



Bayesian Projected Calibration of Computer Models

Fangzheng Xie & Yanxun Xu

To cite this article: Fangzheng Xie & Yanxun Xu (2020): Bayesian Projected Calibration of Computer Models, Journal of the American Statistical Association, DOI: [10.1080/01621459.2020.1753519](https://doi.org/10.1080/01621459.2020.1753519)

To link to this article: <https://doi.org/10.1080/01621459.2020.1753519>



View supplementary material [↗](#)



Accepted author version posted online: 13 Apr 2020.
Published online: 13 May 2020.



Submit your article to this journal [↗](#)



Article views: 206



View related articles [↗](#)



View Crossmark data [↗](#)



Bayesian Projected Calibration of Computer Models

Fangzheng Xie and Yanxun Xu

Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD

ABSTRACT

We develop a Bayesian approach called the Bayesian projected calibration to address the problem of calibrating an imperfect computer model using observational data from an unknown complex physical system. The calibration parameter and the physical system are parameterized in an identifiable fashion via the L_2 -projection. The physical system is imposed a Gaussian process prior distribution, which naturally induces a prior distribution on the calibration parameter through the L_2 -projection constraint. The calibration parameter is estimated through its posterior distribution, serving as a natural and nonasymptotic approach for the uncertainty quantification. We provide rigorous large sample justifications of the proposed approach by establishing the asymptotic normality of the posterior of the calibration parameter with the efficient covariance matrix. In addition to the theoretical analysis, two convenient computational algorithms based on stochastic approximation are designed with strong theoretical support. Through extensive simulation studies and the analyses of two real-world datasets, we show that the proposed Bayesian projected calibration can accurately estimate the calibration parameters, calibrate the computer models well, and compare favorably to alternative approaches. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received February 2019
Accepted April 2020

KEYWORDS

Asymptotic normality;
Computer experiment;
 L_2 -projection;
Semiparametric efficiency;
Uncertainty quantification

1. Introduction

With the rapid development of computational techniques and mathematical tools, computer models have been widely adopted by researchers to study large and complex physical systems. One can think of computer models as complicated nonlinear functions designed by experts using their scientific knowledge (Sacks et al. 1989; Fang, Li, and Sudjianto 2005). Compared to real physical experiments, *in silico* computer models are typically much faster and cheaper to run. Furthermore, computer models can be used to generate data that are infeasible to collect in practice. For example, a publicly available computer model called TITAN2D (Sheridan et al. 2002) that simulates granular mass flows over digital elevation models of natural terrain was developed to better understand the loss of life and disruption of infrastructure due to volcanic phenomena. Such data are impossible to collect in real life. For more applications of computer models, we refer to Fang, Li, and Sudjianto (2005), Santner, Williams, and Notz (2013), and the April 2018 issue of *Statistica Sinica* (<http://www3.stat.sinica.edu.tw/statistica>), which are devoted to computer experiments and uncertainty quantification.

In this article, we consider the *calibration* problem in computer models when they include not only the variables that can be measured, often referred to as the *design*, but also unknown parameters that are not directly available in the physical system. These parameters are referred to as the *calibration parameters* in the literature (Kennedy and O'Hagan 2001). The goal of calibration is to estimate the calibration parameters by combining observational data from the physical systems and

simulated data from the computer models, so that the computer models with the estimated calibration parameters plugged-in provide reasonable approximations to the underlying physical systems. Formally, we model the outputs $(y_i)_{i=1}^n$ of the physical system η at design $(\mathbf{x}_i)_{i=1}^n$ through a nonparametric regression model

$$y_i = \eta(\mathbf{x}_i) + e_i, \quad i = 1, \dots, n,$$

where $(e_i)_{i=1}^n$ are independent $N(0, \sigma^2)$ noise. The computer model $y^s(\cdot, \theta)$, also known as the simulator, is a function designed by scientific experts to model the unknown physical system $\eta(\cdot)$ when the calibration parameter θ is appropriately estimated.

Despite the success of computer models in many scientific studies, researchers often ask the following question: is the computer model a suitable surrogate for the real physical system? Compared to the physical systems, the traditional computer models are rarely perfect or exact due to their fixed parametric nature or simplifications of the complex physical phenomenon (Tuo and Wu 2015). This implies that there exists some discrepancy between a physical system $\eta(\cdot)$ and its corresponding computer model $y^s(\cdot, \theta)$ even if the computer model is well calibrated. Kennedy and O'Hagan (2001) first tackled this discrepancy issue under a Bayesian framework, which has been influential among many other statisticians and quality control engineers. For an incomplete list of references for the computer model calibration problem, we refer to Higdon et al. (2004), Bayarri et al. (2007), Qian and Wu (2008), Joseph and Melkote (2009), Wang, Chen, and Tsui (2009), Chang and Joseph (2014),

Brynjarsdóttir and O’Hagan (2014), Storlie et al. (2015), among others.

Theoretical properties of the calibration problem were not well understood until Tuo and Wu (2015, 2016), which pointed out that the calibrated computer models estimated by Kennedy and O’Hagan (2001) could lead to poor approximations to the physical systems. The identifiability issue of the calibration parameter in Kennedy and O’Hagan (2001) was also noticed by H. P. Wynn, among several other discussants, in the written discussion of Kennedy and O’Hagan (2001). In short, the identifiability issue refers to the phenomenon that the distribution of the observed data from a physical system does not uniquely determine the corresponding calibration parameter value given the computer model. There are several Bayesian approaches to tackle the identifiability issue. For example, Bayarri et al. (2007) suggested to incorporate the experts’ information of the calibration parameter θ into the prior distribution to reduce the confounding between θ and the model discrepancy caused by the nonidentifiability. Brynjarsdóttir and O’Hagan (2014) presented a concrete example in which the derivative information of the model discrepancy was incorporated through a constrained Gaussian process prior. These Bayesian approaches, however, lack theoretical guarantees and mathematical rigor. In contrast to Bayesian methods, which are traditionally applied to solve the calibration problem, Tuo and Wu (2015, 2016) and Wong, Storlie, and Lee (2017) addressed the identifiability issue rigorously in frequentist frameworks and provided the corresponding theoretical justifications.

We propose a Bayesian method for computer model calibration called the Bayesian projected calibration. To the best of our knowledge, our work is the first one in the literature that simultaneously achieves the following objectives:

- (a) **Model identifiability:** The proposed approach is formulated in a rigorously identifiable fashion. Tuo and Wu (2015, 2016) and Wong, Storlie, and Lee (2017) defined the “true” value of the calibration parameter to be the one that minimizes the L_2 -distance between the computer model $y^s(\cdot, \theta)$ and the physical system $\eta(\cdot)$. Following their work, the proposed Bayesian projected calibration provides a Bayesian method to estimate this “true” value of the calibration parameter consistently.
- (b) **Uncertainty quantification:** The Bayesian projected calibration serves as a natural way for the uncertainty quantification of the calibration parameter through its full posterior distribution. Tuo and Wu (2015) showed the asymptotic normality of the L_2 -projected calibration estimator for the uncertainty quantification of the calibration parameter, which may not work in practice because the amount of the physical data is usually very limited (Tuo 2017). Hence a Bayesian approach is desired, especially when the data are scarce.
- (c) **Theoretical guarantee:** We show that the full posterior distribution of the calibration parameter is asymptotically normal with the efficient covariance matrix. Earlier literature either only provide asymptotic results of specific point estimators (Tuo and Wu 2015, 2016; Tuo 2017; Wong, Storlie, and Lee 2017), or formulate a Bayesian methodology for calibration problems without large sample evaluation (Plumlee

2017). Our method represents the first effort in providing the theoretical guarantee for the full posterior distribution of a Bayesian method for the computer model calibration problem.

- (d) **Convenient computational algorithms:** We design two algorithms based on stochastic approximation to tackle the computation issue for obtaining the posterior distribution of the calibration parameter. Unlike the orthogonal Gaussian process approach for computer model calibration proposed in Plumlee (2017), which is typically computationally expensive, the proposed two algorithms are computationally cheaper. This is illustrated in Section 5. Furthermore, the theoretical properties of these algorithms, including the convergence analyses, are discussed, justifying their usefulness.

The rest of the article is organized as follows. In Section 2, we formulate the calibration problem rigorously in an identifiable fashion and introduce the Bayesian projected calibration method. Section 3 elaborates on the asymptotic properties of the posterior distribution of the calibration parameter. We discuss the computational strategies for computing the projected calibration and its approximation in Section 4, in which two algorithms based on stochastic approximation are designed with theoretical support. In Section 5, we demonstrate the advantages of the Bayesian projected calibration in terms of the estimation accuracy and the uncertainty quantification via simulation studies and two real-world data examples. Potential extensions beyond the current projected calibration framework are considered in Section 6, and we conclude the article with a discussion in Section 7.

2. Problem Formulation

2.1. Background

We first briefly review the frequentist L_2 -projected calibration approach proposed by Tuo and Wu (2015) before introducing the proposed Bayesian projected calibration method, which can be regarded as the Bayesian version of the L_2 -projected calibration.

Suppose one has collected responses $(y_i)_{i=1}^n$ from a physical system η on a set of design points $(\mathbf{x}_i)_{i=1}^n \subset \Omega \subset \mathbb{R}^p$, where $\eta : \Omega \rightarrow \mathbb{R}$ is a deterministic function, and the design space Ω is the closure of a connect bounded convex open set in $\mathbb{R}^p$. The physical responses $(y_i)_{i=1}^n$ are noisy due to measurement or observational errors, and are hence modeled by the following nonparametric regression model:

$$y_i = \eta(\mathbf{x}_i) + e_i, \quad i = 1, \dots, n, \quad (1)$$

where e_i ’s are independent $N(0, \sigma^2)$ noise. Such a model has been widely adopted in the literature of calibration (Kennedy and O’Hagan 2001; Tuo and Wu 2015; Tuo 2017; Wong, Storlie, and Lee 2017).

Let Θ be the parameter space of the calibration parameter θ . We assume that $\Theta \subset \mathbb{R}^q$ is compact. A computer model is a deterministic function $y^s : \Omega \times \Theta \rightarrow \mathbb{R}$ that produces an output $y^s(\mathbf{x}, \theta)$ given a controllable input $\mathbf{x} \in \Omega$ and the calibration parameter $\theta \in \Theta$. The goal of calibration is to estimate θ

given the computer model y^s and the physical data $(y_i)_{i=1}^n$, such that the calibrated computer model approximates the physical system well. However, as pointed out by Tuo and Wu (2016) and Wong, Storlie, and Lee (2017), the calibration parameter θ cannot be identified without further restriction, in the sense that θ cannot be uniquely determined by the distribution of the physical data $(\mathbf{x}_i, y_i)_{i=1}^n$. More precisely, by alternatively expressing the physical system η in terms of the computer model $y^s(\mathbf{x}, \theta)$ and a discrepancy $\delta(\mathbf{x})$ as the equation (Kennedy and O'Hagan 2001; Plumlee 2017; Tuo 2017; Wong, Storlie, and Lee 2017)

$$\eta(\mathbf{x}) = y^s(\mathbf{x}, \theta) + \delta(\mathbf{x}),$$

where the discrepancy function δ is completely nonparametric, it is clear that (θ, δ) cannot be uniquely identified by the physical system η . Therefore, the “true” value of the calibration parameter that gives rise to the physical data is not well-defined.

The computer model calibration problem was firstly studied in Kennedy and O'Hagan (2001) using a Bayesian approach by imposing a Gaussian process prior on the discrepancy function $\delta(\cdot)$. Although this approach did not address the identifiability issue directly, a related later work (Tuo and Wu 2016) tackled the identifiability issue by modifying the Kennedy and O'Hagan (abbreviated as KO) approach in a simplified setting. Specifically, if the discrepancy function follows a mean-zero Gaussian process prior with covariance function $\Psi(\cdot, \cdot)$, θ follows a uniform prior, and the physical data are noise-free (i.e., e_i 's are zeros), then the posterior density of θ is proportional to

$$\begin{aligned} \pi(\theta \mid (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)) \\ \propto \exp \left[-\frac{1}{2}(\mathbf{y} - \mathbf{y}_\theta^s)^T \Psi(\mathbf{x}_{1:n}, \mathbf{x}_{1:n})^{-1}(\mathbf{y} - \mathbf{y}_\theta^s) \right], \end{aligned}$$

where $\mathbf{y} = [y_1, \dots, y_n]^T$, $\mathbf{y}_\theta^s = [y^s(\mathbf{x}_1, \theta), \dots, y^s(\mathbf{x}_n, \theta)]^T$, and $\Psi(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) = [\Psi(\mathbf{x}_i, \mathbf{x}_j)]_{n \times n}$. Instead of using a fully Bayesian approach, Tuo and Wu (2016) argued for a simplified KO approach, which is to compute the maximum a posteriori estimator $\hat{\theta}$ that maximizes the above display. Under certain regularity conditions, Tuo and Wu (2016) proved that $\hat{\theta}$ converges to a point θ^* that minimizes the reproducing kernel Hilbert space norm of δ associated with the covariance function Ψ . Therefore, in this simplified KO approach, the “true” value of θ can be defined to be θ^* . However, when the physical data are noisy, such an approach is no longer valid for defining θ^* , and the resulting estimator $\hat{\theta}$ does not converge to the desired θ^* (Tuo and Wu 2015, 2016).

Alternatively, as pointed out in Tuo and Wu (2016, sec. 4.2), in comparison with the definition of θ^* in the simplified KO approach, it is also reasonable to define the “true” value of θ through the more straightforward L_2 -projection:

$$\begin{aligned} \theta^* &= \arg \min_{\theta \in \Theta} \|\eta(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2 \\ &= \arg \min_{\theta \in \Theta} \int_{\Omega} [\eta(\mathbf{x}) - y^s(\mathbf{x}, \theta)]^2 d\mathbf{x}. \end{aligned} \quad (2)$$

The L_2 -projected calibration method provides an estimator $\hat{\theta}_{L_2}$ for θ^* using a two-step procedure. First, an estimator $\hat{\eta}$ of the

physical system η is obtained via the *kernel ridge regression* (Wahba 1990) given the physical data $(\mathbf{x}_i, y_i)_{i=1}^n$:

$$\hat{\eta} = \arg \min_{f \in \mathbb{H}_{\Psi}(\Omega)} \frac{1}{n} \sum_{i=1}^n [y_i - f(\mathbf{x}_i)]^2 + \lambda \|f\|_{\mathbb{H}_{\Psi}(\Omega)}, \quad (3)$$

where $\Psi : \Omega \times \Omega \rightarrow \mathbb{R}$ is a positive definite covariance function, $\mathbb{H}_{\Psi}(\Omega)$ is the *reproducing kernel Hilbert space* (RKHS) associated with Ψ , and $\|\cdot\|_{\mathbb{H}_{\Psi}(\Omega)}$ is the native norm of $\mathbb{H}_{\Psi}(\Omega)$. We refer to Wahba (1990) and Wendland (2004) for detailed treatment of these concepts. Then, the L_2 -projected calibration estimator $\hat{\theta}_{L_2}$ for θ^* is given by

$$\hat{\theta}_{L_2} := \arg \min_{\theta \in \Theta} \|\hat{\eta}(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2. \quad (4)$$

The L_2 -projected calibration has very nice theoretical properties: $\hat{\theta}_{L_2}$ is not only $\sqrt{n}$ -consistent for θ^* , but is also semiparametric efficient (Tuo and Wu 2016). In other words, it provides an optimal estimator to the “true” calibration parameter. More importantly, compared to the simplified KO approach, the L_2 -calibration approach can directly deal with noisy physical data.

