

Received 27 December 2022, accepted 7 January 2023, date of publication 16 January 2023, date of current version 24 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3237395

RESEARCH ARTICLE

Curve Fitting Algorithm of Functional Radiation-Response Data Using Bayesian Hierarchical Gaussian Process Regression Model

KWANG-WOO JUNG¹, JAEHO KIM^{1,2}, HO-JIN JUNG³, SEUNG-WON SEO³,
JI-MAN HONG³, HYOUNG-WOO BAI¹, AND SEONGIL JO^{1,4}

¹Radiation Research Division, Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute, Jeongseup 56212, Republic of Korea

²Department of Data Science, Inha University, Incheon 22212, Republic of Korea

³Insilicogen Inc., Yongin 16954, Republic of Korea

⁴Department of Statistics, Inha University, Incheon 22212, Republic of Korea

Corresponding authors: Seongil Jo (bstatsjo@gmail.com) and Hyoung-Woo Bai (hbai@kaeri.re.kr)

This work was supported by Inha University Research Grant and the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MIST) under Grant 2016M2B8A7952915.

ABSTRACT We present a nonparametric Bayesian hierarchical (NBH) model and develop a variational approximation (VA) algorithm for the curve fitting of the functional radiation response data. The NBH model is based on a Bayesian hierarchical (BH) model with a Gaussian-Inverse Wishart process (G-IWP) prior, which simultaneously smooths multiple functional observations and estimates mean-covariance functions. We use the automatic differentiation variational inference (ADVI) algorithm with a Gaussian distribution as the variational distribution for approximating the posterior distribution of parameters of interest, which is applicable to a wide class of probabilistic models and can also be implemented in Stan (a probabilistic programming system). Using the NBH model and the Gaussian ADVI algorithm, we fit a dataset for the semiconductor obtained from the radiation response map (RRM) of South Korea.

INDEX TERMS Bayesian hierarchical model, curve fitting, functional radiation data, Gaussian-inverse Wishart process, Gaussian variational approximation algorithm.

I. INTRODUCTION

Radiation technology is usually used to refer to the use of ionization and transmission force characteristics, a phenomenon in which orbital electrons of neutral atoms or molecules are reversibly separated by the action of radiation. Radiation technology is deeply intertwined with both basic and applied sciences, and it is very fashionable thanks to its property to be integrated across industry sectors ranging from traditional businesses to high-tech industries such as new material development, advanced medical technology, biotechnology, and military science. This is because, in the past, radiation was limitedly used around material conversion through radiation irradiation, but radiation technology has lately moved to fusion technology. Radiation technology is

largely divided into three generations as follows. The use of simple radiation irradiation technology in agriculture and medical fields in the 1960s is classified as the first generation, and the second generation was a period of active application of radiation fusion technology, which incorporates existing radiation irradiation technology into various fields such as chemistry, biology, nano, environment, and informatics. Currently, radiation technology is being incorporated into the fourth industry and is evolving into specific purposes of radiation fusion technology such as radiation response big data and predictable radiation molecular conversion technology, which is categorized as the third generation.

National-level databases for radiation responses, in particular, are being created, and numerous relevant research is being performed as a result of the development of statistical techniques and machine learning-based analysis methods. The United States is conducting various studies and building

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan¹.

data using neutrons at the National Institute of Standards and Technology and Oak Ridge National Laboratory [1]. Germany strategically promotes multidisciplinary research in the field of advanced radiation medicine, centering on the Heidelberg Ion-Beam Therapy Center [2]. Japan conducts comprehensive research in the field of radiation at the National Institute for Quantum Science and Technology [3]. The Korea Atomic Energy Research Institute (KAERI) established a Radiation Response Map (RRM) modeling platform to promote and support new radiation convergence industries as part of its project to expand the radiation research base. RRM is a database that systematically collects and organizes reaction information that appears when irradiated with radiation from food to industrial materials. It is worth highlighting, in particular, that this data was rigorously evaluated by the government and was awarded the highest grade of Data Quality Certification-Value (DQC-V). It is evaluated that the radiation utilization platform, which is tricky to predict and contains huge data, has been officially certified for reliability. RRM has a data integrity and consistency rate of 99.97% or higher and can be expected to boost its efficacy by sophisticated prediction models.

We employ a nonparametric Bayesian hierarchical (NBH) model to analyze the data from the RRM modeling platform. NBH incorporates Bayesian qualities that can quantify uncertainty by estimating the posterior distribution, while also including nonparametric and hierarchical properties. Since [4] proposed a hierarchical Bayesian model, NBH has gained a great deal of attention in complex network analysis and spatiotemporal modeling [5], [6], [7], [8]. The reasons are similar to the descriptions following, which are appropriate for RRM data analysis. First, it deals with nonparametric characteristics. In real-world applications, the parameter form of the probability distribution is unknown, and it is often not easy to assume this. Although Bayesian approaches are regarded as a very powerful statistical tool for measuring uncertainty, parametric Bayesian models based on assumptions about probability distributions are frequently troublesome for inference. As a result, we offer a nonparametric Bayesian approach with a Gaussian-Inverse Wishart process (G-IWP) priority in this paper. The second is about hierarchical characteristics. A hierarchical Bayesian approach is based on the premise that different models share information by sharing common parameters. The conditionally independent hierarchy is the essential foundation of hierarchical modeling, in which a set of parameters is coupled by making their distributions rely on a shared underlying parameter. As data groups with similar statistical properties exchange information, the estimated parameters' uncertainty can be reduced. This generally results in more accurate statistical estimations. Furthermore, the hierarchy is highly valuable for providing a Bayesian interpretation of the concept of shrinkage and random effects from the perspective of frequentists [5].

