.

Figure 5 shows the effect of m on the performance of the method when $N_{con} = 100$ and $k = 200$ are fixed. This

shows that as m increases, the performance of the method improves; that is, the expectation and variance converge to the neighborhoods of the real values, and this expectation trend is particularly pronounced.

Fig. 3 (Color online) Finite cylindrical reactor problem: the effect of increasing N_{con} on the results under three sets of parameters (m, k) , i.e., (100, 100), (150, 150), (200, 200), respectively. The curves in all three cases show that the effect of N_{con} on the results is minimal

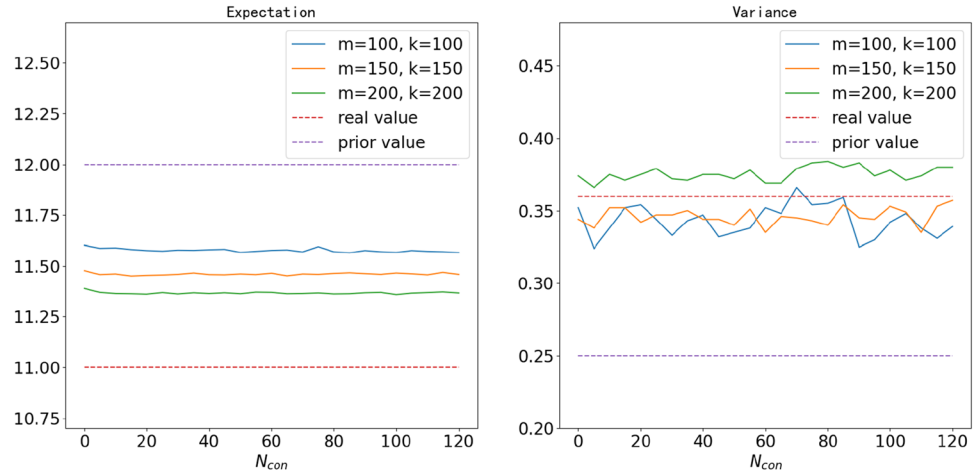


Fig. 4 (Color online) Finite cylindrical reactor problem: when $N_{con} = 100$ is fixed, the effect of increasing number of samples k on the results under three sets of parameters m , i.e., 100, 150, 200, respectively. The curves in all three cases show that the performance of the method improves with the increase of m and quickly converges to the real situation

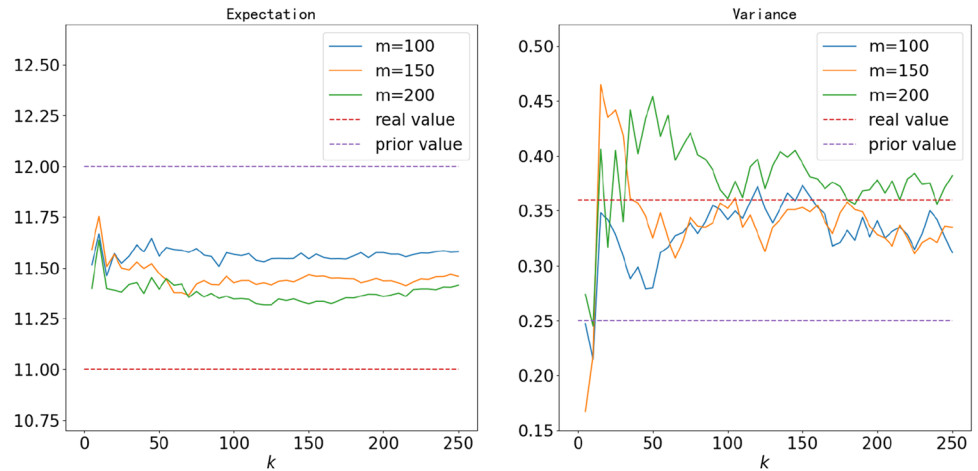
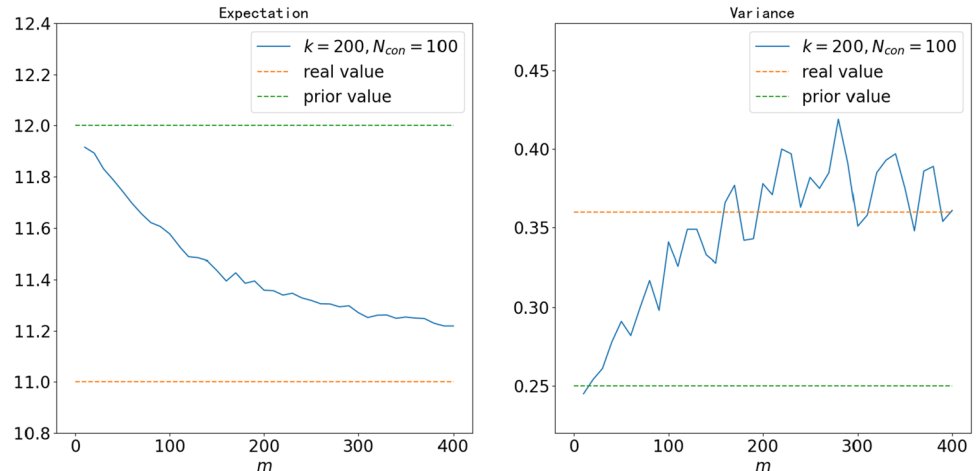