2.2. Bayesian Projected Calibration

The L_2 -projected calibration estimator $\hat{\theta}_{L_2}$ enjoys nice asymptotic properties. Nevertheless, it is a frequentist approach, the uncertainty quantification of which needs to be assessed via additional procedures, for example, the bootstrap (Wong, Storlie, and Lee 2017). A convenient way to incorporate the uncertainty is to develop a Bayesian model with carefully selected prior distributions, and assess the uncertainty via the posterior distribution of the parameter of interest. Although the Bayesian philosophy originates from the belief that the parameters of interest are random variables following certain prior distributions, Bayesian methods have been gaining popularity among frequentist community as well, as it can be regarded as a class of flexible approaches to estimate deterministic parameters. Furthermore, there has been a rapid development in the field of frequentist justification of Bayesian estimation for deterministic parameters in nonparametric and high-dimensional problems (Xie and Xu 2019, 2020). The readers are referred to Ghosal and van der Vaart (2017) for a thorough review of the related literature. The key to the success of Bayesian methods in estimating deterministic parameters lies in the delicate construction of an appropriate prior model. In what follows, we achieve this goal in the context of the computer model calibration problem by constructing a prior model.

We follow the definition of the “true” value θ^* of θ given in (2), as it minimizes the uncertainty beyond the computer model for explaining the physical system. There are two unknown parameters: the physical system η , taking values in some function space $\mathcal{F}$, and the calibration parameter $\theta \in \Theta$. The statistical model for calibration can be defined by

$$\begin{aligned} \mathcal{P} &= \left\{ \phi_{\sigma}(y - \eta(\mathbf{x})) : \eta \in \mathcal{F}, \theta^* \right. \\ &= \arg \min_{\theta \in \Theta} \|\eta(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2 \left. \right\}, \end{aligned}$$

where $\phi_\sigma(\cdot)$ is the density function of $N(0, \sigma^2)$. Namely, the parameter (η, θ^*) is constrained on a manifold in $\mathcal{F} \times \Theta$ defined by

$$\mathcal{M} = \left\{ (\eta, \theta^*) \in \mathcal{F} \times \Theta : \theta^* = \arg \min_{\theta \in \Theta} \|\eta(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2 \right\}. \quad (5)$$

We will rigorously show in Section 3 that under certain regularity conditions, $\mathcal{M}$ is a differentiable Banach manifold. It is therefore natural to treat the “true” calibration parameter θ^* as a functional $\theta^* : \mathcal{F} \rightarrow \Theta$, $\eta \mapsto \arg \min_{\theta} \|\eta(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2$, of a physical system η , and we denote this functional by θ_η^* . To distinguish the parameter (η, θ_η^*) and the truth that generates the data, we denote η_0 to be the true physical system producing physical data $(y_i)_{i=1}^n$, and $\theta_0^* = \theta_{\eta_0}^*$.

We now introduce the *Bayesian projected calibration*. The unknown physical process η is imposed a mean-zero Gaussian process prior $\Pi = \text{GP}(0, \tau^2 \Psi)$, where $\Psi : \Omega \times \Omega \rightarrow \mathbb{R}_+$ is a positive definite covariance function, and $\tau > 0$ is a scaling factor. We will discuss later regarding the choice of an appropriate covariance function Ψ . Let $\mathcal{D}_n$ denote the physical data $(\mathbf{x}_i, y_i)_{i=1}^n$, and $\Pi(\cdot | \mathcal{D}_n)$ denote the posterior distribution given $\mathcal{D}_n$. It is straightforward to show that the posterior distribution of η is also a Gaussian process with mean function $\tilde{\eta}$ and covariance function $\tilde{\Psi}$, where

$$\tilde{\eta}(\mathbf{x}) = \tau^2 \Psi(\mathbf{x}_{1:n}, \mathbf{x})^T (\tau^2 \Psi(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y}, \quad (6)$$

$$\begin{aligned} \tilde{\Psi}(\mathbf{x}, \mathbf{x}') &= \tau^2 \Psi(\mathbf{x}, \mathbf{x}') - \tau^2 \Psi(\mathbf{x}_{1:n}, \mathbf{x})^T (\tau^2 \Psi(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) \\ &\quad + \sigma^2 \mathbf{I}_n)^{-1} \tau^2 \Psi(\mathbf{x}_{1:n}, \mathbf{x}'). \end{aligned} \quad (7)$$

Here $\Psi(\mathbf{x}_{1:n}, \mathbf{x}) = [\Psi(\mathbf{x}_1, \mathbf{x}), \dots, \Psi(\mathbf{x}_n, \mathbf{x})]^T \in \mathbb{R}^n$, $\Psi(\mathbf{x}_{1:n}, \mathbf{x}_{1:n}) = [\Psi(\mathbf{x}_i, \mathbf{x}_j)]_{n \times n} \in \mathbb{R}^{n \times n}$, and $\mathbf{y} = [y_1, \dots, y_n]^T \in \mathbb{R}^n$. Note here the predictive mean $\tilde{\eta}(\mathbf{x})$ given physical data $\mathcal{D}_n$ coincides with the kernel ridge regression estimator $\hat{\eta}$ for some suitably chosen τ (see, e.g., Rasmussen and Williams 2006). The Gaussian process prior $\text{GP}(0, \tau^2 \Psi)$ on η naturally induces a prior distribution on θ_η^* through the constrained manifold $\mathcal{M}$ in (5). Therefore, after collecting noisy physical responses $(\mathbf{x}_i, y_i)_{i=1}^n$ from the regression model (1), one can obtain the posterior distribution of η , and hence, that of θ_η^* , given the physical responses and the computer model y^s . The procedure of computing the posterior distribution of θ_η^* is referred to as the Bayesian projected calibration. It can be regarded as a Bayesian version of the L_2 -projected calibration method, because the two approaches are to estimate the “true” value of θ over the constrained manifold $\mathcal{M}$ using a Bayesian and a frequentist approach, respectively. Furthermore, in Section 3 we will show that the posterior of θ_η^* in the Bayesian projected calibration is asymptotically centered at the L_2 -projected calibration estimator $\hat{\theta}_{L_2}$.

The choice of an appropriate covariance function Ψ for the Gaussian process prior is of fundamental importance in computer model calibration. One of the most popular choices is the class of the Matérn covariance functions. Formally, given $\alpha > p/2$, the Matérn covariance function with a smoothness

parameter α and a range parameter ψ is given by

$$\begin{aligned} \Psi_\alpha(\mathbf{x}, \mathbf{x}' | \psi) &= \frac{1}{\Gamma(\alpha) 2^{\alpha-1}} \left(\frac{\sqrt{2\alpha} \|\mathbf{x} - \mathbf{x}'\|}{\psi} \right)^\alpha K_\alpha \left(\frac{\sqrt{2\alpha} \|\mathbf{x} - \mathbf{x}'\|}{\psi} \right), \end{aligned} \quad (8)$$

where K_α is the modified Bessel function of the second kind. Throughout this work, we shall assume that all Gaussian processes involved are with the Matérn covariance function for the convenience of theoretical analyses. As will be seen in Section 3, when the smoothness parameter α matches the smoothness level of the underlying true physical system, the resulting convergence rate is minimax-optimal. In practice the Matérn covariance functions with smoothness parameters $\alpha = 3/2$ and $\alpha = 5/2$ are popular due to their closed-form expressions, but the practitioners are welcome to select other covariance functions based on their expert knowledge depending on the specific application domains.

Remark 1 (Expensive computer model). In the literature of computer experiments, it is common that the computer model y^s is not directly available to us or time-consuming to run, in which case the computer model can be only computed at the given design points. Thus, finding an emulator $\hat{y}^s$ for y^s using the computer outputs data at the given design points is needed. One first collects a set of data $(\mathbf{x}_j^s, \theta_j^s, y_j^s)_{j=1}^m$ from m runs of the computer model, where $y_j^s = y^s(\mathbf{x}_j^s, \theta_j^s)$ is the output at the design point $\mathbf{x}_j^s$, then build the emulator $\hat{y}^s$ using the data $(\mathbf{x}_j^s, \theta_j^s, y_j^s)_{j=1}^m$. There are varieties of methods for constructing emulators for computer experiments, including Gaussian process models (Santner, Williams, and Notz 2013), the radial basis function interpolation (Wendland 2004), the polynomial chaos approximation (Xiu 2010), and the smoothing spline ANOVA (Wahba 1990). To perform computer model calibration when the computer model y^s is not directly available or time-consuming to run, the calibration parameter θ_η^* can be estimated by replacing y^s with the corresponding emulator $\hat{y}^s$.

3. Theoretical Properties

In this section, we provide large sample justifications of the proposed Bayesian projected calibration by characterizing the asymptotic behavior of the posterior distribution $\Pi(\theta_\eta^* \in \cdot | \mathcal{D}_n)$. The posterior of θ_η^* has the similar behavior as the L_2 -projected calibration estimator $\hat{\theta}_{L_2}$: $\Pi(\theta_\eta^* \in \cdot | \mathcal{D}_n)$ is not only $\sqrt{n}$ -consistent, but also asymptotically normal with the efficient covariance matrix. The asymptotic normality of the full posterior distribution is also known as the Bernstein-von Mises (BvM) limit (see, e.g., van der Vaart 2000, chap. 10). The development of a semiparametric BvM theorem had not been established until Bickel and Kleijn (2012). For a thorough treatment of BvM limits of smooth functionals in semiparametric models in general, we refer to Castillo and Rousseau (2015a).

Before proceeding to the main results, we introduce some notations and definitions. Given an integer vector $\mathbf{k} = [k_1, \dots, k_p]^T$ and a function $f(x_1, \dots, x_p) : \Omega \rightarrow \mathbb{R}$, denote $D^{\mathbf{k}}$ to be the mixed partial derivative operator defined by $D^{\mathbf{k}} f = \partial^{|\mathbf{k}|} f / \partial x_1^{k_1} \dots \partial x_p^{k_p}$, where $|\mathbf{k}| := \sum_{j=1}^p k_j$. Let $\alpha > 0$ be positive,

and $\underline{\alpha}$ be the greatest integer strictly smaller than α . The α -Hölder norm of a function $f : \Omega \rightarrow \mathbb{R}$ is defined by

$$\|f\|_{\mathfrak{C}_\alpha(\Omega)} := \max_{\mathbf{k}: |\mathbf{k}| \leq \underline{\alpha}} \|D^{\mathbf{k}} f\|_{L_\infty(\Omega)} + \max_{\mathbf{k}: |\mathbf{k}| = \alpha} \sup_{\mathbf{x} \neq \mathbf{x}'} \frac{|D^{\mathbf{k}} f(\mathbf{x}) - D^{\mathbf{k}} f(\mathbf{x}')|}{\|\mathbf{x} - \mathbf{x}'\|^{\alpha - \underline{\alpha}}}.$$

The α -Hölder space of functions on Ω , denoted by $\mathfrak{C}_\alpha(\Omega)$, is the set of functions with finite α -Hölder norms. The α -Sobolev space of functions, denoted by $\mathcal{H}_\alpha(\Omega)$, is the set of functions $f : \Omega \rightarrow \mathbb{R}$ that can be extended to $\mathbb{R}^p$ such that the Fourier transforms $\widehat{f}(\mathbf{t}) = (2\pi)^{-p} \int_{\mathbb{R}^p} e^{it^\top \mathbf{x}} f(\mathbf{x}) d\mathbf{x}$ satisfy (van der Vaart and van Zanten 2011)

$$\int_{\mathbb{R}^p} (1 + \|\mathbf{t}\|^2)^\alpha |\widehat{f}(\mathbf{t})|^2 d\mathbf{t} < \infty.$$

To study the asymptotic behavior of $\Pi(\theta_\eta^* \in \cdot \mid \mathcal{D}_n)$, we first explore the convergence property of the physical system η . In practice it is common to encounter design points that are either randomly sampled or fixed a priori at certain prespecified locations, and the design space can be not as regular as a hypercube, but in this section, we shall assume for the ease the mathematical treatment that the design space Ω is the unit hypercube $[0, 1]^p$, and the design points $(\mathbf{x}_i)_{i=1}^n$ are independently and uniformly sampled from Ω . The theory developed here can be easily extended to the case where the design points are independently drawn from a distribution with a density that is bounded away from 0 and ∞ . The computer model y^s is assumed to be directly available to us or cheap to run. Such a simplified assumption is also adopted in Wong, Storlie, and Lee (2017) and it does no harm to the theoretical analyses, as the amount of data from the computer experiments is typically much larger than the sample size of the physical data. In addition, the computer data $(\mathbf{x}_j^s, \theta_j^s, y_j^s)_{j=1}^m$ are deterministic and the approximation error between y^s and $\widehat{y}^s$, when sufficiently small as m gets large, does not affect the stochastic analyses here. Therefore, one may assume that the error between $\widehat{y}^s$ and y^s is negligible. The true but unknown physical system η_0 is assumed to lie in the intersection of the α -Hölder space $\mathfrak{C}_\alpha(\Omega)$ and α -Sobolev space $\mathcal{H}_\alpha(\Omega)$ for some $\alpha > p/2$. We assume that the prior Π for η is the mean-zero Gaussian process $\text{GP}(0, \tau^2 \Psi_\alpha(\cdot, \cdot \mid \psi))$ and without loss of generality, the scaling factor τ is fixed at 1. We shall also assume that the range parameter ψ is fixed, as fixing the range parameter does not affect the asymptotic analyses of both η and θ . When $\psi = 1$, we use $\Psi_\alpha(\cdot, \cdot)$ to denote $\Psi_\alpha(\cdot, \cdot \mid \psi)$.

We now present the convergence result for η . In particular, the first result (9) in the following theorem directly follows from van der Vaart and van Zanten (2011, Theorem 5) and the proof of the second result is given in the supplementary materials.

Theorem 1 (Convergence of η). Suppose η is imposed a Gaussian process prior $\Pi = \text{GP}(0, \Psi_\alpha)$, and $\eta_0 \in \mathfrak{C}_\alpha(\Omega) \cap \mathcal{H}_\alpha(\Omega)$, where $\alpha > p/2$. Then for any sequence $M_n \rightarrow \infty$,

$$\mathbb{E}_0 \left[\Pi \left(\|\eta - \eta_0\|_{L_2(\Omega)} > M_n n^{-\alpha/(2\alpha+p)} \mid \mathcal{D}_n \right) \right] \rightarrow 0, \quad (9)$$

and there exists some constant $M > 0$ such that

$$\Pi \left(\|\eta - \eta_0\|_{L_\infty(\Omega)} > M \mid \mathcal{D}_n \right) \rightarrow 0$$

in $\mathbb{P}_0$ -probability.

The resulting rate $n^{-\alpha/(2\alpha+p)}$ is proven to be optimal when the underlying true function η_0 is an α -Hölder or α -Sobolev function (see, e.g., Stone 1982; van der Vaart and Wellner 1996; Ghosal and van der Vaart 2017).

We next discuss the property of θ_η^* as a functional: $\eta \mapsto \theta_\eta^*$. Under certain regularity conditions to be stated next, θ_η^* yields a first-order Taylor expansion with respect to η locally around η_0 . Such a smooth property of the functional θ_η^* serves as the building block to derive the asymptotic normality of $\Pi(\theta_\eta^* \in \cdot \mid \mathcal{D}_n)$.

A1 θ_η^* is the unique solution to (2) and is in the interior of Θ for η in an L_2 -neighborhood of η_0 .

A2 $\sup_{\theta \in \Theta} \|y^s(\cdot, \theta)\|_{L_2(\Omega)} < \infty$.

A3 The Hessian matrix

$$\mathbf{V}_\eta = \int_{\Omega} \left\{ \frac{\partial^2}{\partial \theta \partial \theta^\top} [\eta(\mathbf{x}) - y^s(\mathbf{x}, \theta)]^2 \right\} d\mathbf{x} \Big|_{\theta = \theta_\eta^*}$$

is strictly positive definite for all η in an L_2 -neighborhood of η_0 .