In general, complex models with higher parameters make it difficult to estimate the posterior distribution, which can-

not be obtained directly or, if obtained, is not derived in a well-known form in Bayesian approaches. For this reason, the Markov chain Monte Carlo (MCMC) method, which generates samples from probability distributions proportional to the posterior distribution and infers, is widely used. MCMC delivers reasonably accurate results, yet because it involves a huge amount of calculation formulas, the computational cost grows extremely long as the amount of data increases or the model becomes more intricate. In this paper, we employ variational inference, a Bayesian inference approach that applies a variational method, to address the MCMC difficulty. [17] suggested variational inference as a method for discovering and inferring specific distribution which is closest to posterior distribution among easily obtained distributions. Because it applies a distribution that is relatively straightforward to calculate rather than a posterior distribution that is tricky to calculate, this method is suitable for use not just in high-dimensional complicated models like NBH, but also in large-scale data. However, obtaining a probability distribution that is similar to the posterior distribution for variational inference is frequently problematic. [20] suggested automated differentiation variable inference (ADVI) to overcome this, which is a method of transforming parameters into real coordinate space and approximating posterior distributions to Gaussian distributions. In this paper, we estimate the model described above using ADVI, which does not need to find a probability distribution similar to the posterior distribution.

The purpose of this paper is to contribute to the growth of associated sectors via the statistical analysis of radiation response data on semiconductors. To that purpose, we propose an advanced prediction model based on the NBH model. In addition, by properly presenting the computation algorithm for this method, it becomes relatively easy for readers to use. Our main contributions are summarized as follows:

- We provide the applicability of statistical models for curve fitting of functional radiation response data. In addition to the NBH model for further sophisticated fitting, we describe the ADVI algorithm that is adaptable to a wide range of probabilistic models for computation. It also facilitates the implementation of the ADVI algorithm for the NBH model by providing Stan programming code for the convenience of readers.
- We demonstrate usefulness of radiation response data from the RRM modeling platform via data analysis on semiconductors. This can help us understand the properties of semiconductors while also encouraging innovation in radiation technology.

The rest of the paper is organized as follows. Section II presents preparations for building a model. Section III describes in detail the main model and its algorithm to be applied in this study, and the analysis of a real data set is presented in Section IV. Section VI summarizes the results of this paper and considers several possible directions for future research.

II. PRELIMINARIES

A. GAUSSIAN-INVERSE WISHART PROCESS

G-IWPs are conjugate prior distributions for the mean and covariance functions of GPs. Here we give a brief introduction about GP and IWP.

GPs are commonly used as a natural method for defining prior distributions on the space of functions with continuous domains in nonparametric Bayesian (NPB) inference [9], [10], [11]. Let $f = (f(t), t \in \mathcal{T})$ be a collection of function valued random variables at an any index set \mathcal{T} . Then f is called a GP if the marginal distribution of any finite-dimensional $f(t_1), \dots, f(t_m)$, where $m > 0$ and $t_1, \dots, t_m \in \mathcal{T}$, is a joint Gaussian distribution by the Kolmogorov's existence theorem [12]. The GP is completely determined by the mean function $\mu(t) = \mathbb{E}[f(t)]$, $t = (t_1, \dots, t_m)$ and the covariance function $k(t, t') = \mathbb{E}[(f(t) - \mu(t))(f(t') - \mu(t'))]$ for $t, t' \in \mathcal{T}$. Note that $k(\cdot, \cdot)$ is a symmetric and positive definite Mercer kernel. We will hereafter write the GP as

$$f(t) \sim \mathcal{GP}(\mu(t), k(t, t')). \quad (1)$$

Yang et al. [13] and Yang et al. [14] proposed an IWP prior for the covariance function of the GP prior to reduce biases caused by misspecifying the covariance kernel and showed that the prior provides more flexible covariance estimation and accurate fitting of functional observations. The IWP is defined such that the finite-dimensional projection $k(t, t')$ follows inverse-Wishart distribution [15] given by

$$p(k(t, t')) \propto \det(k(t, t'))^{-\frac{1}{2}(\delta+2m)} e^{-\frac{1}{2}\text{tr}(k(t, t')^{-1}\Psi(t, t'))}$$

with $\delta + m - 1$ degrees of freedom. Here $\delta > 4$ denotes a positive-integer valued shape parameter, $\text{tr}(\cdot)$ is the trace, and $\psi(t, t')$ is a symmetric and positive definite scaling parameter. We denote the IWP as

$$K \equiv k(t, t') \sim \mathcal{IW}\mathcal{P}(\delta, \Psi), \quad \Psi \equiv \psi(t, t'). \quad (2)$$