Fig. 5 (Color online) Finite cylindrical reactor problem: when $N_{con} = 100, k = 200$ is fixed, the effect of increasing number of sensors m on the results. The curve clearly shows the convergence process of algorithm performance. Compared with N_{con} and k , m has the greatest influence on the method



After applying the ensemble Bayesian method by setting $m = 200, k = 200, N_{\text{con}} = 100$, we obtained the frequency distribution shown in Fig. 2 (b). Intuitively, this frequency distribution provides an effective estimation of the real distribution. Equation (19) describes the KL divergence calculations. From the perspective of the KL divergence, we again conclude that the distribution for the estimation derived using the ensemble Bayesian method is much closer to the real distribution than the prior one.

$$KL(p_{\text{prior}}||p_{\text{real}}) = 1.418, KL(p_{\text{est}}||p_{\text{real}}) = 0.180. \tag{19}$$

4.1.3 Estimation of several parameters

In this section, we test the effectiveness of the method for estimating several parameters simultaneously because several parameters are simultaneously uncertain in many cases. Now, we use the ensemble Bayesian method proposed in this study to simultaneously estimate the distributions of three parameters A, R_e and H_e . Consider the following system:

$$\begin{cases} y(x, \mu^{(1)}, \mu^{(2)}, \mu^{(3)}) = \mu^{(3)} \cdot J_0\left(\frac{2.405x}{\mu^{(1)}}\right) \cos\left(\frac{\pi x}{\mu^{(2)}}\right), \\ z(x, \mu^{(1)}, \mu^{(2)}, \mu^{(3)}) = y(x, \mu^{(1)}, \mu^{(2)}, \mu^{(3)})(1 + \epsilon) \end{cases} \tag{20}$$

where $x \in [0.5, 7.5] \subset \mathbf{R}$ is a one-dimensional spatial position; $\mu^{(1)}, \mu^{(2)}, \mu^{(3)}$ are parameters that we are interested in and assumed to have a real joint distribution; J_0 is the zeroth-order first-order Bessel function. $y \in \mathbf{R}$ is an observable physical quantity, and $z \in \mathbf{R}$ is the observation of y with noise $\epsilon \sim N(0, \sqrt{0.2})$.

We experimented with the simulation datasets. To prepare the observational data for our experiments, we first selected 200 positions at equal intervals in $[0.5, 7.5]$ and repeatedly obtained observations for the corresponding positions

under 200 pairs of $(\mu^{(1)}, \mu^{(2)}, \mu^{(3)})$ sampled from the real joint distribution.