A4 For all $j, k = 1, \dots, q$, it holds that

$$\sup_{\theta \in \Theta} \left\| \frac{\partial y^s}{\partial \theta_j}(\cdot, \theta) \right\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)} < \infty, \quad \frac{\partial^2 y^s}{\partial \theta_j \partial \theta_k}(\cdot, \cdot) \in \mathfrak{C}_1(\Omega \times \Theta).$$

The proof of the following lemma is given in the supplementary materials.

Lemma 1 (Taylor expansion). Assume that $\eta_0 \in \mathfrak{C}_\alpha(\Omega) \cap \mathcal{H}_\alpha(\Omega)$ for some $\alpha > p/2$. Under conditions A1–A4, there exists some $\epsilon > 0$ and some positive constants $L_{\eta_0}^{(1)}$ and $L_{\eta_0}^{(2)}$ depending on η_0 only, such that $\|\theta_\eta^* - \theta_0^*\| \leq L_{\eta_0}^{(1)} \|\eta - \eta_0\|_{L_2(\Omega)}$ and

$$\left\| \theta_\eta^* - \theta_0^* - 2 \int_{\Omega} [\eta(\mathbf{x}) - \eta_0(\mathbf{x})] \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}, \theta_0^*) d\mathbf{x} \right\| \leq L_{\eta_0}^{(2)} \|\eta - \eta_0\|_{L_2(\Omega)}^2 \quad (10)$$

whenever $\|\eta - \eta_0\|_{L_2(\Omega)} < \epsilon$, where $\mathbf{V}_0 = \mathbf{V}_{\eta_0}$. Furthermore, if A1 and A3 hold for all η in an L_2 -neighborhood $\mathcal{U}$ of η_0 , then the set $\mathcal{M}(\mathcal{U}) := \{(\eta, \theta_\eta^*) : \eta \in \mathcal{U}\}$ forms a Banach manifold, and if $\mathcal{U}$ is the entire $L_2(\Omega)$ space, then $\mathcal{M}$ defined by (5) is a Banach manifold.

It follows immediately from the convergence results of the physical system η (Theorem 1) and the Taylor expansion property of θ_η^* (Lemma 1) that the posterior of θ_η^* is consistent.

Corollary 1 (Consistency of θ_η^*). Suppose η is imposed a Gaussian process prior $\Pi = \text{GP}(0, \Psi_\alpha)$, and $\eta_0 \in \mathfrak{C}_\alpha(\Omega) \cap \mathcal{H}_\alpha(\Omega)$. Then the posterior of θ_η^* is consistent, that is, $\Pi(\|\theta_\eta^* - \theta_0^*\| > \epsilon \mid \mathcal{D}_n) \rightarrow 0$ in $\mathbb{P}_0$ -probability for any $\epsilon > 0$.

Now we characterize the asymptotic behavior of $\Pi(\theta_\eta^* \in \cdot \mid \mathcal{D}_n)$, which is the main result of this article. Under certain regularity conditions, the posterior distribution of $\sqrt{n}(\theta_\eta^* - \widehat{\theta}_{L_2})$ is asymptotically normal, where $\widehat{\theta}_{L_2}$ is the frequentist L_2 -projected calibration estimator of θ proposed by Tuo and Wu (2015) (see Section 2). We describe the L_2 -projected calibration

procedure in our context for completeness:

$$\begin{aligned}\hat{\eta} &= \arg \min_{f \in \mathbb{H}_{\Psi_\nu}(\Omega)} \frac{1}{n} \sum_{i=1}^n [y_i - f(\mathbf{x}_i)]^2 + \lambda_n \|f\|_{\mathbb{H}_{\Psi_\nu}(\Omega)}^2, \\ \hat{\theta}_{L_2} &= \arg \min_{\theta \in \Theta} \|\hat{\eta}(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2,\end{aligned}$$

where $\nu = \alpha - p/2$, and $\lambda_n \asymp n^{-2\alpha/(2\alpha+p)}$ is a sequence depending on the sample size of the physical data $\mathcal{D}_n$.

The proof of the following theorem is deferred to the Appendix.

Theorem 2 (Asymptotic normality). Suppose η is imposed a Gaussian process prior $\Pi = \text{GP}(0, \Psi_\alpha)$, and $\eta_0 \in \mathcal{C}_\alpha(\Omega) \cap \mathcal{H}_\alpha(\Omega)$, where $\alpha > p/2$. Under conditions A1–A4, it holds that

$$\begin{aligned}\sup_A \left| \Pi \left(\sqrt{n}(\theta_\eta^* - \hat{\theta}) \in A \mid \mathcal{D}_n \right) - N(0, 4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1})(A) \right| \\ = o_{\mathbb{P}_0}(1),\end{aligned}$$

provided that

$$\mathbf{W} = \int_{\Omega} \left[\frac{\partial y^s}{\partial \theta}(\mathbf{x}, \theta_0^*) \frac{\partial y^s}{\partial \theta^T}(\mathbf{x}, \theta_0^*) \right] d\mathbf{x}$$

is strictly positive definite, where the supremum is taken over all measurable subsets in $\mathbb{R}^q$, and $\hat{\theta}$ is any estimator of θ satisfying

$$\hat{\theta} - \theta_0^* = 2\mathbf{V}_0^{-1} \left[\frac{1}{n} \sum_{i=1}^n e_i \frac{\partial y^s}{\partial \theta}(\mathbf{x}_i, \theta_0^*) \right] + o_{\mathbb{P}_0}(n^{-1/2}).$$

In particular, $\hat{\theta}$ can be taken as the L_2 -calibration estimator $\hat{\theta}_{L_2}$ if $\lambda_n \asymp n^{-2\alpha/(2\alpha+p)}$ and $\nu = \alpha - p/2$ are chosen in the kernel ridge regression (3).

Tuo and Wu (2015) proved that the L_2 -projected calibration estimator $\hat{\theta}_{L_2}$ is also asymptotically normal: $\sqrt{n}(\hat{\theta}_{L_2} - \theta_0^*) \xrightarrow{\mathcal{L}} N(0, 4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1})$. Furthermore, the covariance matrix $4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1}$ achieves the semiparametric efficiency in the sense that there does not exist a regular estimator with a smaller asymptotic covariance matrix (in spectra). The posterior of θ_η^* possesses a similar optimal behavior as the L_2 -calibration in the following two senses: First, $\Pi(\theta_\eta^* \in \cdot \mid \mathcal{D}_n)$ is a posteriori consistent and $\hat{\theta}_{L_2}$ is consistent for θ_0^* in the frequentist sense, and the covariance matrix of the asymptotic posterior of $\sqrt{n}(\theta_\eta^* - \hat{\theta}_{L_2})$ coincides with the asymptotic covariance matrix of $\hat{\theta}_{L_2}$. Second, the following corollary unveils that the coordinate-wise posterior median of θ_η^* , as a Bayes estimator given rise by the full posterior distribution, is asymptotically equivalent to $\hat{\theta}_{L_2}$ in the coordinate-wise sense. The proof of the following corollary is given in the supplementary materials.

Corollary 2. Suppose η is imposed a Gaussian process prior $\Pi = \text{GP}(0, \Psi_\alpha)$, and $\eta_0 \in \mathcal{C}_\alpha(\Omega) \cap \mathcal{H}_\alpha(\Omega)$, where $\alpha > p/2$. Let $\hat{\theta}^*$ be the coordinate-wise marginal posterior median of θ_η^* . Then under the conditions of Theorem 2, for each $k = 1, \dots, q$,

$$\sqrt{n} [\hat{\theta}^* - \theta_0^*]_k \xrightarrow{\mathcal{L}} N(0, 4\sigma^2 [\mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1}]_{kk}),$$

where $[\cdot]_k$ is the k th component of the argument vector and $[\cdot]_{kk}$ is the (k, k) th element of the argument matrix.

We finish this section with the following $\sqrt{n}$ -consistency result of θ_η^* , which is a refinement of Corollary 1. It is a consequence of Theorem 2 and the asymptotic normality of $\hat{\theta}_{L_2}$.

Corollary 3 ($\sqrt{n}$ -Consistency of θ_η^*). Suppose η is imposed a Gaussian process prior $\Pi = \text{GP}(0, \Psi_\alpha)$, and $\eta_0 \in \mathcal{C}_\alpha(\Omega) \cap \mathcal{H}_\alpha(\Omega)$, where $\alpha > p/2$. Under the conditions of Theorem 2, the posterior of θ_η^* is $\sqrt{n}$ -consistent, that is, for any sequence $M_n \rightarrow \infty$, it holds that $\mathbb{E}_0 \left[\Pi \left(\sqrt{n} \|\theta_\eta^* - \theta_0^*\| > M_n \mid \mathcal{D}_n \right) \right] \rightarrow 0$.

4. Computation Strategies

As seen in Section 3, the Bayesian projected calibration enjoys nice theoretical properties. In this section, we develop the computation strategies for obtaining the posterior distribution of θ_η^* given the physical data $\mathcal{D}_n = (\mathbf{x}_i, y_i)_{i=1}^n$.

In principle, computing the posterior distribution of θ_η^* is quite straightforward: To obtain T independent samples from $\Pi(\theta_\eta^* \in \cdot \mid \mathcal{D}_n)$, one first draw T independent sample paths $\eta^{(1)}, \dots, \eta^{(T)}$ from the posterior distribution of η from $\text{GP}(\tilde{\eta}, \tilde{\Psi})$ using formula (6) and (7), and then for each $\eta^{(t)}$, $t = 1, \dots, T$, compute the minimizer $\theta_{\eta^{(t)}}^* = \arg \min_{\theta} \|y^s(\cdot, \theta) - \eta^{(t)}(\cdot)\|_{L_2(\Omega)}^2$. Although drawing sample paths from the posterior distribution η is easy, it is, however, nontrivial to compute the corresponding $\theta_{\eta^{(t)}}^*$ s due to the generally intractable integral $\|y^s(\cdot, \theta) - \eta(\cdot)\|_{L_2(\Omega)}^2$ for any sample path η . A natural but naive strategy is to discretize the integral by the Monte Carlo approximation. Namely, one first draw N independent samples $\mathbf{x}_1^d, \dots, \mathbf{x}_N^d$ uniformly from Ω , then approximately compute θ_η^* for a sample path η by minimizing the discretized integral:

$$\theta_\eta^* \approx \arg \min_{\theta \in \Theta} \frac{1}{N} \sum_{j=1}^N \left[y^s(\mathbf{x}_j^d, \theta) - \eta(\mathbf{x}_j) \right]^2. \quad (11)$$

This strategy becomes accurate as $N \rightarrow \infty$ by the law of large numbers, but is not recommended in practice for the following reason: The discretized integral is still a function of θ , the minimizer of which is typically not in a closed-form except in rare cases, and finding the minimizer often requires the use of iterative optimization algorithms. Assuming that at least R iterations are needed to obtain $\theta_{\eta^{(t)}}^*$ for each $\eta^{(t)}$, $t = 1, \dots, T$, we see that the total complexity of the computation procedure becomes $O(NTR)$. To ensure the quality of the discretization approximation, N is typically made sufficiently large, especially when Ω is multidimensional. We will see next that the computation burden can be reduced by stochastic approximation methods.

4.1. Stochastic Approximation for the Projected Calibration

We first briefly introduce the basic idea of stochastic approximation before applying it to the projected calibration procedure. Stochastic approximation methods can be dated back to Robbins and Monro (1951) and have been gaining enormous progress in the recent decade thanks to the emergence of big data problems and the rapid development of advanced machine

learning techniques. This collection of methods focuses on minimizing objective functions $f(\theta)$ that can be written as an expected value $f(\theta) = \mathbb{E}_{\mathbf{w}}[g(\mathbf{w}, \theta)]$, where $\mathbf{w}$ is a random variable following a distribution $p(\mathbf{w})$. The major difference between stochastic approximation and its deterministic counterpart is that rather than observing f directly, one only has access to the noisy version $g(\mathbf{w}, \cdot)$. The basic idea of stochastic approximation is to generate a sequence of iterates $\theta^{(1)}, \theta^{(2)}, \dots$, using the stochastic gradient descent (SGD) method as follows: Let $\theta^{(t)}$ be the updated value of θ in the t th iteration. Then $\theta^{(t+1)}$ is updated by $\theta^{(t+1)} = \theta^{(t)} - \alpha_t \nabla_{\theta} g(\mathbf{w}_t, \theta^{(t)})$, where $(\alpha_t)_{t \geq 1} \subset (0, \infty)$ is a sequence of suitable step sizes and $(\mathbf{w}_t)_{t \geq 1}$ are independent copies of $\mathbf{w} \sim p(\mathbf{w})$. There is vast literature discussing the choice of the step sizes $(\alpha_t)_{t \geq 1}$ for convex and nonconvex f , among which the *AdaGrad* method (Duchi, Hazan, and Singer 2011) is one of the most popular ones. In the form of Li and Orabona (2018), the authors proposed to use the following form of the coordinate-wise step sizes:

$$\alpha_{tk} = a_0 \left\{ b_0 + \sum_{j=1}^{t-1} \left[\frac{\partial g(\mathbf{w}_j, \theta^{(j)})}{\partial \theta_k} \right]^2 \right\}^{-(1/2+\epsilon)}, \quad k = 1, \dots, q, \quad (12)$$

where $a_0, b_0 > 0$, $\epsilon \in (0, 1/2]$ are some constants, and then update $\theta^{(t+1)}$ by

$$\theta^{(t+1)} = \theta^{(t)} - \text{diag}(\alpha_{t1}, \dots, \alpha_{tq}) \frac{\partial g}{\partial \theta}(\mathbf{w}_t, \theta^{(t)}). \quad (13)$$

Convergence of AdaGrad for convex and nonconvex f was addressed in Li and Orabona (2018). In what follows, we modify the AdaGrad method for the projected calibration to reduce the aforementioned computation burden.

Recall that in the projected calibration procedure, we are interested in computing $\theta_{\eta}^* = \arg \min_{\theta \in \Theta} \|y^s(\cdot, \theta) - \eta(\cdot)\|_{L_2(\Omega)}^2$ given a sample path η drawn from the posterior distribution $\Pi(\eta \in \cdot \mid \mathcal{D}_n)$. Denote the integral $f_{\eta}(\theta) = \|y^s(\cdot, \theta) - \eta(\cdot)\|_{L_2(\Omega)}^2$. Clearly, by introducing a uniform random variable $\mathbf{w} \sim \text{Unif}(\Omega)$, $f_{\eta}(\theta) = \mathbb{E}_{\mathbf{w}}\{[y^s(\mathbf{w}, \theta) - \eta(\mathbf{w})]^2\}$ can be expressed as the expected value of a function of $\mathbf{w}$. Note that the parameter space Θ for the calibration parameter is compact, and AdaGrad needs to be modified to avoid searching out of the boundary of Θ . In the current context, whenever the updated $\theta^{(t+1)}$ strays outside the parameter space, we repeatedly take step-halving procedures until it falls back to Θ . We formally organize the modified AdaGrad for the projected calibration in Algorithm 1.

We see that by calling Algorithm 1 for computing $\theta_{\eta^{(t)}}^*$ for $t = 1, \dots, T$ rather than repeatedly optimizing the discretized integral in (11), the computation complexity is reduced to $O(NT)$. Furthermore, the convergence to a stationary point can be guaranteed by the following theorem, the proof of which is given in the supplementary materials. Although it is challenging to provide a theory for finding the global minimizer of the nonconvex objective functions involved, this can be typically addressed by trying multiple starting points in practice.

Theorem 3. Assume that the sample path η is continuous over Ω . Then under conditions A2 and A4, the output $\theta^{(N)}$ of Algorithm 1 converges to a stationary point of $f_{\eta}(\theta)$ as $N \rightarrow \infty$ a.s. with respect to the distribution of $(\mathbf{w}_t)_{t \geq 0}$.