B. VARIATIONAL APPROXIMATION

Variational approximation (VA) methods (or variational Bayes methods) [16], [17], [18] are deterministic algorithms, which are alternatives to simulation-based MCMC algorithms, approximating the posterior distribution given by

$$p(\theta | y) \propto p(\theta)p(y | \theta),$$

where θ is a parameter of interest, $p(\theta)$ is the prior distribution, and $p(y | \theta)$ is a likelihood of data y . VA methods assume a class of tractable distributions, $\mathcal{Q} = \{q(\theta); \theta \in \Theta\}$, and pick the best approximation distribution (called as variational distribution) minimizing the Kullback-Leibler (KL) divergence [19] in that class as follows:

$$q^*(\theta) = \arg \min_{q \in \mathcal{Q}} KL(q(\theta) || p(\theta | y)), \quad (3)$$

where the KL divergence is non-negative and defined by

$$KL(q(\theta) || p(\theta | y)) = \int \log \frac{q(\theta)}{p(\theta | y)} q(\theta) d\theta.$$

Generally, it is difficult to minimize the KL divergence because of involving the true posterior distribution. Instead, we seek the variational distribution maximizing the so-called variational lower bound $\mathcal{L}(q)$ obtained from the following identity

$$\begin{aligned} KL(q(\theta) || p(\theta | y)) &= \int \log \frac{q(\theta)}{p(\theta)p(y | \theta)/p(y)} q(\theta) d\theta \\ &= \int \log \frac{q(\theta)}{p(\theta)p(y | \theta)} q(\theta) d\theta + \log p(y) \\ &= -\mathcal{L}(q) + \log p(y), \end{aligned}$$

where $\log p(y) = \log \int p(\theta)p(y | \theta) d\theta$ is the log marginal likelihood function and can be often used as the evidence of the model. Note that the variational lower bound is also called the evidence lower bound (ELBO).

III. NBH MODEL

In this section, we present a NBH model and develop a VA algorithm for approximating the posterior distribution. The model is the NBH model with G-IWP prior proposed by [13] and the algorithm is based on the automatic differentiation variational inference (ADVI) [20], [21] that is the VA method with Gaussian distributions as variational distributions of parameters of interest.

A. MODEL DESCRIPTION

We now describe the NBH model with G-IWP prior (hGIWP). To do this, let $\{Y_i(\cdot); i = 1, \dots, N\}$ be N independent radiation response trajectories and the i th radiation response trajectory $Y_i(\cdot)$ has m measurements on the radiation level $t_i = \{t_{i1}, \dots, t_{im}\}$ with $t_{ij} \in \mathcal{T}$ for $j = 1, \dots, m$ and $t_{i1} < t_{i2} < \dots < t_{im}$. We assume that the curve of the i th radiation response is modeled as

$$Y_i(t_{ij}) = f_i(t_{ij}) + \epsilon_i(t_{ij}), \quad (4)$$

where the errors $\{\epsilon_i(\cdot)\}$ are independent and identically distributed (i.i.d.) Gaussian random variables with mean zero and variance $\sigma^2 > 0$, i.e., $\epsilon_i(t_{ij}) \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$, and the mean curves $\{f_i(\cdot)\}$ are defined by

$$f_i(t_i) = (f_i(t_{i1}), \dots, f_i(t_{im}))^\top \stackrel{i.i.d.}{\sim} \mathcal{GP}(\mu(t_i), k(t_i, t_i)).$$

The hGIWP model assigns the G-IWP prior for the mean function $\mu(\cdot)$ and covariance function $k(\cdot, \cdot)$ as follows:

$$\begin{aligned} \mu | K &\sim \mathcal{GP}\left(\mu_0, \frac{1}{c}K\right), \\ K &\sim \mathcal{IW}\mathcal{P}(\delta, \Psi), \end{aligned} \quad (5)$$

where $\mu = \mu(t_i)$, $\mu_0 = (\mu_{01}, \dots, \mu_{0m})^\top$, c is a positive scale parameter that is set at a fixed value, and the scaling parameter

Ψ is defined by the Matérn covariance kernel [9]

$$\psi(t_i, t_l) = \sigma_z^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{|t_i - t_l|}{\rho} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{|t_i - t_l|}{\rho} \right).$$

Here σ_z^2 and ρ are the positive scale parameters, ν is the order, and $K_\nu(\cdot)$ is the modified Bessel function of second kind [22]. Note that the parameters ρ and ν are often fixed at certain values to estimate a stable covariance. For more details of the Matérn covariance kernel, see [9], [13], [14], and references therein.