Suppose the real joint distribution of $(\mu^{(1)}, \mu^{(2)}, \mu^{(3)})$ be three-dimensional Gaussian distribution with expectation $\mathbf{E}(\mu^{(1)}, \mu^{(2)}, \mu^{(3)})$ and covariance matrix $\mathbf{Cov}(\mu^{(1)}, \mu^{(2)}, \mu^{(3)})$ as follows:

$$\mathbf{E}(\mu^{(1)}, \mu^{(2)}, \mu^{(3)}) = \begin{pmatrix} 11 \\ 18 \\ 20 \end{pmatrix},$$

$$\mathbf{Cov}(\mu^{(1)}, \mu^{(2)}, \mu^{(3)}) = \begin{pmatrix} 0.25 & 0.1 & 0 \\ 0.1 & 0.25 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It should be noted that the correlations between them are assumed here.

Prior distribution $\pi(\mu_1, \mu_2, \mu_3)$ is also a three-dimensional Gaussian distribution with expectation $\mathbf{E}_p(\mu_1, \mu_2, \mu_3)$ and covariance matrix $\mathbf{Cov}_p(\mu_1, \mu_2, \mu_3)$ as follows:

$$\mathbf{E}_p(\mu^{(1)}, \mu^{(2)}, \mu^{(3)}) = \begin{pmatrix} 12 \\ 19 \\ 22 \end{pmatrix},$$

$$\mathbf{Cov}_p(\mu^{(1)}, \mu^{(2)}, \mu^{(3)}) = \begin{pmatrix} 0.16 & 0 & 0 \\ 0 & 0.16 & 0 \\ 0 & 0 & 1.44 \end{pmatrix}.$$

After applying the ensemble Bayesian method by setting $m = 200, k = 200, N_{\text{con}} = 100$, we obtained three frequency distributions, as shown in Fig. 6. We find that the expectations and variances of all these frequency distributions are significantly close to the expectations and variances of the real marginal distributions. Intuitively, the frequency distributions provide an effective estimation; therefore, the ensemble Bayesian method is also suitable for the simultaneous estimation of several parameters. From the perspective

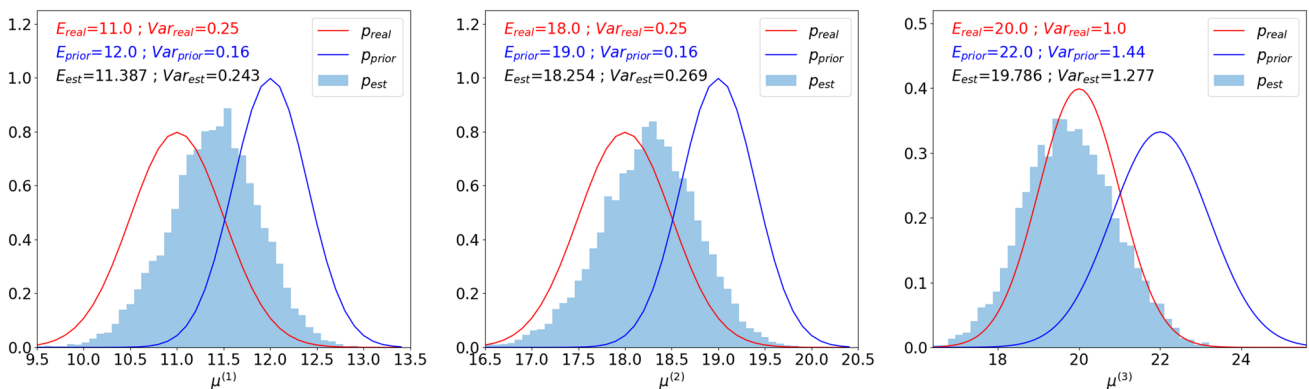


Fig. 6 (Color online) Finite cylindrical reactor problem: result of three marginal distributions when setting the number of sensors $m = 200$, number of samples $k = 200$, and number of convergence steps of Metropolis–Hastings algorithm $N_{\text{con}} = 100$

of the KL divergence, the estimation also shows a significant improvement compared to the prior distribution. Equation (21) provides the calculations for the KL divergences.