Algorithm 1 Modified AdaGrad for the projected calibration

- 1: **Input:** Computer model $y^s(\cdot, \cdot)$ and its derivative $\nabla_{\theta} y^s(\cdot, \cdot)$;
Sample path $\eta(\cdot)$;
 - 2: **Initialize:** Initialize $\theta^{(1)} \sim \text{Unif}(\Theta)$; Set N to be number of samples from Ω ;
 - 3: **For** $t = 1 : (N - 1)$
 - 4: **Draw** $\mathbf{w}_t \sim \text{Unif}(\Omega)$;
 - 5: **For** $k = 1 : q$
 - 6: **Compute** α_{tk} using formula (12) with $g(\mathbf{w}, \theta) = [\eta(\mathbf{w}) - y^s(\mathbf{w}, \theta)]^2$;
 - 7: **End For**
 - 8: **Compute**

$$\theta^{(t+1)} = \theta^{(t)} - 2[y^s(\mathbf{w}_t, \theta^{(t)}) - \eta(\mathbf{w}_t)] \text{diag}(\alpha_{t1}, \dots, \alpha_{tq}) \frac{\partial y^s}{\partial \theta}(\mathbf{w}_t, \theta^{(t)});$$
 - 9: **If** $\theta^{(t+1)} \notin \Theta \setminus \partial\Theta$, then set $\alpha_{tk} \leftarrow \alpha_{tk}/2$ for $k = 1, \dots, q$ and go to line 8;
 - 10: **End For**
 - 11: **Output:** the last iterate $\theta^{(N)}$.
-

4.2. Approximate Computation of the Projected Calibration

We have seen in Section 4.1 that by borrowing the idea of stochastic approximation and AdaGrad, the computation complexity of computing the projected calibration can be reduced from $O(NTR)$ to $O(NT)$, where N is the number of samples required for discretization of the integral or AdaGrad, T is the number of independent copies drawn from the posterior of θ_{η}^* , and R is the number of iterations required to find θ_{η}^* given η using any iterative optimization algorithm. Although Algorithm 1 adopts certain stochastic approximation techniques, the resulting samples of $\theta_{\eta^{(t)}}^*$'s are drawn exactly from the posterior distribution of θ_{η}^* for any sample size n as $N \rightarrow \infty$ (recall that N can be made arbitrarily large). In contrast, in this subsection we seek an approximate computation method that can further reduce the computation cost of the projected calibration for a relatively large sample size.

The major computation bottleneck in finding θ_{η}^* by either minimizing the discretized integral (11), which requires $O(NR)$ operations, or AdaGrad, which needs $O(N)$ operations, is that there does not exist a closed-form formula to exactly compute θ_{η}^* using y^s and η . It is, however, feasible to approximate θ_{η}^* in certain ways. Recall that by Lemma 1 θ_{η}^* can be linearly approximated by a Taylor's expansion locally around η_0 :

$$\theta_{\eta}^* = \theta_0^* + 2 \int_{\Omega} [\eta(\mathbf{x}) - \eta_0(\mathbf{x})] \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}, \theta_0^*) d\mathbf{x} + \text{remainder},$$

where the remainder term is of the order $O(\|\eta - \eta_0\|_{L_2(\Omega)}^2)$. Since η_0 is unknown, a kernel ridge regression estimator $\hat{\eta}$, in the form of Section 3 (i.e., Ψ is the Matérn covariance function Ψ_{ν} with the smoothness parameter $\nu = \alpha - p/2$, and $\lambda_n \asymp n^{-2\alpha/(2\alpha+p)}$, where we assume that $\eta_0 \in \mathcal{C}_{\alpha}(\Omega) \cap \mathcal{H}_{\alpha}(\Omega)$), can be applied in place of η_0 to estimate θ_0^* and $\mathbf{V}_0$. Consequently, we have the

following approximation:

$$\theta_\eta^* \approx \tilde{\theta}_\eta := \hat{\theta}_{L_2} + 2 \int_{\Omega} [\eta(\mathbf{x}) - \hat{\eta}(\mathbf{x})] \mathbf{V}_{\hat{\eta}}^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}, \hat{\theta}_{L_2}) d\mathbf{x}. \quad (14)$$

The above approximation can be treated as a Taylor's expansion of θ_η^* locally around $\hat{\eta}$ alternatively. As illustrated later this section, computing the posterior distribution of $\tilde{\theta}_\eta$ is much more convenient than computing that of θ_η^* . More importantly, $\tilde{\theta}_\eta$, as a cheap approximation to θ_η^* , is asymptotically equivalent to θ_η^* for a large sample size. This can be formally put in the following theorem, which is proved in the supplementary materials:

Theorem 4. Assume the conditions in Theorem 2 hold, and $\tilde{\theta}_\eta$ is computed using formula (14). Then it holds that

$$\sup_A \left| \Pi \left(\sqrt{n}(\tilde{\theta}_\eta - \hat{\theta}_{L_2}) \in A \mid \mathcal{D}_n \right) - \mathcal{N} \left(\mathbf{0}, 4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1} \right) (A) \right| = o_{\mathbb{P}_0}(1),$$

where $\mathbf{W}$ is given in Theorem 2.

Now we discuss the practical consideration of computing the posterior distribution of $\tilde{\theta}_\eta$. Thanks to the closed-form expression (14), we can compute $\tilde{\theta}_\eta$ by discretizing the integral involved using N independent uniform samples $\mathbf{x}_1^d, \dots, \mathbf{x}_N^d$ from Ω :

$$\begin{aligned} \tilde{\theta}_\eta \approx \tilde{\theta}_\eta^{(N)} &:= \hat{\theta}_{L_2} - \frac{2}{N} \sum_{j=1}^N \hat{\eta}(\mathbf{x}_j^d) \mathbf{V}_{\hat{\eta}}^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}_j^d, \hat{\theta}_{L_2}) \\ &+ \frac{2}{N} \sum_{j=1}^N \eta(\mathbf{x}_j^d) \mathbf{V}_{\hat{\eta}}^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}_j^d, \hat{\theta}_{L_2}). \end{aligned}$$

Note that the posterior distribution of η is $\text{GP}(\tilde{\eta}, \tilde{\Psi})$ by formulas (6) and (7), it follows that a posteriori, $\tilde{\theta}_\eta^{(N)}$ follows a normal distribution with mean

$$\hat{\theta}_{L_2} + \frac{2}{N} \sum_{j=1}^N \left[\tilde{\eta}(\mathbf{x}_j^d) - \hat{\eta}(\mathbf{x}_j^d) \right] \mathbf{V}_{\hat{\eta}}^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}_j^d, \hat{\theta}_{L_2}) \quad (15)$$

and covariance matrix

$$\frac{4}{N^2} \sum_{j=1}^N \sum_{\ell=1}^N \mathbf{V}_{\hat{\eta}}^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}_j^d, \hat{\theta}_{L_2}) \tilde{\Psi}(\mathbf{x}_j^d, \mathbf{x}_\ell^d) \mathbf{V}_{\hat{\eta}}^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}_\ell^d, \hat{\theta}_{L_2}). \quad (16)$$

The complete algorithm of computing the approximate projected calibration is detailed in Algorithm 2. It can be clearly seen that this procedure further reduces the overall computation complexity from $O(NT)$ to $O(N)$. In the numerical studies, we find that Algorithm 2 provides a satisfactory approximation to the exact posterior when $n \geq 30$.

5. Numerical Examples

This section provides numerical examples to evaluate the proposed Bayesian projected calibration. Section 5.1 presents simulation studies via three synthetic examples. Two real-world data examples are included in Sections 5.2 and 5.3, respectively.

Algorithm 2 Approximate computation for the projected calibration

- 1: **Input:** Computer model $y^s(\cdot, \cdot)$ and its derivative $\nabla_{\theta} y^s(\cdot, \cdot)$; Physical data $(\mathbf{x}_i, y_i)_{i=1}^n$;
- 2: **Compute** the kernel ridge regression estimator $\hat{\eta}$;
- 3: **Call** Algorithm 1 with input y^s , $\nabla_{\theta} y^s(\cdot, \cdot)$, and $\hat{\eta}$ and output $\hat{\theta}_{L_2}$;
- 4: **Generate** N independent samples $\mathbf{x}_1^d, \dots, \mathbf{x}_N^d$ uniformly from Ω ;
- 5: **Compute** the mean vector $\hat{\theta}_{\text{APC}}$ using formula (15);
- 6: **Compute** the covariance matrix $\hat{\Sigma}_{\text{APC}}$ using formula (16);
- 7: **Output:** $\hat{\theta}_{\text{APC}}$ and $\hat{\Sigma}_{\text{APC}}$.

5.1. Simulation Studies

For simulated examples, we consider three configurations.

- Configuration 1. The computer model is

$$y^s(x, \theta) = 7[\sin(2\pi\theta_1 - \pi)]^2 + 2[(2\pi\theta_2 - \pi)^2 \sin(2\pi x - \pi)],$$

and the physical system coincides with the computer model when $\theta_0^* = [0.2, 0.3]^T$, that is, $\eta_0(x) = y^s(x, \theta_0^*)$. The design space Ω is $[0, 1]$, and the parameter space Θ for θ is $[0, 0.25] \times [0, 0.5]$. We simulate $n = 50$ observations from the randomly perturbed physical system $y_i = \eta_0(x_i) + e_i$, where $(x_i)_{i=1}^n$ are uniformly sampled from Ω , and the variance for the noise $(e_i)_{i=1}^n$ is set to 0.2^2 .

- Configuration 2. We follow an example provided in Gu and Wang (2017). The computer model is $y^s(x, \theta) = \sin(5\theta x) + 5x$, and the physical system is $\eta_0(x) = 5x \cos(15x/2) + 5x$. The design space Ω is $[0, 1]$, and the parameter space Θ for θ is $[0, 3]$. We simulate $n = 30$ observations from $y_i = \eta_0(x_i) + e_i$ with $\text{var}(e_i) = 0.2^2$, and $(x_i)_{i=1}^n$ are equidistant on Ω . The L_2 -discrepancy $\|\eta_0(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}$ between the computer model y^s and the physical system η_0 as a function of θ is depicted in Figure 1. The minimizer of the L_2 -discrepancy is $\theta_0^* = 1.8771$.
- Configuration 3. We use the pedagogical example in Plumlee (2017). The physical system is $\eta_0(x) = 4x + x \sin(5x)$ and the computer model is $y^s(x, \theta) = \theta x$, where $x \in \Omega = [0, 1]$ and $\theta \in \Theta = [2, 4]$. We take $(x_i)_{i=1}^n = \{0, 0.05, 0.1, 0.15, 0.2, \dots, 0.8\}$, and the responses are given by $y_i = \eta_0(x_i) + e_i$ with $\text{var}(e_i) = 0.02^2$. The L_2 -discrepancy as a function of θ is given by

$$\begin{aligned} \|\eta(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)} \\ = \sqrt{0.33(4 - \theta)^2 - 0.2898(4 - \theta) + 0.201714}, \end{aligned}$$

and is minimized at $\theta_0^* = 3.5609$.

For the three configurations described above, we impose the Matérn Gaussian process prior $\text{GP}(0, \tau^2 \Psi_{\alpha})$ on η , where Ψ_{α} is the Matérn covariance function given by (8) with $\alpha = 5/2$. Here the scaling factor τ is set to $\tau = 1$ for all 3 configurations for the ease of implementation. To draw posterior samples of θ_η^* , we first draw posterior samples of η using formula (6) and (7), then compute θ_η^* by $\theta_\eta^* = \arg \min_{\theta} \|\eta(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2$ using Algorithm 1. For all three configurations, 1000 samples of

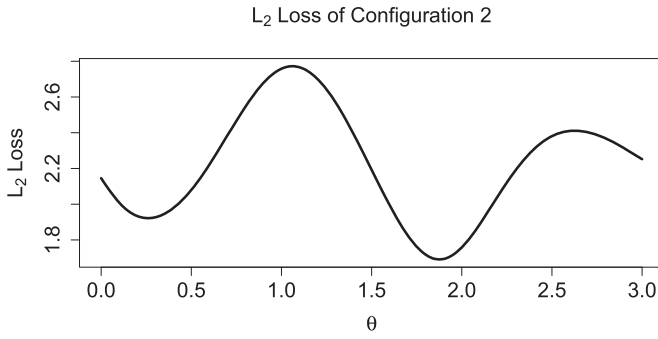


Figure 1. The L_2 -discrepancy $\|\eta_0(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}$ between the computer model y^s and the physical system η_0 as a function of θ for configuration 2.

θ_η^* are drawn from the posterior distribution for the subsequent analysis and the number of random samples N for AdaGrad in Algorithm 1 is set to 2000. For both configurations 1 and 2, we also draw 1000 samples using Algorithm 2.

For comparison, we implement the calibration method by Kennedy and O’Hagan (2001) (abbreviated as KO) and the orthogonal Gaussian process method by Plumlee (2017) (abbreviated as OGP). For the KO calibration approach, we follow the suggestion of van der Vaart and van Zanten (2009), let the range parameter ψ follow an inverse-Gamma prior distribution $\pi(\psi) \propto \psi^{-a_\psi-1} \exp(-b_\psi/\psi)$ for some $a_\psi, b_\psi > 0$, and set $a_\psi = b_\psi = 2$ in all numerical examples. For both the KO method and the OGP method in all three configurations, Markov chain Monte Carlo samplers are implemented to draw 1000 posterior samples after discarding 1000 burn-in samples.

For configuration 1, the summary statistics of the posterior distribution of θ_η^* are provided in Table 1, together with those obtained via Algorithm 2, the KO method, and the OGP method. We can see that the Bayesian projected calibration, the approximate projected calibration, and the OGP method all outperform the KO method in terms of both the accuracy of the point estimators (the posterior means) and the uncertainty quantification (the lengths of credible intervals and the standard deviations of posterior samples). Although the OGP method provides the slightly better posterior inference result compared to the Bayesian projected calibration, the computation runtime is significantly longer than the other two methods. The computational bottleneck of the OGP method was also mentioned in Plumlee (2017, sec. 6). Figure 2(a) presents the scatterplot of the posterior samples of $\sqrt{n}(\theta_\eta^* - \hat{\theta}_{L_2})$. The level curves of the BvM limit shows that the asymptotic distribution of $\Pi(\sqrt{n}(\theta_\eta^* - \hat{\theta}_{L_2}) | \mathcal{D}_n)$ developed in Section 3 offers a decent approximation to the exact posterior. Figure 2(b) presents the scatterplot

of $\sqrt{n}(\theta_\eta^* - \hat{\theta}_{L_2})$ against the level curves of the approximate projected calibration density, showing that the approximation is accurate. We provide the trace plot of the loss function $f_\eta(\theta) = \|\hat{\eta}(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2$ and the trajectory of the calibration parameter θ in Figure 3 to demonstrate the convergence behavior of the modified AdaGrad in Algorithm 2 for computing the approximate projected calibration. Comparing Figures 2(a) and (c), we see that the Bayesian projected calibration outperforms the KO in terms of the uncertainty quantification. We also investigate the performance of the calibrated computer model in Figures 2(d) and (e). The point-wise 95%-credible bands for the computer model also validate that the Bayesian projected calibration provides a better estimate to y^s in contrast to the KO approach.

Similarly, for configuration 2, the advantages of Bayesian/approximate projected calibration in terms of the uncertainty quantification and the computational cost are summarized from the statistics reported in Table 2. It can be seen that the Bayesian/approximate projected calibration produce smaller uncertainty compared to the KO calibration. We also provide the histogram of the projected calibration and the density of the approximate projected calibration (blue curve) in Figure 4(a), showing that the approximate projected calibration density provides a satisfactory approximation to the exact posterior. Furthermore, the red curve in Figure 4(a) shows that the asymptotic BvM limit approximates the exact posterior well even though the sample size is only $n = 30$. The convergence of the modified AdaGrad in Algorithm 1 can be assessed via the trace plot of the loss function $f_\eta(\theta)$ in Figure 4(b).