For the scale parameter σ^2 of the model (4) and the hyper-parameter σ_z^2 of the Matérn covariance kernel Ψ , the hGIWP model assumes an inverse-gamma distribution and a gamma distribution as the prior distributions, respectively:

$$\sigma^2 \sim \mathcal{IGa}(a, b), \quad \sigma_z^2 \sim \mathcal{Ga}(a_z, b_z),$$

where $\mathcal{IGa}(a, b)$ denotes the inverse-gamma distribution with shape parameter a and scale parameter b and $\mathcal{Ga}(a_z, b_z)$ is the gamma distribution with shape parameter a_z and rate parameter b_z .

To facilitate posterior computation, we represent the hGIWP model (4) with the following equivalent hierarchical model:

$$\begin{aligned} Y_i(t) | f_i(t), \sigma^2 &\sim \mathcal{N}(f_i(t), \sigma^2 I_m), \quad i = 1, \dots, N, \\ f_i(t) | \mu(t), k(t, t) &\sim \mathcal{N}(\mu(t), k(t, t)), \\ \mu | K &\sim \mathcal{N}\left(\mu_0, \frac{1}{c}K\right), \\ K &\sim \mathcal{IW}(\delta, \Psi), \\ \sigma_z^2 &\sim \mathcal{Ga}(a_z, b_z), \\ \sigma^2 &\sim \mathcal{IGa}(a, b), \end{aligned} \quad (6)$$

where I_m is a $m \times m$ identity matrix and $t = \bigcup_{i=1}^N t_i$ is the pooled grid.

B. POSTERIOR INFERENCE WITH GVA ALGORITHM

We introduce a variational algorithm, which is based on ADVI method, for approximating the posterior distribution of the hGIWP model (6) proportional to, up to a constant that does not depend on the parameters of interest,

$$\begin{aligned} &\prod_{i=1}^N \mathcal{N}(Y_i(t) | f_i(t), \sigma^2 I_m) \mathcal{N}(f_i(t) | \mu(t), k(t, t)) \\ &\times \mathcal{N}(\mu | \mu_0, \frac{1}{c}K) \mathcal{IW}(K | \delta, \Psi) \\ &\times \mathcal{Ga}(\sigma_z^2 | a_z, b_z) \mathcal{IGa}(\sigma^2 | a, b). \end{aligned} \quad (7)$$

Let $\theta = (f_1^\top, \dots, f_N^\top, \mu^\top, K, \sigma_z^2, \sigma^2)^\top$ be the parameters of interest and $p(Y, \theta)$ be the joint distribution (7), where $Y = (Y_1, \dots, Y_N)^\top$, $Y_i = Y_i(t)$, and $f_i = f_i(t)$ for $i = 1, \dots, N$. Let $T(\cdot)$ be also a one-to-one differentiable function transforming the support of parameters θ to the real coordinate space \mathbb{R}^D , where D is the number of parameters. Then, the transformed joint distribution is given as

$$p(Y, \phi) = p(Y, \theta = T^{-1}(\phi)) |\det J_{T^{-1}}(\phi)|,$$

and the ELBO is

$$\mathcal{L}(q) \mathbb{E}_{q(\phi)} \left[\log p(Y, T^{-1}(\phi)) + \log |\det J_{T^{-1}}(\phi)| \right] + \mathbb{E}_{q(\phi)} [\log q(\phi)], \quad (8)$$

where $J_{T^{-1}}$ is the Jacobian of $T^{-1}(\cdot)$ and $\det(\cdot)$ is the determinant of a matrix.

To find the variational distribution $q^*(\phi)$ maximizing the ELBO (8), the ADVI considers a fully factorized (mean-field) Gaussian family

$$\mathcal{Q} = \left\{ q_\eta(\phi) : q_\eta(\phi) = \prod_{d=1}^D \mathcal{N}(\phi_d | \nu_d, \tau_d^2), \right\},$$

and chooses the variational parameter η^* maximizing the ELBO parameterized by $\eta \in \mathbb{R}^D$, i.e.,

$$\eta^* = \arg \max_{\eta \in \mathbb{R}^D} \mathcal{L}(\eta), \quad (9)$$

where $\eta = (\nu_1, \dots, \nu_D, \kappa_1, \dots, \kappa_D)^\top$ and $\kappa_d = \log \tau_d$, $d = 1, \dots, D$. Note that the ADVI uses the re-parametrization trick [24], [25], [26] and a stochastic gradient ascent method with an adaptive step-size satisfying the Robbins and Monro condition [23] for solving the above optimization problem (9). The below algorithm (Algorithm 1) describes the specific steps of the ADVI. The gradients are given as follows:

Algorithm 1: ADVI Algorithm

```

Initialization  $\nu = (\nu_1, \dots, \nu_D)^\top = 0$  and
 $\kappa = (\kappa_1, \dots, \kappa_D)^\top = 0$ ;
while change in ELBO is above some threshold do
    Generate  $S$  samples  $\xi_s \sim \mathcal{N}(0, I_D)$ ;
    Compute the unbiased estimate of  $\nabla_\nu \mathcal{L}(\eta)$  using
        Monte Carlo method;
    Compute the unbiased estimate of  $\nabla_\kappa \mathcal{L}(\eta)$  using
        Monte Carlo method;
    Update  $\nu \leftarrow \nu + \rho \nabla_\nu \mathcal{L}(\eta)$ ;
    Update  $\kappa \leftarrow \kappa + \rho \nabla_\kappa \mathcal{L}(\eta)$ ;
end

```

$$\nabla_\nu \mathcal{L}(\eta) = \mathbb{E} [\nabla_\theta \log p(Y, \phi)],$$

$$\nabla_\kappa \mathcal{L}(\eta) = \mathbb{E} \left[\nabla_\theta \log p(Y, \phi) \eta^\top \text{diag}(\exp(\kappa)) \right] + 1,$$

$$\begin{aligned} \nabla_\theta \log p(Y, \phi) &= \nabla_\theta \log p(Y, \theta) \nabla_\phi T^{-1}(\phi) \\ &\quad + \nabla_\phi \log |\det J_{T^{-1}}(\phi)|. \end{aligned}$$

In this paper, we implement the ADVI algorithm using Stan that is a probabilistic programming system [27]. The Stan code is available in Appendix and GitHub repository (<https://github.com/statjts>).

IV. A REAL CASE STUDY: SEMICONDUCTOR DATA

In this section, we present the performance of the ADVI algorithm for the hGIWP model to a competitor, the

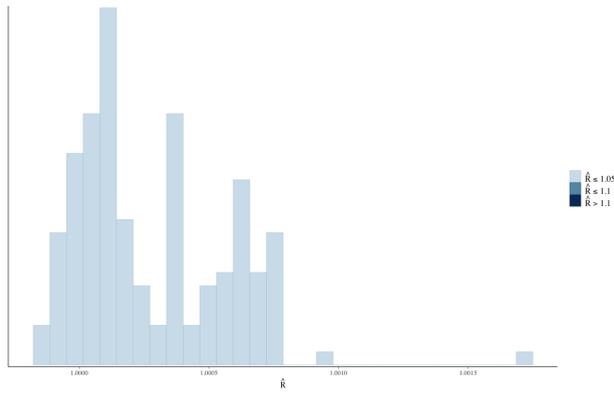


FIGURE 1. The PSRF values.

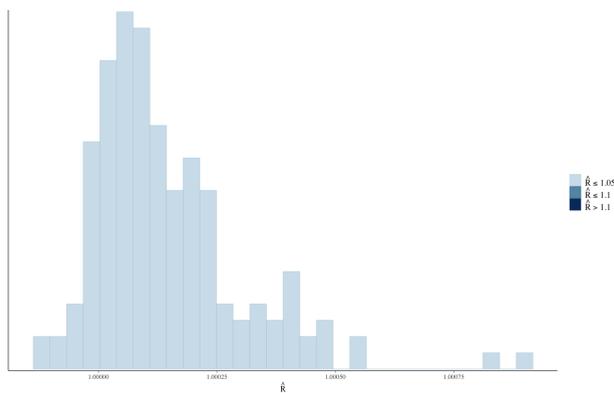


FIGURE 2. The improved PSRF values.

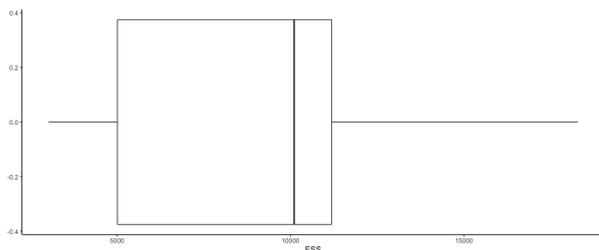


FIGURE 3. The effective sample size.

No-U-Turn Sampler (NUTS) [28] that is an extension to a simulation-based Hamiltonian Monte Carlo (HMC) method [29]. The NUTS uses a Hamiltonian dynamics simulation based on numerical integration (*leapfrog*) without tuning the step size parameter and number of steps to generate a candidate and then accept one by performing a Metropolis acceptance-rejection step. To this end, we use a dataset obtained from Radiation Response Map Modeling Platform developed by Korea Atomic Energy Research Institute (KAERI). The address of the webpage is <https://www.kaeri.re.kr/rrm/#!/home>.