$$\begin{aligned}
 KL(p_{\text{prior}}(\mu_1)||p_{\text{real}}(\mu_1)) &= 2.431 \\
 KL(p_{\text{est}}(\mu_1)||p_{\text{real}}(\mu_1)) &= 0.300 \\
 KL(p_{\text{prior}}(\mu_2)||p_{\text{real}}(\mu_2)) &= 2.431 \\
 KL(p_{\text{est}}(\mu_2)||p_{\text{real}}(\mu_2)) &= 0.130 \\
 KL(p_{\text{prior}}(\mu_3)||p_{\text{real}}(\mu_3)) &= 2.038 \\
 KL(p_{\text{est}}(\mu_3)||p_{\text{real}}(\mu_3)) &= 0.039
 \end{aligned}
 \tag{21}$$

4.2 the 2D IAEA benchmark problem

[42] gives the definition of the 2D IAEA benchmark problem and [44] gives its implementation with neutronic code. Figure 7 illustrates the geometry of the reactor. Note that only one-quarter is given because the rest can be inferred by symmetry along the x and y axes. We denote this quarter by Ω , which is composed of four subregions with different physical properties: Neumann boundary conditions for the

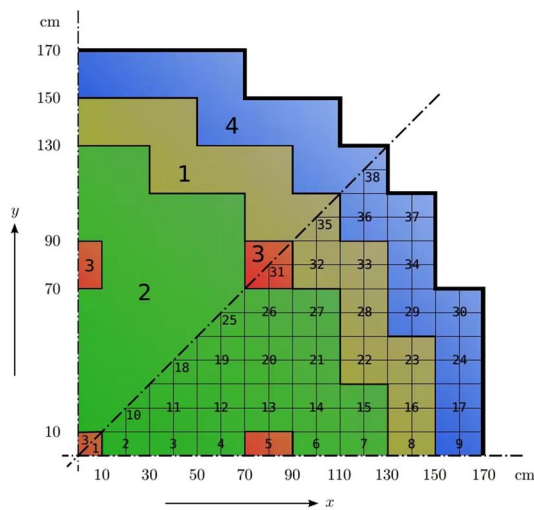


Fig. 7 (Color online) Geometry of 2D IBP, upper octant: region assignments, lower octant: fuel assembly identification [44]

Table 1 General settings of the parameters of the IAEA 2D benchmark problem

Region	D_1 (cm)	D_2 (cm)	$\Sigma_{1 \rightarrow 2}$ (cm^{-1})	$\Sigma_{a,1}$ (cm^{-1})	$\Sigma_{a,2}$ (cm^{-1})	$\nu\Sigma_{f,1}$ (cm^{-1})	$\nu\Sigma_{f,2}$ (cm^{-1})	χ_1	χ_2	Material
Ω_1	1.50	0.40	0.02	0.01	0.080	0.00	0.135	1	0	Fuel 1
Ω_2	1.50	0.40	0.02	0.01	0.085	0.00	0.135	1	0	Fuel 2
Ω_3	1.50	0.40	0.02	0.01	0.130	0.00	0.135	1	0	Fuel 2 + Rod
Ω_4	2.00	0.30	0.04	0.00	0.010	0.00	0.000	0	0	Reflector

left and bottom boundaries and the mixed boundary condition for the external border.

The 2D IAEA benchmark problem as modeled by two-dimensional two-group diffusion equations. Specifically, the flux $\phi = (\phi_1, \phi_2)^T$ (indexes 1 and 2 denote the high and thermal energy, respectively) satisfies the following eigenvalue problem [5]: Find $(\lambda, \phi) \in C \times L^\infty(\Omega) \times L^\infty(\Omega)$, s.t.

$$\begin{cases}
 -\nabla(D_1 \nabla \phi_1) + (\Sigma_{a,1} + \Sigma_{1 \rightarrow 2})\phi_1 = \lambda \chi_1 (\nu \Sigma_{f,1} \phi_1 + \nu \Sigma_{f,2} \phi_2) \\
 -\nabla(D_2 \nabla \phi_2) + \Sigma_{a,2} \phi_2 - \Sigma_{1 \rightarrow 2} \phi_1 = \lambda \chi_2 (\nu \Sigma_{f,1} \phi_1 + \nu \Sigma_{f,2} \phi_2)
 \end{cases}
 \tag{22}$$

where $\Sigma_{1 \rightarrow 2}$ is the macroscopic scattering cross section from groups 1 to 2; D_i , $\Sigma_{a,i}$, $\Sigma_{f,i}$, and χ_i are the diffusion coefficient, macroscopic absorption cross section, macroscopic fission cross section, and fission spectrum of group i , $i \in \{1, 2\}$, respectively; and ν is the average number of neutrons emitted per fission. In other words, these quantities are the model parameters. The general settings are listed in Table 1.