Configuration 3 is slightly challenging due to the fact that no physical data are available in $(0.8, 1]$, and the physical data are relatively sparse (see Figure 5). In such a scenario, we do not recommend using Algorithm 2 for an approximate posterior inference. We provide the corresponding summary statistics for the Bayesian projected calibration, the KO method, and the OGP method, in Table 3. It can be seen that when the design

Table 2. Summary statistics of the posterior of θ for configuration 2 (simulation truth is $\theta_0^* = 1.8771$); projected refers to the projected calibration, and approximate refers to the approximate projected calibration using Algorithm 2.

	Projected	KO	OGP	Approximate
Mean	1.8816	1.8805	1.8825	1.8822
Standard deviation	0.0047	0.0661	0.0023	0.0047
97.5%-Quantile	1.8907	2.0089	1.8766	1.8915
2.5%-Quantile	1.8725	1.7480	1.8678	1.8731
Runtime	237.289 sec	1.034 sec	31,843 sec	6.269 sec

Table 1. Summary statistics of the posterior of θ for configuration 1 (the simulation truth is $\theta_0^* = [0.2, 0.3]^T$); projected refers to the projected calibration, and approximate refers to the approximate projected calibration using Algorithm 2.

θ	Projected		KO		OGP		Approximate	
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
Mean	0.1984	0.3009	0.1934	0.2988	0.2068	0.3024	0.1986	0.3004
Standard deviation	0.0011	0.0013	0.0269	0.0025	0.0005	0.0006	0.0011	0.0013
97.5%-Quantile	0.2006	0.3034	0.2439	0.3182	0.2013	0.2999	0.2007	0.3029
2.5%-Quantile	0.1963	0.2984	0.1445	0.2938	0.1992	0.2975	0.1965	0.2979
Runtime	279 sec		0.834 sec		40,562 sec		7.365 sec	

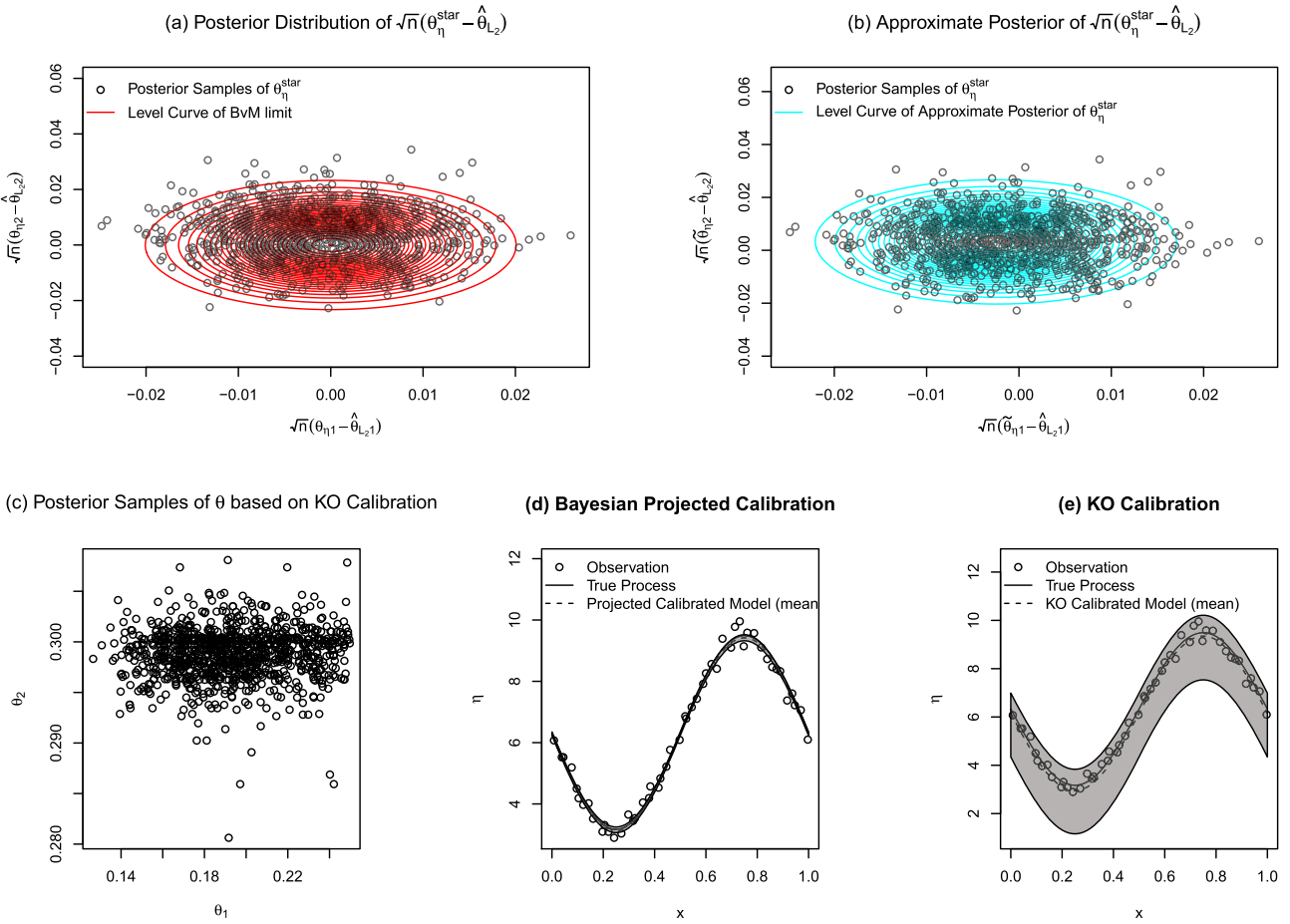


Figure 2. Visualization of the posterior inference for configuration 1 in the simulation studies. Panels (a) and (b) show the scatterplot of the posterior samples of $\sqrt{n}(\theta_\eta^* - \hat{\theta}_{L_2})$ and the level curves of the corresponding BvM limit / approximate projected calibration density from Algorithm 2. Panel (c) presents the scatterplot of the posterior samples of θ using the KO approach. Panels (d) and (e) display the calibrated computer models (in dashed lines) using the Bayesian projected calibration and the KO approach, respectively, together with their corresponding point-wise 95%-credible intervals (in shaded area).

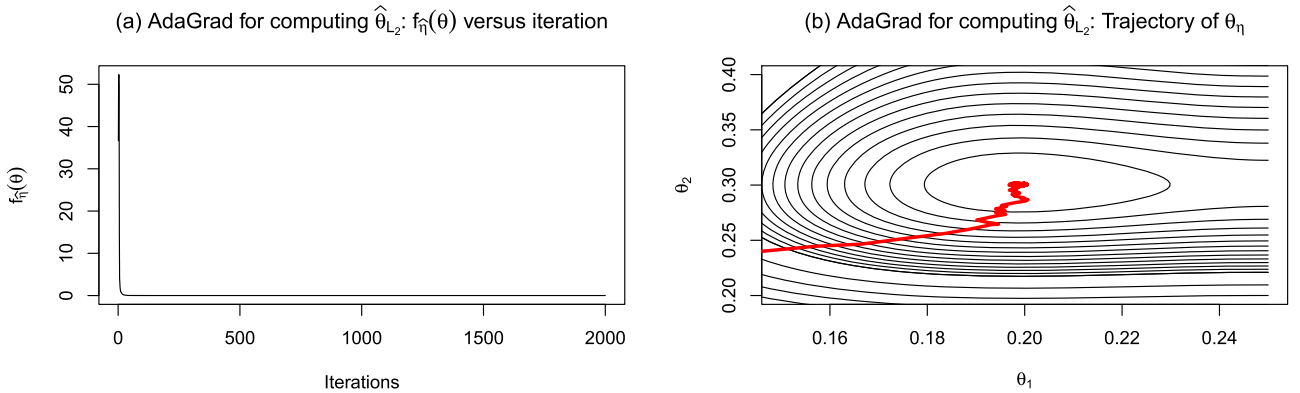


Figure 3. Convergence behavior of the modified AdaGrad for computing $\hat{\theta}_{L_2}$ for configuration 1 in the simulation studies.

points are not regularly spread over Ω , the KO method results in larger uncertainty when estimating θ compared to the Bayesian projected calibration and the OGP method. Note that it is unfair to compare the point estimator of the KO method with those of the other two competitors, since the “true” values of θ are different according to their respective formulations. For the uncertainty quantification performance measured by the widths of credible intervals and standard deviations, the OGP method and the Bayesian projected calibration are similar, and both

outperform the KO approach. The calibrated computer models are visualized in Figure 5.

5.2. Ion Channel Example

We apply the Bayesian projected calibration to the ion channel example used in Plumlee, Joseph, and Yang (2016). The dataset involves measurements from experiments concerning ion channels of cardiac cells. Specifically, the output of the experiment

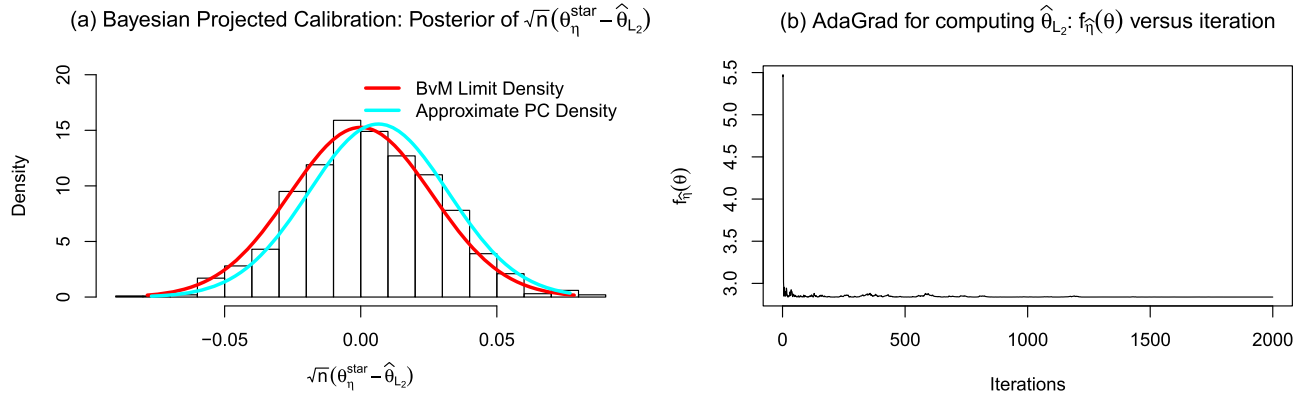


Figure 4. Simulation study configuration 2: the left panel is the histogram of the posterior samples of $\sqrt{n}(\theta_\eta^* - \hat{\theta}_{L_2})$, together with the theoretical BvM limit density (red solid line) and the approximate projected calibration density (cyan solid line); the right panel is the trace plot of the loss function $f_{\hat{\eta}}(\theta)$ values of Algorithm 1 along the iterations.

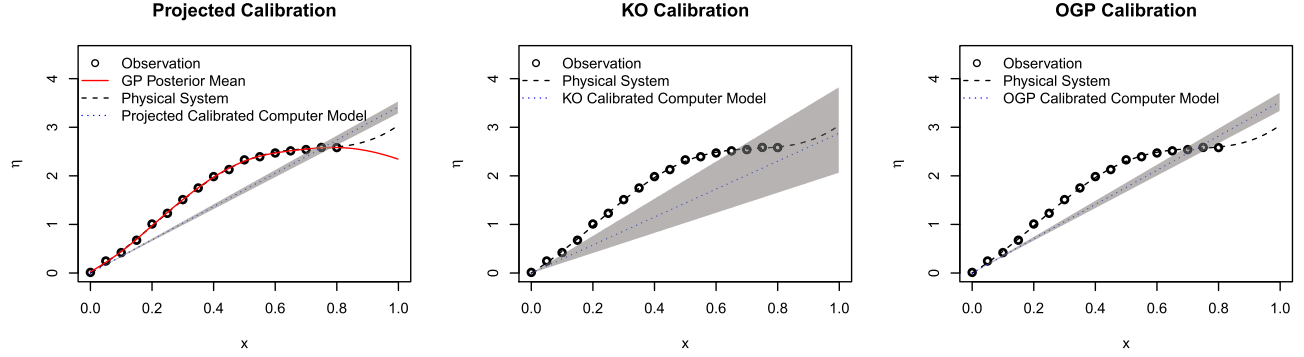


Figure 5. Visualization of the posterior inference for configuration 3 in the simulation studies. The three panels show the calibrated computer models (in dotted lines) using the Bayesian projected calibration, the KO calibration method, and the OGP calibration method, respectively, together with their corresponding point-wise 95%-credible intervals (in shaded area). The dashed lines represent the true physical system $\eta_0(x) = 4x + x \sin(5x)$.

Table 3. Summary statistics of the posterior of θ for configuration 3 (simulation truth is $\theta_0^* = 3.5609$).

	Projected calibration	KO calibration	OGP calibration
Mean	3.4064	3.1109	3.6001
Standard deviation	0.0614	0.4760	0.0911
97.5%-Quantile	3.5964	3.9385	3.7733
2.5%-Quantile	3.3624	2.1467	3.4167

is the current through sodium channels in a cardiac cell membrane, and the input is the time. For a detailed description of the experiment, we refer to Plumlee, Joseph, and Yang (2016). We consider a subset of the original dataset, which consists of 19 normalized current records needed to maintain the membrane potential fixed at -35 mV as the outputs, together with the logarithm of the corresponding time as the inputs. The same dataset was also studied in Plumlee (2017). For the computer model, Clancy and Rudy (1999) suggested the following Markov model for ion channels:

$$y^s(x, \theta) = \mathbf{e}_1^T \exp[\exp(x)A(\theta)]\mathbf{e}_4,$$

where $\mathbf{e}_i$ is the column vector with 1 at the i th coordinate and 0 for the rest components, the outer exp is the matrix exponential function, $\theta = [\theta_1, \theta_2, \theta_3]^T$, and

$$A(\theta) = \begin{bmatrix} -\theta_2 - \theta_3 & \theta_1 & 0 & 0 \\ \theta_2 & -\theta_1 - \theta_2 & \theta_1 & 0 \\ 0 & \theta_2 & -\theta_1 - \theta_2 & \theta_1 \\ 0 & 0 & \theta_2 & -\theta_1 \end{bmatrix}.$$

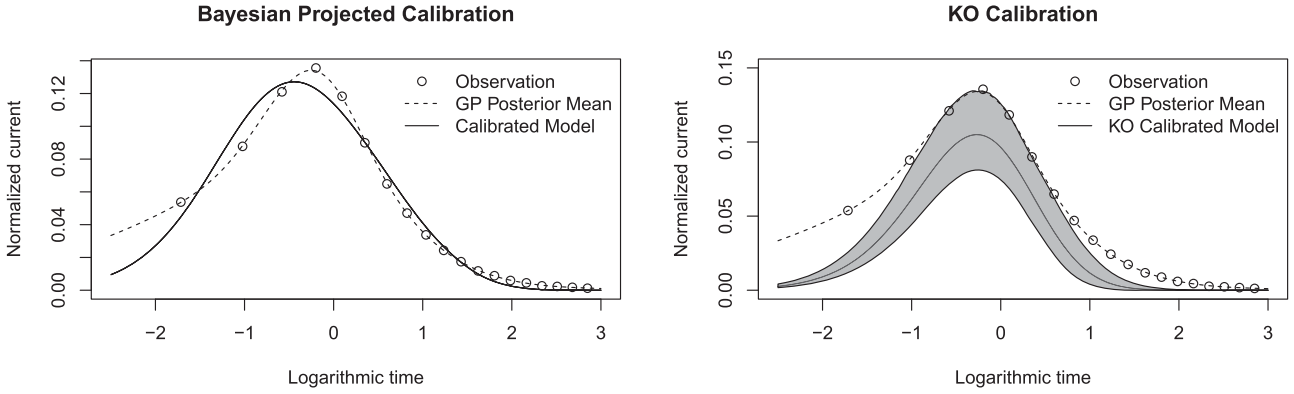
We follow the Monte Carlo methods described in Section 5.1 to collect 1000 posterior samples of θ using the Bayesian projected calibration and the KO calibration method. The OGP calibration method, however, fails to provide adequate samples from the posterior distribution within 20 hours. The smoothness parameter α for the Matérn covariance function is set to $5/2$, and we set $\tau = 0.02$, $\sigma = 0.001$, as suggested by Plumlee (2017). Table 4 presents the comparison of the corresponding summary statistics. The calibrated computer models are also visualized in Figure 6. Clearly, the Bayesian projected calibration provides better estimates to both the calibration parameter θ and the computer model in terms of lower uncertainty (a smaller standard deviation and thinner credible intervals, respectively). It can also be seen that the Bayesian projected calibrated computer model provides a better approximation to the physical data than the KO method does.