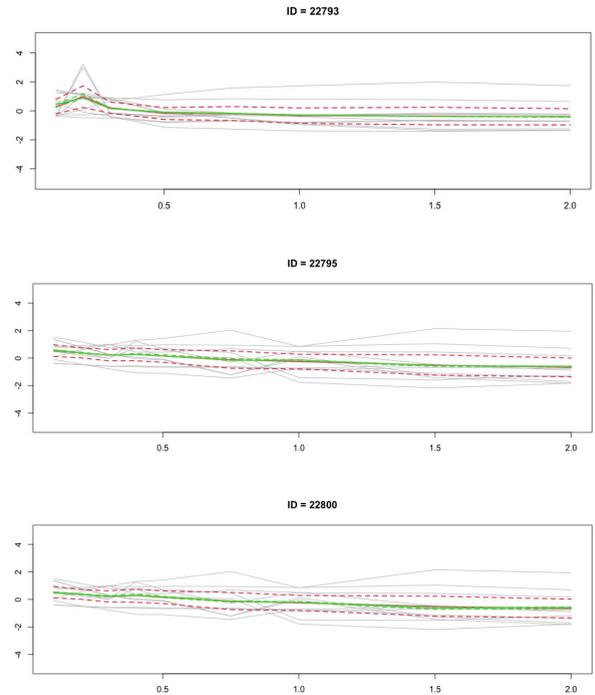


FIGURE 4. Posterior estimates for curves.

We set hyperparameters $c = 1$, $\delta = 5$, $a_z = 1.515$, $a = 2.418$, and $b = b_z = 1$ using a heuristic empirical Bayesian method as described in [13]. For inference via the NUTS algorithm, we run simultaneously four Markov chains with different sets of initial values and use 5,000 posterior samples obtained by saving every 5th sample after a 20,000 burn-in period for each chain. The total number of posterior samples for inference is 40,000. We check the convergence of the algorithm using the potential scale reduction factor (PSRF) introduced in [30] and its improved version proposed in [31]. We run the NUTS and ADVI algorithms on an Intel 10-Core Xeon W CPU@3GHz 128 GB Ram.

Figure 1 and Figure 2 show the histograms for the PSRF values of [30] and the improved PSRF values of [31], respectively. From these plots, we can see that the Markov chains are converged and mixed well. We also computed the effective sample size (ESS) in order to check the efficiency of the posterior samples and found that the minimum of ESS is about 3,000. It means that the posterior estimates obtained from these samples are reliable. We display the boxplot of ESS in Figure 3.

Table 1 shows summary statistics (mean and standard deviation) of R-squared (R^2) values and root mean squared error (RMSE) values over radiation levels (see, e.g., [33]) and Figure 4 displays results of fitting functions on the radiation-response of three semiconductors. Here the ID number indicates semiconductor. In the figure, the solid grey lines denote real observed functions. The solid and dotted red lines represent the curves and 95% credible intervals fitted by

TABLE 1. Mean and standard deviation (parenthesis) of R^2 and root mean squared error (RMSE) values over radiation levels V .

	R^2		RMSE	
	NUTS	ADVI	NUTS	ADVI
ID = 22793	0.98 (0.02)	0.97 (0.03)	0.10 (0.03)	0.13 (0.05)
ID = 22795	0.98 (0.01)	0.98 (0.01)	0.10 (0.03)	0.12 (0.02)
ID = 22800	0.99 (0.01)	0.99 (0.00)	0.10 (0.02)	0.07 (0.02)

TABLE 2. Computation times (sec).

	NUTS	ADVI
ID = 22793	4531(s)	19(s)
ID = 22795	5264(s)	23(s)
ID = 22800	5123(s)	22(s)

the NUTS method and the solid and dotted green lines are the fitted curves and 95% credible intervals approximated from the ADVI algorithm. These results indicate that the ADVI algorithm estimates the mean curves considerably accurately as the NUTS algorithm but underestimates the uncertainty (This same phenomenon appears in [32]).

Table 2 shows the computation times in seconds. From the table, we can see that the ADVI method is faster than the NUTS method. That is, the ADVI algorithm reduces the computational demands dramatically. These results suggest that the fully factorized ADVI algorithm can be used to quickly estimate the mean curves when uncertainty is not important.

V. CONCLUSION

In this paper, we have presented a nonparametric Bayesian hierarchical model with the Gaussian-Inverse Wishart process prior (hGIWP) and developed a Gaussian variational approximation (GVA) algorithm based on the automatic differentiation variational inference (ADVI) method. We have compared the GVA algorithm with the well-known No-U-Turn sampler (NUTS), a modified version of the Hamiltonian Monte Carlo (HMC) method, by fitting functions on the radiation-response of semiconductors. The results suggest that the GVA algorithm can be used for fitting the mean curves of functional data and when quick calculations are required because of reducing the computational demands.

VI. FUTURE WORK

Since the fully factorized (mean-field) GVA algorithm tends to underestimate the uncertainty, the additional approaches are needed to calculate accurate posterior variances (see, e.g., [34]). This is an important research area for future work.

APPENDIX A

In this appendix, we give the residual plots for checking normal error assumptions in model (4) of Section III. The residual plots (Figure 5) below look fine in that there do not appear to be any strong patterns.