Later, we assumed D_2 in Ω_4 was uncertain and used the proposed ensemble Bayesian method to estimate its distribution, as it is probably one of the most important parameters. We assume that all other parameters are certain and set them according to Table 1 except for D_2 in Ω_4 . We denote this as μ , which is what we are interested in.

We experimented with simulation datasets. To prepare the observation data for our experiments, we first selected 45 positions (i.e., (10, 10), (30, 10), (30, 30), ..., (170, 10), ..., (170, 170)) in area Ω and repeatedly obtained the observations for the corresponding positions under 200 pairs of μ sampled from the real distribution.

Moreover, the two-dimensional two-group diffusion Eq. (22) must be solved in the algorithm process. This was achieved by employing the generic high-quality finite element solver FreeFem++ [45].

Suppose that the real distribution of μ is a Gaussian distribution with expectation $\mathbf{E}(\mu)$ and variance $\mathbf{Var}(\mu)$ as follows:

$$\mathbf{E}(\mu) = 2.00, \mathbf{Var}(\mu) = 0.04,
 \tag{23}$$

where $\mathbf{E}(\mu)$ is the value listed in Table 1.

The prior distribution $\pi(\mu)$ is also a Gaussian distribution with expectation $\mathbf{E}_p(\mu)$ and variance $\mathbf{Var}_p(\mu)$ as follows:

$$E_p(\mu) = 2.20, \text{Var}_p(\mu) = 0.05. \quad (24)$$

After applying the ensemble Bayesian method with setting $N_{\text{con}} = 100$, we obtained the frequency distribution shown in Fig. 8. We find that the expectation of the frequency distribution is significantly close to the real expectation, and the variance is also closer to the real variance than the prior distribution. Furthermore, the frequency distribution intuitively provided an effective estimation. From the perspective of the KL divergence, the estimation also shows a significant improvement compared to the prior estimation. Equation (25) describes the KL divergence calculations. In this case, we used the engineering software FreeFem++ to implement our algorithm because this model only has numerical solutions; therefore, the proposed ensemble Bayesian method has potential for engineering applications.

$$KL(p_{\text{prior}}||p_{\text{real}}) = 0.5134, KL(p_{\text{est}}||p_{\text{real}}) = 0.0213. \quad (25)$$

For the finite cylindrical reactor, the computation time of one-dimensional case (Fig. 2b) is 111.7 s; the computation time of three-dimensional case (Fig. 6) is 359.6s. For the 2D IAEA benchmark (Fig. 8), the computation time is about 130000 s. The main reason why 2D IAEA benchmark takes more time is that FreeFem++ takes up most of the runtime, and the most important problem is that FreeFem++ is not called or connected to the code efficiently enough. Therefore, if there is a way to call FreeFem++ efficiently, it can significantly save computing time.

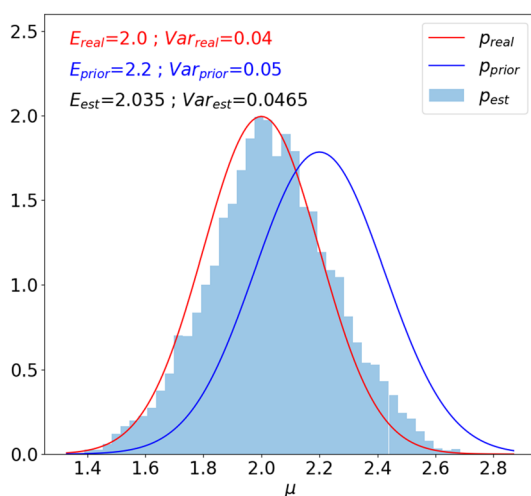


Fig. 8 (Color online) 2D IAEA benchmark problem: Result of the method when setting the number of samples $k = 200$ and the number of convergence steps of Metropolis–Hastings algorithm $N_{\text{con}} = 100$