5.3. Spot Welding Example

Now we consider the spot welding example studied in Bayarri et al. (2007) and Chang and Joseph (2014). The three control variables in the experiment are the load, the current, and the gauge. The physical experiments are only conducted for gauge being 1 and 2. Since the computer model fails to produce sufficient meaningful outputs when the gauge is set to 1, we only consider the case where the gauge is 2, that is, the control variables are the load and the current only. The physical data are provided in

Table 4. Summary statistics of the posterior of θ for the ion channel example.

θ	Projected calibration			KO calibration		
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Mean	6.011166	5.578567	3.500813	3.4713447	0.9325514	6.7811932
Standard deviation	0.000012	0.000006	0.000006	0.2974497	0.5369031	1.1803662
97.5%-Quantile	6.011191	5.578578	3.500824	4.154933	2.034486	9.148351
2.5%-Quantile	6.011143	5.578556	3.500802	3.009278	0.114780	4.536802

**Figure 6.** Visualization of computer model calibration for the ion channel example. The left and right panels present the calibrated computer models (dashed lines) using the proposed approach and the KO calibration approach, respectively. The shaded area is the point-wise 95%-credible intervals for the KO calibrated computer model. The physical data (circles) and the Gaussian process (GP) estimates of the physical system (dashed lines) are also displayed.

Bayarri et al. (2007, Table 4). For each fixed design point, the mean of the 10 outputs replicates is taken as the response.

The computer model for the spot welding system is not directly available to us. The computer model consists of a time-consuming finite element method (FEM) for numerically solving a system of partial differential equations (PDEs). There are 21 available runs for the computer code, as presented in Bayarri et al. (2007, Table 3). Besides the three control variables (the load, the current, and the gauge) in the physical experiment, the computer model also involves another unknown parameter θ (denoted as u in Bayarri et al. 2007) that summarizes the material and surface. This parameter needs to be tuned with the physical data and is exactly the calibration parameter in the current context. As discussed in Section 2, an emulator is needed as a surrogate for the computer model when the code is expensive. Here we apply the RobustGaSP package (Gu, Palomo, and Berger 2018) to emulate the expensive FEM computer model. For theoretical properties of the RobustGaSP emulator, we refer to Gu, Wang, and Berger (2018).

We draw posterior samples using the Bayesian projected calibration and the KO method. The only difference is that the expensive-to-run computer model y^s is replaced by the predictive mean of the RobustGaSP emulator based on the results of 21 runs on the FEM computer code. The summary statistics for θ are presented in Table 5, indicating that the Bayesian projected calibration outperforms the KO approach in terms of the uncertainty quantification for estimating θ , that is, a smaller standard deviation and a thinner credible interval. The calibrated computer models using the Bayesian projected calibration and the KO approach, together with their corresponding point-wise 95%-credible intervals, are depicted in Figure 7. We can see that in terms of computer model calibration, both approaches behave similarly. The point-wise credible intervals, however, indicate that the Bayesian projected calibration method outperforms the

Table 5. Summary statistics of the posterior of θ for the spot welding example.

	Projected calibration	KO calibration
Mean	4.385933	4.126239
Standard deviation	0.08455849	1.440555
97.5%-Quantile	4.505187	7.164378
2.5%-Quantile	4.183981	1.604301

KO approach regarding the uncertainty quantification for y^s . The imperfection of the computer model can also be seen from the discrepancy presented on the right two panels of Figure 7.

6. Beyond the Projected Calibration

In this work, we follow the definition in Tuo and Wu (2015) and Wong, Storlie, and Lee (2017) to define the true calibration parameter θ_0^* as the minimizer of the L_2 -distance between the physical system $\eta(\cdot)$ and the computer model $y^s(\cdot, \theta)$. However, it may not be always appropriate to define θ_0^* via minimizing the L_2 -distance between $\eta(\cdot)$ and $y^s(\cdot, \theta)$ in practice. For example, in some scientific problems, the calibration parameter could have a physical meaning and there also exist circumstances where the physical system is more complicated than a computer mode plus model discrepancy. In this section, we discuss several scenarios in which the L_2 -projection approach may not apply.

First, the definition of $\theta_0^* = \arg \min_{\theta \in \Theta} \|\eta_0(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2$ requires that the minimizer is uniquely defined, which may not hold in practice. In addition, in certain applications, experts have scientific knowledge on the calibration parameter and the corresponding computer model. In such scenarios θ_0^* could be defined via alternative loss functions. For example, if scientists have certain knowledge on the relationship between the computer model mechanism and the calibration parameter through a penalty function $\mathcal{P}(\theta, y^s)$, one may define

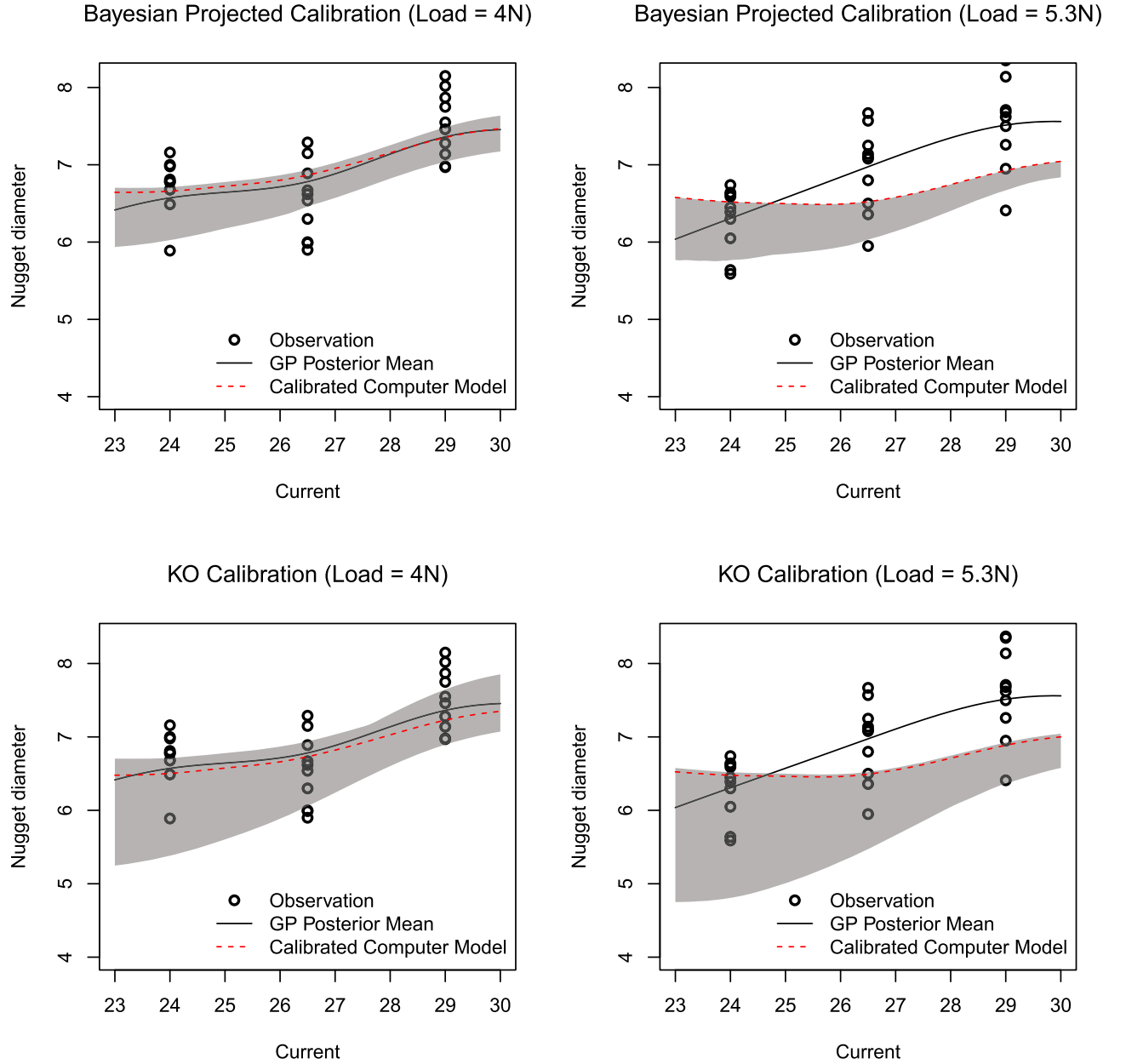


Figure 7. Visualization of computer model calibration for the spot welding example. The left and right panels present the calibrated computer models (red dashed lines) as a function of the current with the load fixed at 4 N and 5.3 N, respectively. The shaded areas are the point-wise 95%-credible intervals for the corresponding calibrated computer models. The physical data (circles) and the Gaussian process (GP) estimates of the physical system (solid lines) are also displayed.

θ_0^* to be the minimizer of a penalized L_2 -function

$$\|\eta_0(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2 + \mathcal{P}(\theta, y^s),$$

which allows us to reduce the discrepancy between the physical system and the computer model, and integrate the experts' knowledge within the calibration procedure through the penalty function $\mathcal{P}$ simultaneously. We expect that the corresponding asymptotic theory and Bayes estimators' efficiency can be developed following the same technique adopted in this work, provided that $\mathcal{P}$ is twice continuously differentiable. We defer this portion of extension to future work.

Second, when model discrepancy between the true physical system η and the computer model y^s cannot be modeled by an additive stochastic process, the projected calibration can be extended to accommodate this scenario. For instance, consider

the following nonlinear state space model

$$\eta''(x) = \theta_1 \eta'(x) + \theta_2 \eta(x) + \theta_3 \eta^3(x) + q(x) + \delta(x),$$

where $q(x)$ is some known process, δ is model discrepancy, $\theta = [\theta_1, \theta_2, \theta_3]^T$ is the calibration parameter, and noisy measurements y_i 's are generated from $y_i = \eta(\mathbf{x}_i) + e_i$ with $e_i \sim N(0, \sigma^2)$ independently. For simplicity let q be deterministic. One may define θ_η^* by

$$\theta_\eta^* = \arg \min_{\theta \in \Theta} \|\eta''(x) - \theta^T \mathbf{v}_\eta(\cdot) - q(\cdot)\|_{L_2(\Omega)}^2,$$

where $\mathbf{v}_\eta(x) = [\eta'(x), \eta(x), \eta^3(x)]^T$. It is convenient to model η , η' , and η'' jointly by imposing a GP prior on η with sufficient smoothness (see, e.g., Rasmussen and Williams 2006, sec.

9.4). Furthermore, simple algebra directly leads to the following closed-form formula for θ_η^* :

$$\theta_\eta^* = \left[\int_{\Omega} \mathbf{v}_\eta(x) \mathbf{v}_\eta(x)^T dx \right]^{-1} \int_{\Omega} \mathbf{v}_\eta(x) [\eta''(x) - q(x)] dx.$$

The above formula can be used to compute θ_η^* once η is appropriately modeled through a well-behaved GP, but theoretical properties would require a separate treatment.

Last but not least, taking the L_2 -minimization criterion as the definition of the true calibration parameter may not be always appropriate when θ has certain physical meaning. Let us take the free fall example in physics introduced in Plumlee (2017) for illustration. Suppose a ball is released at vertical height y_0 with an unknown initial velocity v_0 , and the vertical height of the ball y decreases due to gravity as the time x increases. Then Newtonian mechanism in a vacuum condition for describing the ball falling process leads to the following quadratic formula:

$$y(x) = y_0 + v_0 x - \frac{1}{2} g x^2, \quad 0 \leq x \leq 1,$$

where g is the gravitational acceleration. Denoting $\theta = [v_0, g]^T$, we can treat the preceding display as a computer model $y^\delta(x, \theta) = y(x)$ for describing the ball falling process. In reality free fall experiments are rarely done in a vacuum condition, and the true free fall system is usually affected by air resistance. Consequently, the acceleration (namely, the second derivative of the height) of the ball at time x equals $-g + c v(x)^2$, where $v(x)$ is the velocity of the ball at time x and c is the coefficient of air resistance. Taking $c = 0.02$, $y_0 = 8$ m, $v_0 = -1$ m/s, $g = 10$ m/s², where we assume that experiments are done on earth, we obtain the following true physical system of the free fall (Plumlee 2017):

$$\eta(x) = \frac{5}{2} \log \left\{ \frac{50}{49} - \frac{50}{49} \tanh^2 \left[\sqrt{2} x + \tanh^{-1}(\sqrt{0.02}) \right] \right\} + 8, \\ 0 \leq x \leq 1.$$

The physical meaning of the calibration parameter $\theta = [v_0, g]^T$ is clear in the free fall experiment: v_0 represents the initial velocity of the ball when it is released, and g is the gravitational acceleration. Taking θ_0^* as the minimizer of $\|\eta(\cdot) - y^\delta(\cdot, \theta)\|_{L_2(0,1)}^2$ gives rise to $v_0 = -1.6377$ m/s and $g = 5.8965$ m/s². Although the resulting calibrated computer model can be used as an emulator for the true physical system of the free fall with air resistance, the calibration parameter values are questionable: The ball is never released at the initial velocity $v_0 = -1.6377$ m/s (which should be $v_0 = -1$ m/s ideally), and the gravitational acceleration on earth is far from 5.8965 m/s² (which should be $g \approx 9.8$ m/s² according to the physics literature). This example shows that taking the true values of v_0 and g as the minimizer of $\|\eta(\cdot) - y^\delta(\cdot, \theta)\|_{L_2(0,1)}^2$ for best fitting is inappropriate in light of the physical meaning of these parameters.

To accommodate the physical meaning of v_0 and g , we can modify the definition of $\theta_0^* = \arg \min_{\theta} \|\eta_0(\cdot) - y^\delta(\cdot, \theta)\|_{L_2(0,1)}^2$ as follows. Recall that v_0 is the velocity of the ball at time $x = 0$. We expect that the true free fall system gives rise to the initial velocity of the ball being close to v_0 , namely, $[\eta'(0) - v_0]^2$ should

be sufficiently small, where $\eta'(x)$ is the velocity of the ball in the true system at time x . Similarly, the second derivative of the height of the ball in the true free fall system satisfies $\eta''(x) = -g + c[\eta'(x)]^2$. When $x = 0$, $\eta''(0)$ is expected to be close to the gravitational acceleration $-g$ because $[\eta''(x) + g]^2$ is minimized when $x = 0$. Hence, we modify the L_2 -minimization criterion after taking into account the physical meaning of θ :

$$[v_0^*, g_0^*]^T = \arg \min_{v_0, g} \left\{ \|\eta(\cdot) - y^\delta(\cdot, \theta)\|_{L_2(0,1)}^2 + \lambda_1 [\eta'(0) - v_0]^2 + \lambda_2 [\eta''(0) + g]^2 \right\},$$

where λ_1 and λ_2 are tuning parameters representing the importance of learning the initial velocity and the gravitational acceleration, respectively. For example, setting $\lambda_1 = \lambda_2 = 10$ results in $v_0^* = -0.9735$ m/s, $g_0^* = 9.7888$ m/s², which is more interpretable in terms of physical meaning compared to the naive L_2 -projection formulation, as the initial velocity was set to be $v_0 = -1$ m/s and the gravitational acceleration on earth is $g \approx 9.8$ m/s². Therefore, the proposed L_2 -projected calibration approach is flexible for extensions to handle scenarios where calibration parameters have specific physical meaning.