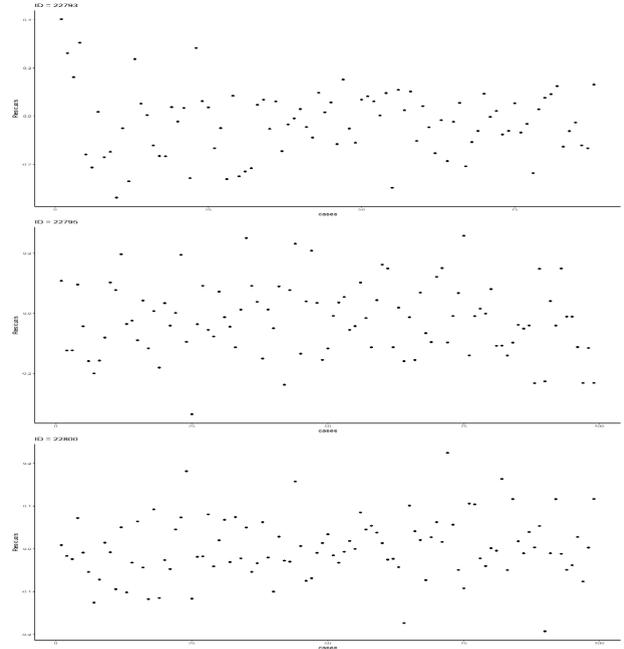


FIGURE 5. Residual plots.

APPENDIX B

We provide a Stan code for the hGIWP model.

```

data {
  int<lower=1> n;
  int<lower=1> p;
  vector[p] Y[n];
  vector[p] T[n];
  matrix[p,p] I;
  vector[p] mu0;
  matrix[p,p] A;
  real<lower=0> delta;
  real<lower=0> cc;
  real<lower=0> a;
  real<lower=0> b;
  real<lower=0> as;
  real<lower=0> bs;
}

parameters {
  real<lower=0> sigmae_sq;
  real<lower=0> sigmas_sq;
  vector[p] Z[n];
  vector[p] mu;
  cov_matrix[p] S;
}

model {
  sigmae_sq~inv_gamma(a, b);
  sigmas_sq~inv_gamma(as, bs);
  mu~multi_normal(mu0, (1/cc)*S);
  S~inv_wishart(delta+p-1, sigmas_sq*A);
}

```

```

for (i in 1:n) {
  Z[i]~multi_normal(mu, S);
}
for (i in 1:n) {
  Y[i]~multi_normal(Z[i], sigmae_sq*I);
}
}

```

ACKNOWLEDGMENT

(Kwang-Woo Jung and Jaeoh Kim contributed equally to this work.)