5 Conclusion

In this study, we considered model parameters from the perspective of random variables in the context of nuclear reactor engineering and proposed a general form of parameter distribution inference problem. In the context of this parameter distribution estimation problem, we conducted a preliminary exploration and proposed an ensemble Bayesian method to estimate the parameters by obtaining frequency distributions, combined with a special type of experimental setup. Simultaneously, we introduced the KL divergence theory to quantify the estimation performance of the method. Various numerical experiments were conducted, including a finite cylindrical reactor model, which has an analytical solution, and the 2D IAEA benchmark problem, which only has a numerical solution. From the results of these tests, the following conclusions were drawn:

- For those two models, the frequency distributions intuitively provide effective estimation for the parameters. The expectations are significantly close to the real ones, and the variances are also more accurate than the prior distributions. From the perspective of KL divergence, the method also has good performance;
- The ensemble Bayesian method can estimate several parameters simultaneously, even if there are correlations between them;
- When the model has only numerical solution, we use engineering software FreeFem++ to implement our algorithm. The method also works well in this case, which means it has potential for engineering applications;
- The convergence speed of the method is fast. Thus, in general, the ensemble Bayesian method we propose has optimistic application prospects.

The proposed method shows high potential for engineering applications to correct prior distributions when using the data assimilation technique for parameter estimation.

Further studies could investigate the performance of the proposed method under moderate-scale parameters, the performance for non-Gaussian parameters, and characteristic parameters that are likely to be difficult to reproduce. As an important review of machine learning (ML) in nuclear physics, [46] not only describes the different methodologies used in ML algorithms and techniques and some applications, particularly for low- and intermediate-energy nuclear physics, but also provides a valuable summary and outlook on the possible application directions and improvements of ML algorithms in low- and intermediate-energy nuclear physics. Inspired by the review, we will

improve the efficiency of the method, and our initial idea is to improve the efficiency of the Metropolis–Hastings sampling method used in this study. Furthermore, we will determine how to involve more physics-informed or physics-guided to develop new machine learning algorithms.

Author Contributions All authors contributed to the study conception and design. Material preparation and data collection and analysis were performed by Jia-Qin Zeng, He-Lin Gong, and Hai-Xiang Zhang. The first draft of the manuscript was written by Jia-Qin Zeng, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The data that support the findings of this study are openly available in Science Data Bank at <https://www.doi.org/10.57760/sciencedb.j00186.00333> and <https://cstr.cn/31253.11.sciencedb.j00186.00333>.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