7. Conclusion and Discussion

We develop a novel Bayesian projected calibration method following the frequentist L_2 -projected calibration method in Tuo and Wu (2015). The proposed approach is formulated in an identifiable way and naturally quantifies the uncertainty in the calibration problem through the posterior distribution. Theoretical justification of the Bayesian projected calibration is provided: the marginal posterior distribution of the calibration parameter is not only $\sqrt{n}$ -consistent, but also asymptotically normal with the efficient covariance matrix. We also provide the practitioners with two easy-to-implement and efficient computational algorithms for the computation of the Bayesian projected calibration with theoretical support.

To obtain sensible estimators of the true calibration parameter θ_0^* (defined as the minimizer of the L_2 -distance between the physical system and the computer model), the OGP calibration method proposed in Plumlee (2017) and the Bayesian projected calibration proposed in this work can be applied. Alternatively, Gu and Wang (2018) proposed to directly apply a modified GP prior, referred to as the scaled Gaussian process (S-GaSP), to the discrepancy function $\delta(\mathbf{x}) = \eta(\mathbf{x}) - y^\delta(\mathbf{x}, \theta)$ for computer model calibration. The scaled Gaussian process is defined by modifying the eigenvalues of the covariance function of some classical GP (e.g., the Matérn process or the squared-exponential process) such that the sample paths have smaller L_2 -norms than the original GP. The construction of the S-GaSP is slightly involved, but the resulting maximum a posteriori estimator of θ and δ can be expressed as the following doubly penalized kernel ridge regression problem (Gu, Xie, and Wang 2018):

$$(\hat{\theta}, \hat{\delta}) = \arg \min_{\theta \in \Theta, \delta \in \mathbb{H}(\Omega)} \frac{1}{n} \sum_{i=1}^n [y_i - y^\delta(\mathbf{x}_i, \theta) - \delta(\mathbf{x}_i)]^2 + \lambda_1 \|\delta\|_{\mathbb{H}(\Omega)}^2 + \lambda_2 \|\delta\|_{L_2(\Omega)}^2,$$

where $\mathbb{H}(\Omega)$ is the RKHS associated with the original GP, and $\lambda_1, \lambda_2 > 0$ are the tuning parameters. The motivation of the extra penalty term $\lambda_2 \|\delta\|_{L_2(\Omega)}^2$ in comparison with the classical kernel ridge regression exactly comes from the idea of minimizing the L_2 -norm of δ . When λ_1 and λ_2 are carefully selected, the resulting estimators $\hat{\theta}$ converges to θ_0^* , but the rate is slower than $1/\sqrt{n}$ (Gu, Xie, and Wang 2018) compared to that of the Bayesian projected calibration and the L_2 -calibration, which may not be desired when the efficient estimation of θ is needed.

The estimation methods in this work and Tuo and Wu (2015) can be viewed as the following two-step procedure: first estimate the physical system through Gaussian process models; then estimate the calibration parameter using the L_2 -projection criterion. On the other hand, it is possible to jointly estimate the calibration parameter and the discrepancy between the physical system and the computer model. The aforementioned OGP calibration method (Plumlee 2017) exactly tackles this joint estimation issue. The theoretical development for OGP, nevertheless, is only restricted to a point estimator (Tuo 2017): the maximum a posteriori (MAP) estimator of θ is asymptotically normal and semiparametric efficient. It will be nontrivial to apply the technical results developed here to the OGP calibration method, and the asymptotic characterization of the corresponding full posterior distribution will be an interesting topic.

Similar to the OGP calibration method, the Bayesian projected calibration also involves intractable integrals, and we propose to use stochastic approximation methods to reduce the computation complexity in Section 4. For moderately large sample sizes, one can apply Algorithm 2 to compute a decent approximation to the projected calibration efficiently, but for sparse data, one has to rely on the slightly more cumbersome Algorithm 1 to perform the exact posterior inference. It is therefore desired that the computational barrier of Algorithm 1 can be tackled via more efficient algorithms.

Appendix A: Auxiliary Results

In this section, we list some auxiliary results that are used to prove Theorem 2. The proofs of the lemmas stated in this section are deferred to the supplementary materials. Before proceeding, we introduce some notions and definitions that are widely used in the study of empirical processes. Suppose $\mathcal{F}$ is a function space equipped with metric d . Given two functions $l, u \in \mathcal{F}$, a bracket $[l, u]$ is a set of functions f such that $l \leq f \leq u$ everywhere, and the size of the bracket is defined to be $d(l, u)$. The ϵ -bracketing number of $\mathcal{F}$ with respect to the metric d , denoted by $\mathcal{N}_{[\cdot, \cdot]}(\epsilon, \mathcal{F}, d)$, is the minimum number of brackets of size ϵ that are needed to cover $\mathcal{F}$. The bracketing integral $J_{[\cdot, \cdot]}(\epsilon, \mathcal{F}, d)$ is defined to be the integral of the logarithmic bracketing number as follows:

$$J_{[\cdot, \cdot]}(\epsilon, \mathcal{F}, d) = \int_0^\epsilon \sqrt{\log \mathcal{N}_{[\cdot, \cdot]}(\delta, \mathcal{F}, d)} d\delta.$$

Suppose $\mathcal{X}$ is the space where random variables take values. Given a sequence $(\mathbf{x}_i)_{i=1}^n$ of independent and identically distributed random variables, the empirical measure and the empirical process of a function $f : \mathcal{X} \rightarrow \mathbb{R}$, denoted by $\mathbb{P}_n f$ and $\mathbb{G}_n f$, are defined by

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i), \quad \mathbb{G}_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^n [f(\mathbf{x}_i) - \mathbb{E}f(\mathbf{x}_i)],$$

respectively. For two variables a and b , we use $a \lesssim b$ and $a \gtrsim b$ to denote the inequalities up to a universal multiplicative constant, and write $a \asymp b$ if $a \lesssim b$ and $a \gtrsim b$.

In the empirical processes theory, maximum inequalities are widely adopted to study the asymptotic behavior of nonparametric estimators. Here we cite one of them that is used in the proof of Theorem 2 (see, e.g., van der Vaart 2000, Lemma 19.36).

Theorem 5. Let $(\mathbf{x}_i)_{i=1}^n$ be independent and identically distributed according to a distribution $\mathbb{P}_{\mathbf{x}}$ over $\mathcal{X}$, and let $\mathcal{F}$ be a class of measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$. If $\|f\|_{L_2(\mathbb{P}_{\mathbf{x}})}^2 < \delta^2$ and $\|f\|_\infty \leq M$ for all $f \in \mathcal{F}$, where δ and M does not depend on $\mathcal{F}$, then

$$\mathbb{E} \left[\sup_{f \in \mathcal{F}} |\mathbb{G}_n f| \right] \lesssim J_{[\cdot, \cdot]}(\delta, \mathcal{F}, \|\cdot\|_{L_2(\mathbb{P}_{\mathbf{x}})}) \times \left[1 + \frac{M}{\delta^2 \sqrt{n}} J_{[\cdot, \cdot]}(\delta, \mathcal{F}, \|\cdot\|_{L_2(\mathbb{P}_{\mathbf{x}})}) \right].$$

The following lemma is the modification of a standard probabilistic theorem for Gaussian processes. For the related literature, we refer to van der Vaart and van Zanten (2008) and Ghosal and van der Vaart (2017).

Lemma 2. Suppose η is imposed the Matérn Gaussian process with smoothness parameter α , and $\eta_0 \in \mathcal{C}_\alpha(\Omega) \cap \mathcal{H}_\alpha(\Omega)$, where $\alpha > p/2$. Let $\epsilon_n = n^{-\alpha/(2\alpha+p)}$. Then there exists a measurable set $\mathcal{B}_n$ in $\mathcal{C}(\Omega)$ (the space of all continuous functions on Ω) such that for sufficiently large n , the following hold:

$$\begin{aligned} \Pi(\mathcal{B}_n \mid \mathcal{D}_n) &= 1 - o_{\mathbb{P}_0}(1), \\ J_{[\cdot, \cdot]}(\epsilon_n \log n, \mathcal{B}_n, \|\cdot\|_{L_2(\Omega)}) &\lesssim (\log n)^{2\alpha/(2\alpha+p)} \sqrt{n} \epsilon_n^2. \end{aligned}$$

Now denote

$$\ell_n(\eta) = \sum_{i=1}^n \log p_\eta(y_i, \mathbf{x}_i) = \sum_{i=1}^n \log \phi_\sigma(y_i - \eta(\mathbf{x}_i))$$

to be the log-likelihood function of η given the physical data $(\mathbf{x}_i, y_i)_{i=1}^n$. Define the event

$$\mathcal{A}_n = \{ \|\eta - \eta_0\|_{L_2(\Omega)} \leq M_n \epsilon_n \} \cap \{ \|\eta - \eta_0\|_{L_\infty(\Omega)} \leq M \} \cap \mathcal{B}_n,$$

where $M_n = \log n$, M is given by Theorem 1, and $\mathcal{B}_n$ is given by Lemma 2. Then by Theorem 1 and Lemma 2 we know that $\Pi(\mathcal{A}_n \mid \mathcal{D}_n) = 1 - o_{\mathbb{P}_0}(1)$.

Lemma 3. Suppose the conditions of Theorem 2 hold. For each vector $\mathbf{t} \in \mathbb{R}^q$ and each $\eta \in \mathcal{F}$ define

$$\eta_{\mathbf{t}}(\mathbf{x}) = \eta(\mathbf{x}) - \frac{2\sigma^2}{\sqrt{n}} \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^\delta}{\partial \theta}(x, \theta_0^*).$$

Given a realization η of the Matérn Gaussian process $\text{GP}(0, \Psi_\alpha)$, define the following isometry associated to η :

$$\begin{aligned} U : \mathbb{H}_0 &= \left\{ \sum_{k=1}^K a_k \Psi(\cdot, \mathbf{t}_k) : \mathbf{t}_k \in \Omega, a_k \in \mathbb{R}, K \in \mathbb{N}_+ \right\} \rightarrow L_2(\mathbb{P}_0), \\ \sum_{k=1}^K a_k \Psi(\cdot, \mathbf{t}_k) &\mapsto \sum_{k=1}^K a_k \eta(\mathbf{t}_k), \end{aligned}$$

and extend U from $\mathbb{H}_0$ to $\overline{\mathbb{H}}_0 = \mathbb{H}_{\Psi_\alpha}(\Omega)$ continuously. Define the event

$$\mathcal{C}_n = \{ |U(g)| \leq L \sqrt{n} \epsilon_n \|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)} \}.$$

Then there exists a sufficiently large L such that $\Pi(\mathcal{C}_n^c \mid \mathcal{D}_n) = o_{\mathbb{P}_0}(1)$, and the following holds:

$$\begin{aligned} &\int_{\mathcal{A}_n \cap \mathcal{C}_n} \exp[\ell_n(\eta) - \ell_n(\eta_0)] \Pi(d\eta) \\ &= [1 + o_{\mathbb{P}_0}(1)] \left\{ \int \exp[\ell_n(\eta) - \ell_n(\eta_0)] \Pi(d\eta) \right\}. \end{aligned}$$

The asymptotic normality result of the L_2 -projected calibration estimator $\widehat{\theta}_{L_2}$ from Tuo and Wu (2015) is also useful to study the asymptotic behavior of $\Pi(\sqrt{n}(\theta_\eta^* - \widehat{\theta}_{L_2}) \in \cdot \mid \mathcal{D}_n)$.

Theorem 6. Under the conditions of Theorem 2, it holds that

$$\widehat{\theta}_{L_2} - \theta_0^* = 2\mathbf{V}_0^{-1} \left[\frac{1}{n} \sum_{i=1}^n e_i \frac{\partial y^s}{\partial \theta}(\mathbf{x}_i, \theta_0^*) \right] + o_{\mathbb{P}_0}(n^{-1/2}).$$

Appendix B: Proof of Theorem 2

Theorem 1 and Lemma 2 imply that $\Pi(\mathcal{A}_n \cap \mathcal{C}_n \mid \mathcal{D}_n) = 1 - o_{\mathbb{P}_0}(1)$. Let $\Theta_n = \{\theta_\eta^* : \eta \in \mathcal{A}_n \cap \mathcal{C}_n\}$. It follows directly that $\Pi(\theta_\eta^* \in \Theta_n \mid \mathcal{D}_n) = 1 - o_{\mathbb{P}_0}(1)$. Denote

$$\Pi(\theta_\eta^* \in \cdot \mid \mathcal{D}_n, \Theta_n) = \frac{\Pi(\theta_\eta^* \in \cdot \cap \Theta_n \mid \mathcal{D}_n)}{\Pi(\theta_\eta^* \in \Theta_n \mid \mathcal{D}_n)}.$$

Following the argument in Castillo and Rousseau (2015a), it suffices to show that

$$\sup_A \left| \Pi \left(\sqrt{n}(\theta_\eta^* - \widehat{\theta}_{L_2}) \in A \mid \mathcal{D}_n, \Theta_n \right) - N \left(\mathbf{0}, 4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1} \right) \right| \xrightarrow{\mathbb{P}_0} 0.$$

We prove the result by the method of moment generating function, namely, showing that for any fixed vector $\mathbf{t} \in \mathbb{R}^q$, it holds that

$$\begin{aligned} & \int_{\mathcal{A}_n \cap \mathcal{C}_n} \exp \left[\mathbf{t}^T \sqrt{n} (\theta_\eta^* - \widehat{\theta}_{L_2}) \right] \Pi(d\eta \mid \mathcal{D}_n) \\ & \rightarrow \exp \left[\frac{1}{2} \mathbf{t}^T (4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1}) \mathbf{t} \right] \end{aligned}$$

in $\mathbb{P}_0$ -probability. The rest part of the proof is completed by Castillo and Rousseau (2015b, Lemmas 1 and 2).