REFERENCES

- [1] D. L. Poster, C. C. Miller, R. A. Martinello, N. R. Horn, M. T. Postek, T. E. Cowan, Y. S. Obeng, and J. J. Kasianowicz, "Ultraviolet radiation technologies and healthcare associated infections: Standards and metrology needs," *J. Res. Nat. Inst. Standards Technol.*, vol. 126, no. 3, pp. 1–33, 2021.
- [2] L. Wegent, A. Munscher, and N. Mockelmann, "Adenoidzystisches Karzinom unklaren Ursprungs mit Infiltration der Schilddrüse: CUP-Syndrom versus primäres der Trachea," *Laryngo-Rhino-Otologie*, vol. 97, no. 2, p. 10371, 2018.
- [3] T. Bando, H. Tojo, M. Takechi, N. Aiba, T. Wakatsuki, M. Yoshida, S. Inoue, and G. Matsunaga, "On collapses in strong reversed shear plasmas during or just after plasma current ramp-up in JT-60U," *Plasma Fusion Res.*, vol. 16, Jul. 2021, Art. no. 1402089.
- [4] C. Kemp, A. Perfors, and J. B. Tenenbaum, "Learning overhypotheses with hierarchical Bayesian models," *Develop. Sci.*, vol. 10, no. 3, pp. 307–321, May 2007.
- [5] Y. W. Teh and M. I. Jordan, "Hierarchical Bayesian nonparametric models with applications," *Bayesian Nonparametrics*, vol. 28, no. 158, p. 42, 2009.
- [6] R. Salakhutdinov, J. Tenenbaum, and A. Torralba, "One-shot learning with a hierarchical nonparametric Bayesian model," in *Proc. ICML Workshop Unsupervised Transf. Learn.*, 2012, pp. 195–206.
- [7] W. Fan, L. Yang, and N. Bouguila, "Unsupervised grouped axial data modeling via hierarchical Bayesian nonparametric models with Watson distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9654–9668, Dec. 2022.
- [8] W. Xie, P. Zhang, R. Chen, and Z. Zhou, "A nonparametric Bayesian framework for short-term wind power probabilistic forecast," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 371–379, Jan. 2019.
- [9] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
- [10] J. Q. Shi and T. Choi, *Gaussian Process Regression Analysis for Functional Data*. Boca Raton, FL, USA: CRC Press, 2011.
- [11] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2022.
- [12] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ, USA: Wiley, 1995.
- [13] J. Yang, H. Zhu, T. Choi, and D. D. Cox, "Smoothing and mean-covariance estimation of functional data with a Bayesian hierarchical model," *Bayesian Anal.*, vol. 11, no. 3, pp. 649–670, Sep. 2016.
- [14] J. Yang, D. D. Cox, J. S. Lee, P. Ren, and T. Choi, "Efficient Bayesian hierarchical functional data analysis with basis function approximations using Gaussian-Wishart processes," *Biometrics*, vol. 73, no. 4, pp. 1082–1091, Dec. 2017.
- [15] A. P. Dawid, "Some matrix-variate distribution theory: Notational considerations and a Bayesian application," *Biometrika*, vol. 68, no. 1, pp. 265–274, 1981.
- [16] H. Attias, "Inferring parameters and structure of latent variable models by variational Bayes," in *Proc. 15th Conf. Uncertainty Artif. Intell.*, L. Laskey and H. Prade, Eds. San Francisco, CA, USA: Morgan Kaufmann, 1999, pp. 21–30.
- [17] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol. 37, no. 2, pp. 183–233, Nov. 1999.
- [18] M. I. Jordan and M. J. Wainwright, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, nos. 1–2, pp. 1–305, 2007.
- [19] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
- [20] A. Kucukelbir, R. Ranganath, A. Gelman, and D. M. Blei, "Automatic variational inference in Stan," in *Proc. Neural Inf. Process. Syst.*, 2015, pp. 1–9.
- [21] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei, "Automatic differentiation variational inference," *J. Mach. Learn. Res.*, vol. 18, no. 14, pp. 1–45, Jul. 2017.
- [22] M. Abramowitz and A. I. Stegun, *Handbook of Mathematical Functions*. Town, U.K.: Dover, 1965.
- [23] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, Sep. 1951.
- [24] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–14.
- [25] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1278–1286.
- [26] M. Titsias and M. Lázaro-Gredilla, "Doubly stochastic variational Bayes for non-conjugate inference," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1971–1979.
- [27] Stan Development Team, *Stan Modeling Language Users Guide and Reference Manual*, 2022.
- [28] M. D. Hoffman and A. Gelman, "The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo," *J. Mach. Learn. Res.*, vol. 15, pp. 1351–1381, Apr. 2014.
- [29] R. Neal, "MCMC using Hamiltonian dynamics," in *Handbook of Markov Chain Monte Carlo*, A. Gelman, L. Jones, and X.-L. Meng, Eds. Orange, CA, USA: Chapman, 2011, pp. 62–116.
- [30] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Stat. Sci.*, vol. 7, no. 4, pp. 457–472, Nov. 1992.
- [31] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Burkner, "Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC (with discussion)," *Bayesian Anal.*, vol. 16, no. 2, pp. 667–718, Jun. 2021.
- [32] L. S. L. Tan and D. J. Nott, "Gaussian variational approximation with sparse precision matrices," *Statist. Comput.*, vol. 28, no. 2, pp. 259–275, Mar. 2018.
- [33] J. Ramsay and B. W. Silverman, *Functional Data Analysis*. Berlin, Germany: Springer, 2005, 2005.
- [34] R. Giordano, T. Broderick, and M. I. Jordan, "Covariances, robustness, and variational Bayes," *J. Mach. Learn. Res.*, vol. 19, pp. 1–49, Jan. 2018.



KWANG-WOO JUNG received the Ph.D. degree in fungal genetics from Yonsei University, in 2015. He is currently a Senior Researcher with the Advanced Radiation Technology Institute (ARTI), Korea Atomic Energy Research Institute (KAERI). His research interest includes data collection for radiation response and its application.



JAEOH KIM received the Ph.D. degree in statistics from Korea University, in 2018. He is currently an Assistant Professor with the Department of Data Science and an Adjunct Professor with the Department of Statistics, Inha University. His research interests include data mining, data science, and statistical computation and its applications.



HO-JIN JUNG received the M.S. degree in biology from Konkuk University, in 2000. She is currently a Principal Researcher with Insilicogen Inc. Her research interests include bioinformatics, and data science and its applications.



HYOUNG-WOO BAI received the Ph.D. degree in pharmacology from the University of Kansas Medical Center, in 2010. He is currently a Principal Researcher with the Department of Radiation Science, Korea Atomic Energy Research Institute, and a Professor with the Department of Radiation Science and Technology, University of Science and Technology. His research interest includes radiation biology and its applications.



SEUNG-WON SEO received the M.S. degree in bioinformatics from the University of Science and Technology, in 2013. He is currently a Deputy Principal Developer with Insilicogen Inc. His research interests include bioinformatics, data science, and bio database system and its applications.



JI-MAN HONG received the M.S. degree in bioinformatics from Pusan National University, in 2007. He is currently a Deputy Principal Researcher with Insilicogen Inc. His research interests include bioinformatics, data science, multi-omics, and machine learning and its applications.



SEONGIL JO received the Ph.D. degree in statistics from Seoul National University, in 2014. He is currently an Associate Professor with the Department of Statistics and an Adjunct Professor with the Department of Data Science, Inha University. His research interests include Bayesian statistics, data science, and spatio-temporal models and its applications.

...