- P.J. Turinsky, D.B. Kothe, Modeling and simulation challenges pursued by the consortium for advanced simulation of light water reactors (casl). *J. Comput. Phys.* **313**, 367–376 (2016). <https://doi.org/10.1016/j.jcp.2016.02.043>
- J.P. Argaud, B. Bouriquet, J. Hunt, Data assimilation from operational and industrial applications to complex systems. *Math. Today*. 150–152 (2009)
- B. Bouriquet, J.P. Argaud et al., Robustness of nuclear core activity reconstruction by data assimilation. *Nucl. Instrum. Meth. A.* **629**, 282–287 (2011). <https://doi.org/10.1016/j.nima.2010.09.180>
- J.M. Hite, H.S. Abdel-Khalik, R.C. Smith, et al., Uncertainty quantification and data assimilation (uq/da) study on a vera core simulator component for crud analysis casl-i-2013-0184-000. Milestone Report for L 2,
- A. Hébert, *Applied Reactor Physics*, (Presses inter Polytechnique, 2009)
- S. Marguet, *The Physics of Nuclear Reactors*, (Springer, 2018). <https://doi.org/10.1007/978-3-319-59560-3>
- H. Abdel-Khalik, P. Turinsky, M. Jessee et al., Uncertainty quantification, sensitivity analysis, and data assimilation for nuclear systems simulation. *Nucl. Data. Sheets.* **109**, 2785–2790 (2008). <https://doi.org/10.1016/j.nds.2008.11.010>
- M. Salvatores, G. Palmiotti, G. Aliberti et al., Methods and issues for the combined use of integral experiments and covariance data: Results of a new international collaborative study. *Nucl. Data. Sheets.* **118**, 38–71 (2014). <https://doi.org/10.1016/J.NDS.2014.04.005>
- G. Arbanas, M. Williams, L. Leal et al., Advancing inverse sensitivity/uncertainty methods for nuclear fuel cycle applications. *Nucl. Data. Sheets.* **123**, 51–56 (2015). <https://doi.org/10.1016/j.nds.2014.12.009>
- A. Saltelli, M. Ratto, T. Andres, et al., *Global Sensitivity Analysis. The Primer*, (John Wiley & Sons, Ltd, 2008). <https://doi.org/10.1002/9780470725184>
- G. Evensen, *Data Assimilation: The Ensemble Kalman Filter* (Springer-Verlag, 2006)
- R. Florence, Overview of global data assimilation developments in numerical weather-prediction centres. *Q. J. Roy. Meteor. Soc.* **131**, 3215–3233 (2005). <https://doi.org/10.1256/qj.05.129>
- F. Rabier, H. Järvinen, E. Klinker et al., The ecmwf operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics. *Q. J. Roy. Meteor. Soc.* **126**, 1143–1170 (2000). <https://doi.org/10.1002/qj.49712656415>
- P. Courtier, J.N. Thépaut, A. Hollingsworth, A strategy for operational implementation of 4d-var, using an incremental approach. *Q. J. Roy. Meteor. Soc.* **120**, 1367–1387 (1994). <https://doi.org/10.1002/qj.49712051912>
- G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res-oceans.* **99**, 10143–10162 (1994). <https://doi.org/10.1029/94JC00572>
- P.L. Houtekamer, H.L. Mitchell, Data assimilation using an ensemble kalman filter technique. *Mon. Weather. Rev.* **126**, 796–811 (1998). [https://doi.org/10.1175/1520-0493\(1998\)126<0796:DAUAEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2)
- P.L. Houtekamer, H.L. Mitchell, G. Pellerin et al., Atmospheric data assimilation with an ensemble kalman filter: Results with real observations. *Mon. Weather. Rev.* **133**, 604–620 (2005). <https://doi.org/10.1175/MWR-2864.1>
- M. Asch, M. Bocquet, M. Nodet, *Data assimilation: methods, algorithms, and applications*, (SIAM, 2016)
- S. Cheng, C. Quilodrán-Casas, S. Ouala et al., Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. *IEEE/CAA J. Autom. Sinica.* **10**, 1361–1387 (2023). <https://doi.org/10.1109/JAS.2023.123537>
- S. Cheng, J. Chen, C. Anastasiou et al., Generalised latent assimilation in heterogeneous reduced spaces with machine learning surrogate models. *J. Sci. Comput.* **94**, 11 (2022). <https://doi.org/10.1007/s10915-022-02059-4>
- A.C. Lorenc, Analysis methods for numerical weather prediction. *Q. J. Roy. Meteor. Soc.* **112**, 1177–1194 (1986). <https://doi.org/10.1002/QJ.49711247414>
- A. Tarantola, *Inverse problem theory: methods for data fitting and model parameter estimation* (Elsevier, Amsterdam, 1987)
- C.K. Wikle, L.M. Berliner, A bayesian tutorial for data assimilation. *Physica. D.* **230**, 1–16 (2007). <https://doi.org/10.1016/J.PHYSD.2006.09.017>
- B. Broadhead, R. Childs, C. Parks, Sensitivity and uncertainty analyses applied to criticality safety validation. *Tech. Rep.*
- G. Palmiotti, T. Kugo, F. Trumble, et al., Applications of integral benchmark data. *Nucl. Sci. Eng.* **178**, 295–310. <https://doi.org/10.13182/NSE14-33>
- G. Palmiotti, M. Salvatores, G. Aliberti, A-priori and a-posteriori covariance data in nuclear cross section adjustments: Issues and challenges. *Nucl. Data. Sheets.* **123**, 41–50 (2015). <https://doi.org/10.1016/j.nds.2014.12.008>
- D. Rochman, E. Bauge, A. Vasiliev et al., Nuclear data correlation between different isotopes via integral information. *J. Comput. Phys.* **4**, 7 (2018). <https://doi.org/10.1051/EPJN/2018006>
- A. Hoefer, O. Buss, M. Hennebach et al., Mocaba: A general monte carlo-bayes procedure for improved predictions of integral functions of nuclear data. *Ann Nucl Energy.* **77**, 514–521 (2014). <https://doi.org/10.1016/j.anucene.2014.11.038>
- E. Alhassan, H. Sjöstrand, P. Helgesson et al., On the use of integral experiments for uncertainty reduction of reactor macroscopic parameters within the tmc methodology. *Prog. Nucl. Energ.* **88**, 43–52 (2016). <https://doi.org/10.1016/J.PNUCENE.2015.11.015>
- D. Siefman, Development and application of data assimilation methods in reactor physics. Ph.D. thesis (2019)
- Y. Liu, N. Dinh, R. Smith et al., Uncertainty quantification of two-phase flow and boiling heat transfer simulations through a