Let $\epsilon_n = n^{-\alpha/(2\alpha+p)}$. The left-hand side of the preceding display can be rewritten as

$$\begin{aligned} & \left\{ \int \exp [\ell_n(\eta) - \ell_n(\eta_0)] \Pi(d\eta) \right\}^{-1} \\ & \left\{ \int_{\mathcal{A}_n \cap \mathcal{C}_n} \exp \left[\mathbf{t}^T \sqrt{n} (\theta_\eta^* - \widehat{\theta}_{L_2}) + \ell_n(\eta) - \ell_n(\eta_0) \right] \Pi(d\eta) \right\}. \end{aligned}$$

For the vector $\mathbf{t} \in \mathbb{R}^q$, define

$$\eta_{\mathbf{t}}(\mathbf{x}) = \eta(\mathbf{x}) - \frac{2\sigma^2}{\sqrt{n}} \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}, \theta_0^*),$$

and for each η , define the remainder

$$R_n(\eta, \eta_0) = \frac{n}{2} \|\eta - \eta_0\|_{L_2(\Omega)}^2 - \frac{n}{2} \mathbb{P}_n(\eta - \eta_0)^2.$$

Then simple algebra shows

$$\begin{aligned} & [\ell_n(\eta_{\mathbf{t}}) - \ell_n(\eta_0)] - [\ell_n(\eta) - \ell_n(\eta_0)] \\ & = -\frac{n}{2\sigma^2} \left[\|\eta_{\mathbf{t}} - \eta_0\|_{L_2(\Omega)}^2 - \|\eta - \eta_0\|_{L_2(\Omega)}^2 \right] \\ & \quad - \frac{2}{\sqrt{n}} \sum_{i=1}^n e_i \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}_i, \theta_0^*) \\ & \quad + \frac{1}{\sigma^2} [R_n(\eta_{\mathbf{t}}, \eta_0) - R_n(\eta, \eta_0)] \\ & = 2\sqrt{n} \int_{\Omega} [\eta(\mathbf{x}) - \eta_0(\mathbf{x})] \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}, \theta_0^*) d\mathbf{x} \\ & \quad - \frac{1}{2} \mathbf{t}^T (4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1}) \mathbf{t} \\ & \quad - \frac{2}{\sqrt{n}} \sum_{i=1}^n e_i \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}_i, \theta_0^*) + \frac{1}{\sigma^2} [R_n(\eta_{\mathbf{t}}, \eta_0) - R_n(\eta, \eta_0)]. \end{aligned}$$

Denote the remainder of the Taylor expansion of θ_η^* at θ_0^* by

$$\mathbf{r}(\eta, \eta_0) = \theta_\eta^* - \theta_0^* - 2 \int_{\Omega} [\eta(\mathbf{x}) - \eta_0(\mathbf{x})] \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}, \theta_0^*) d\mathbf{x}.$$

Then by Theorem 6 we have

$$\begin{aligned} & \mathbf{t}^T \sqrt{n} (\theta_\eta^* - \widehat{\theta}_{L_2}) + \ell_n(\eta) - \ell_n(\eta_0) \\ & = \mathbf{t}^T \sqrt{n} (\theta_\eta^* - \theta_0^*) - \frac{2}{\sqrt{n}} \sum_{i=1}^n e_i \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}_i, \theta_0^*) \\ & \quad + o_{\mathbb{P}_0}(1) + \ell_n(\eta) - \ell_n(\eta_0) \\ & = \mathbf{t}^T \sqrt{n} (\theta_\eta^* - \theta_0^*) + o_{\mathbb{P}_0}(1) \\ & \quad - 2\sqrt{n} \int_{\Omega} [\eta(\mathbf{x}) - \eta_0(\mathbf{x})] \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}, \theta_0^*) d\mathbf{x} \\ & \quad + \frac{1}{2} \mathbf{t}^T (4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1}) \mathbf{t} - \frac{1}{\sigma^2} [R_n(\eta_{\mathbf{t}}, \eta_0) - R_n(\eta, \eta_0)] \\ & \quad + \ell_n(\eta_{\mathbf{t}}) - \ell_n(\eta_0) \\ & = \frac{1}{2} \mathbf{t}^T (4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1}) \mathbf{t} + \sqrt{n} \mathbf{t}^T \mathbf{r}(\eta, \eta_0) \\ & \quad + \frac{1}{\sigma^2} [R_n(\eta, \eta_0) - R_n(\eta_{\mathbf{t}}, \eta_0)] \\ & \quad + \ell_n(\eta_{\mathbf{t}}) - \ell_n(\eta_0) + o_{\mathbb{P}_0}(1). \end{aligned}$$

Now set $M_n = \log n$. By Lemma 1 we see that

$$\begin{aligned} \sup_{\eta \in \mathcal{A}_n \cap \mathcal{C}_n} \left| \sqrt{n} \mathbf{t}^T \mathbf{r}(\eta, \eta_0) \right| & \leq L_{\eta_0}^{(2)} \|t\| \sqrt{n} M_n^2 n^{-2\alpha/(2\alpha+p)} \\ & \lesssim M_n^2 n^{(p/2-\alpha)/(2\alpha+p)} = o(1). \end{aligned}$$

In addition, simple algebra and the law of large numbers imply that

$$\begin{aligned} & R_n(\eta, \eta_0) - R_n(\eta_{\mathbf{t}}, \eta_0) \\ & = \frac{2\sigma^4}{n} \sum_{i=1}^n \left[\mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\mathbf{x}_i, \theta_0^*) \right]^2 - 2\sigma^4 \mathbf{t}^T \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1} \mathbf{t} - 2\sigma^2 \mathbb{G}_n \\ & \quad \times \left[(\eta - \eta_0)(\cdot) \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\cdot, \theta_0^*) \right] \\ & = -2\sigma^2 \mathbb{G}_n \left[(\eta - \eta_0)(\cdot) \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\cdot, \theta_0^*) \right] + o_{\mathbb{P}_0}(1). \end{aligned}$$

We now claim that

$$\sup_{\eta \in \mathcal{A}_n} |R_n(\eta, \eta_0) - R_n(\eta_{\mathbf{t}}, \eta_0)| = o_{\mathbb{P}_0}(1).$$

Since over $\mathcal{A}_n$, we know that $\|\eta - \eta_0\|_{L_2(\Omega)} \leq M_n \epsilon_n$, $\|\eta - \eta_0\|_{L_\infty(\Omega)} \leq M$, and by Lemma 2 it holds that

$$\begin{aligned} J_{[\cdot]}(M_n \epsilon_n, \mathcal{A}_n, \|\cdot\|_{L_\infty(\Omega)}) & \lesssim M_n^{2\alpha/(2\alpha+p)} \sqrt{n} \epsilon_n^2 \\ & = (\log n)^{2\alpha/(2\alpha+p)} \sqrt{n} \epsilon_n^2, \end{aligned}$$

it follows by the maximal inequality for empirical process (Theorem 5) that

$$\begin{aligned} & \mathbb{E}_0 \left\{ \sup_{\eta \in \mathcal{A}_n} \left| \mathbb{G}_n \left[(\eta - \eta_0)(\cdot) \mathbf{t}^T \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \theta}(\cdot, \theta_0^*) \right] \right| \right\} \\ & \lesssim J_{[\cdot]}(M_n \epsilon_n, \mathcal{A}_n, \|\cdot\|_{L_2(\Omega)}) \left[1 + M \frac{J_{[\cdot]}(M_n \epsilon_n, \mathcal{A}_n, \|\cdot\|_{L_2(\Omega)})}{M_n^2 \epsilon_n^2 \sqrt{n}} \right] \\ & \lesssim M_n^{2\alpha/(2\alpha+p)} \sqrt{n} \epsilon_n^2 \left[1 + \frac{M_n^{2\alpha/(2\alpha+p)} \sqrt{n} \epsilon_n^2}{M_n^2 \epsilon_n^2 \sqrt{n}} \right] \\ & \lesssim M_n \sqrt{n} \epsilon_n^2 = o(1), \end{aligned}$$

and hence, it holds that $\sup_{\eta \in \mathcal{A}_n \cap \mathcal{C}_n} |R_n(\eta, \eta_0) - R_n(\eta_t, \eta_0)| = o_{\mathbb{P}_0}(1)$. Therefore by applying Lemma 3, we obtain

$$\begin{aligned} & \int_{\mathcal{A}_n \cap \mathcal{C}_n} \exp \left[\mathbf{t}^T \sqrt{n} \left(\boldsymbol{\theta}_\eta^* - \widehat{\boldsymbol{\theta}}_{L_2} \right) + \ell_n(\eta) - \ell_n(\eta_0) \right] \Pi(d\eta) \\ &= \exp \left[\frac{1}{2} \mathbf{t}^T \left(4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1} \right) \mathbf{t} + o_{\mathbb{P}_0}(1) \right] \\ & \quad \int_{\mathcal{A}_n} \exp [\ell_n(\eta_t) - \ell_n(\eta_0)] \Pi(d\eta) \\ &= \exp \left[\frac{1}{2} \mathbf{t}^T \left(4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1} \right) \mathbf{t} + o_{\mathbb{P}_0}(1) \right] [1 + o_{\mathbb{P}_0}(1)] \\ & \quad \int \exp [\ell_n(\eta) - \ell_n(\eta_0)] \Pi(d\eta) \\ &= \left\{ \exp \left[\frac{1}{2} \mathbf{t}^T \left(4\sigma^2 \mathbf{V}_0^{-1} \mathbf{W} \mathbf{V}_0^{-1} \right) \mathbf{t} \right] + o_{\mathbb{P}_0}(1) \right\} \\ & \quad \int \exp [\ell_n(\eta) - \ell_n(\eta_0)] \Pi(d\eta). \end{aligned}$$

The proof is thus completed.

Supplementary Materials

The supplementary materials contain the remaining proofs, additional numerical results, and the R package BayProjected for implementing the computation methods in Section 4.

Funding

The work of Xu was supported by NSF 1918854 and NSF 1940107.

References

- Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007), "A Framework for Validation of Computer Models," *Technometrics*, 49, 138–154. [1,2,11,12]
- Bickel, P., and Kleijn, B. (2012), "The Semiparametric Bernstein-von Mises Theorem," *The Annals of Statistics*, 40, 206–237. [4]
- Brynjarsdóttir, J., and O'Hagan, A. (2014), "Learning About Physical Parameters: The Importance of Model Discrepancy," *Inverse Problems*, 30, 114007. [2]
- Castillo, I., and Rousseau, J. (2015a), "A Bernstein-von Mises Theorem for Smooth Functionals in Semiparametric Models," *The Annals of Statistics*, 43, 2353–2383. [4,16]
- (2015b), Supplement to "A Bernstein-von Mises Theorem for Smooth Functionals in Semiparametric Models," *The Annals of Statistics*, 43, 2353–2383. [16]
- Chang, C.-J., and Joseph, V. R. (2014), "Model Calibration Through Minimal Adjustments," *Technometrics*, 56, 474–482. [1,11]
- Clancy, C. E., and Rudy, Y. (1999), "Linking a Genetic Defect to Its Cellular Phenotype in a Cardiac Arrhythmia," *Nature*, 400, 566–569. [11]
- Duchi, J., Hazan, E., and Singer, Y. (2011), "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *Journal of Machine Learning Research*, 12, 2121–2159. [7]
- Fang, K.-T., Li, R., and Sudjianto, A. (2005), *Design and Modeling for Computer Experiments*, Boca Raton, FL: CRC Press. [1]
- Ghosal, S., and van der Vaart, A. (2017), *Fundamentals of Nonparametric Bayesian Inference* (Vol. 44), Cambridge: Cambridge University Press. [3,5,15]
- Gu, M., Palomo, J., and Berger, J. O. (2018), "RobustGaSP: Robust Gaussian Stochastic Process Emulation in R," arXiv no. 1801.01874. [12]
- Gu, M., and Wang, L. (2017), "An Improved Approach to Bayesian Computer Model Calibration and Prediction," arXiv no. 1707.08215. [8]
- (2018), "Scaled Gaussian Stochastic Process for Computer Model Calibration and Prediction," *SIAM/ASA Journal on Uncertainty Quantification*, 6, 1555–1583. [14]
- Gu, M., Wang, X., and Berger, J. O. (2018), "Robust Gaussian Stochastic Process Emulation," *The Annals of Statistics*, 46, 3038–3066. [12]
- Gu, M., Xie, F., and Wang, L. (2018), "A Theoretical Framework of the Scaled Gaussian Stochastic Process in Prediction and Calibration," arXiv no. 1807.03829. [14,15]
- Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004), "Combining Field Data and Computer Simulations for Calibration and Prediction," *SIAM Journal on Scientific Computing*, 26, 448–466. [1]
- Joseph, V. R., and Melkote, S. N. (2009), "Statistical Adjustments to Engineering Models," *Journal of Quality Technology*, 41, 362. [1]
- Kennedy, M. C., and O'Hagan, A. (2001), "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society, Series B*, 63, 425–464. [1,2,3,9]
- Li, X., and Orabona, F. (2018), "On the Convergence of Stochastic Gradient Descent With Adaptive Stepsizes," arXiv no. 1805.08114. [7]
- Plumlee, M. (2017), "Bayesian Calibration of Inexact Computer Models," *Journal of the American Statistical Association*, 112, 1274–1285. [2,3,8,9,11,14,15]
- Plumlee, M., Joseph, V. R., and Yang, H. (2016), "Calibrating Functional Parameters in the Ion Channel Models of Cardiac Cells," *Journal of the American Statistical Association*, 111, 500–509. [10,11]
- Qian, P. Z., and Wu, C. F. J. (2008), "Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments," *Technometrics*, 50, 192–204. [1]
- Rasmussen, C. E., and Williams, C. K. (2006), *Gaussian Processes for Machine Learning* (Vol. 1), Cambridge, MA: MIT Press. [4,13]
- Robbins, H., and Monroe, S. (1951), "A Stochastic Approximation Method," *The Annals of Mathematical Statistics*, 22, 400–407. [6]
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments," *Statistical Science*, 4, 409–423. [1]
- Santner, T. J., Williams, B. J., and Notz, W. I. (2013), *The Design and Analysis of Computer Experiments*, New York: Springer. [1,4]
- Sheridan, M., Bloebaum, C., Kesavadas, T., Patra, A., and Winer, E. (2002), "Visualization and Communication in Risk Management of Landslides," in *WIT Transactions on Modelling and Simulation* (Vol. 31). [1]
- Stone, C. J. (1982), "Optimal Global Rates of Convergence for Nonparametric Regression," *The Annals of Statistics*, 10, 1040–1053. [5]
- Storlie, C. B., Lane, W. A., Ryan, E. M., Gattiker, J. R., and Higdon, D. M. (2015), "Calibration of Computational Models With Categorical Parameters and Correlated Outputs via Bayesian Smoothing Spline ANOVA," *Journal of the American Statistical Association*, 110, 68–82. [2]
- Tuo, R. (2017), "Adjustments to Computer Models via Projected Kernel Calibration," arXiv no. 1705.03422. [2,3,15]
- Tuo, R., and Wu, C. F. J. (2015), "Efficient Calibration for Imperfect Computer Models," *The Annals of Statistics*, 43, 2331–2352. [1,2,3,5,6,12,14,15,16]
- (2016), "A Theoretical Framework for Calibration in Computer Models: Parametrization, Estimation and Convergence Properties," *SIAM/ASA Journal on Uncertainty Quantification*, 4, 767–795. [2,3]
- van der Vaart, A. W. (2000), *Asymptotic Statistics* (Vol. 3), Cambridge: Cambridge University Press. [4,15]
- van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer. [5]
- van der Vaart, A. W., and van Zanten, J. H. (2008), "Rates of Contraction of Posterior Distributions Based on Gaussian Process Priors," *The Annals of Statistics*, 36, 1435–1463. [15]
- (2009), "Adaptive Bayesian Estimation Using a Gaussian Random Field With Inverse Gamma Bandwidth," *The Annals of Statistics*, 37, 2655–2675. [9]

- (2011), “Information Rates of Nonparametric Gaussian Process Methods,” *Journal of Machine Learning Research*, 12, 2095–2119. [5]
- Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia, PA: SIAM. [3,4]
- Wang, S., Chen, W., and Tsui, K.-L. (2009), “Bayesian Validation of Computer Models,” *Technometrics*, 51, 439–451. [1]
- Wendland, H. (2004), *Scattered Data Approximation* (Vol. 17), Cambridge: Cambridge University Press. [3,4]
- Wong, R. K., Storlie, C. B., and Lee, T. (2017), “A Frequentist Approach to Computer Model Calibration,” *Journal of the Royal Statistical Society, Series B*, 79, 635–648. [2,3,5,12]
- Xie, F., and Xu, Y. (2019), “Optimal Bayesian Estimation for Random Dot Product Graphs,” arXiv no. 1904.12070. [3]
- (2020), “Bayesian Repulsive Gaussian Mixture Model,” *Journal of the American Statistical Association*, 115, 187–203. [3]
- Xiu, D. (2010), *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton, NJ: Princeton University Press. [4]