- data-driven modular Bayesian approach. *Int. J. Heat. Mass. Tran.* **138**, 1096–1116 (2019). <https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2019.04.075>
32. Y. Liu, X. Sun, N. Dinh, Validation and uncertainty quantification of multiphase-CFD solvers: A data-driven Bayesian framework supported by high-resolution experiments. *Nucl. Eng. Des.* **354**, 110200 (2019). <https://doi.org/10.1016/J.NUCENGDDES.2019.110200>
33. M.I. Radaideh, K. Borowiec, T. Kozlowski, Integrated framework for model assessment and advanced uncertainty quantification of nuclear computer codes under bayesian statistics. *Reliab. Eng. Syst. Safte.* **189**, 357–377 (2019). <https://doi.org/10.1016/J.RESS.2019.04.020>
34. S. Geman, D. Geman, Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE. T. Pattern. Anal.* **PAMI-6**, 721–741 (1984). <https://doi.org/10.1109/TPAMI.1984.4767596>
35. S. Duane, A. Kennedy, B.J. Pendleton et al., Hybrid monte carlo. *Phys. Lett. B.* **195**, 216–222 (1987). [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X)
36. M.D. Hoffman, A. Gelman, The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2011). <https://doi.org/10.5555/2627435.2638586>
37. M. Welling, Y.W. Teh, in *International Conference on Machine Learning, Bayesian learning via stochastic gradient langevin dynamics.* (2011)
38. W. Li, X. Zhang, Z. Niu, Studies of nuclear β -decay half-lives with Bayesian neural network approach. *Nucl. Tech.* **46**, 080013 (2023)
39. R. Durrett, *Probability: Theory and Examples* (Cambridge University Press, Cambridge, England, 2010)
40. S. Chib, E. Greenberg, Understanding the metropolis-hastings algorithm. *Am. Stat.* **49**, 327–335 (1995). <https://doi.org/10.1080/00031305.1995.10476177>
41. S. Kullback, R.A. Leibler, On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951). <https://doi.org/10.1214/AOMS/1177729694>
42. IAEA, Argonne code center: benchmark problem book. Tech. Rep. info@nuclear power.com, Diffusion equation - finite cylindrical reactor. <https://www.nuclear-power.com/nuclear-power/reactor-physics/neutron-diffusion-theory/finite-cylindrical-reactor/>
44. G.G. Theler, A. Clause, F.J. Bonetto, Solution of the 2d iaea pwr benchmark with the neutronic code milonga. (2011)
45. F. Hecht, New development in freefem++. *J. Numer. Math.* **20**, 251–265 (2012). <https://doi.org/10.1515/jnum-2012-0013>
46. W. He, Q. Li, Y. Ma et al., Machine learning in nuclear physics at low and intermediate energies. *Sci. China. Phys. Mech.* **66**, 282001 (2023)

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.