

# A THEORETICAL FRAMEWORK FOR CALIBRATION IN COMPUTER MODELS: PARAMETRIZATION, ESTIMATION AND CONVERGENCE PROPERTIES

BY RUI TUO AND C. F. JEFF WU

*Chinese Academy of Sciences and Georgia Institute of Technology*

Calibration parameters in deterministic computer experiments are those attributes that cannot be measured or available in physical experiments. Kennedy and O'Hagan (2001) suggested an approach to estimate them by using data from physical experiments and computer simulations. A theoretical framework is given which allows us to study the issues of parameter identifiability and estimation. It is shown that a simplified version of the original KO method leads to asymptotically inconsistent calibration. This calibration inconsistency can be remedied by modifying the original estimation procedure. A novel calibration method, called the  $L_2$  calibration, is proposed and proven to be consistent and enjoys optimal convergence rate. A numerical example and some mathematical analysis are used to illustrate the source of the inconsistency problem.

**1. Introduction.** Because of the advances in complex mathematical models and fast computer codes, experiments on a computer, or referred to as computer experiments in the statistical literature, have become popular in engineering and scientific investigations. Computer simulations can be much faster or less costly than running physical experiments. Furthermore, physical experiments can be difficult to conduct as in the detonation of explosive materials or even infeasible when only rare events like land slide or hurricane are observed. Therefore computer simulations can be a stand-alone tool or combined with (typically smaller) data from physical experiments or field observations. There are many successful applications of computer experiments as reported in the literature. For a review of the general methodologies and examples, see the books by Santner, Williams and Notz (2003), and Fang, Li and Sudjianto (2005), and the November 2009 issue of *Technometrics*, which was devoted to computer experiments.

In this paper we consider the situations in which both physical experiments/observations and computer simulations are conducted and some input variables in the computer code are either unknown or unmeasured in

---

*AMS 2000 subject classifications:* Primary 62P30, 62A01; secondary 62F12

*Keywords and phrases:* Computer Experiments, Uncertainty Quantification, Gaussian Process, Reproducing Kernel Hilbert Space

the physical experiment. We refer to them as *calibration parameters*. From the responses in physical experiments alone, we cannot estimate the true values of the calibration parameters. We can run the computer codes by choosing selected values of the calibration parameters. From the combined data of the two sources, we can make inference about the parameters. That is, we can use the physical responses to *calibrate* the computer model. An example is to use hydrological models to predict ground water flow like the level of discharge from a stream after rainfall (Kennedy and O'Hagan, 2001). Two calibration parameters in the computer model are  $T_0$ , the average transmissivity of the soil and  $m$ , the level of subsurface drainage. In the physical measurements, the response is the level of discharge and the control variables include rainfall data but neither  $T_0$  nor  $m$  is observable because they are part of the complex geological structures. In their pioneering work Kennedy and O'Hagan (2001) proposed a model to link the two data sources by employing Gaussian process models, which are commonly used in the computer experiments literature. Since its publication, this approach has received a great deal of attention in the statistical literature. See Higdon et al. (2004, 2008), Bayarri et al. (2007a,b), Joseph and Melkote (2009), Wang, Chen and Tsui (2009), Han, Santner and Rawlinson (2009), among others. It has seen a variety of applications, including hydrology, radiological protection, cylinder implosion, spot welding, micro-cutting and climate prediction, which were reported in the papers mentioned above and also in Goldstein and Rougier (2004) and Murphy et al. (2007). In spite of its importance as a methodology and significant practical impact, there has been no theoretical research on its modeling and estimation strategies. The *main purpose* of this paper is to provide a theoretical framework that facilitates the study of issues of parametrization, estimation and modeling in the Kennedy-O'Hagan formulation. For simplicity, we shall refer to Kennedy-O'Hagan as KO in the rest of the paper.

The paper is organized as follows. Some basic notation and terminology are given in the next subsection. A new theoretical framework for the calibration problem and its connection to function approximation via Gaussian process modeling are given in Section 2. In particular, the lack of identifiability of the calibration parameters is discussed and a well-defined notion of calibration parameters is proposed by using the  $L_2$  distance projection. The KO modeling strategy is discussed in Section 3. In order to provide a clean and workable mathematical analysis, we will consider some simplifications of their original formulation. One is to drop the prior, which should not affect the general conclusions of our work because the information in the prior becomes negligible as the data gets larger. Thus we shall refer to calibra-

tion based on this simplification as the *likelihood calibration*. A key result is Theorem 1, which states that the likelihood calibration is asymptotically inconsistent according to our definition of the true calibration parameters. A numerical example is given to show that the inconsistency can have a dramatic effect in small samples and some mathematical analysis is given to shed some light on why this happens. To rectify the inconsistency problem, a modification of the likelihood calibration is proposed in Section 4.1 by introducing a scale parameter into its correlation function. When the scale parameter converges to  $+\infty$  at a certain rate, calibration consistency is restored (see Theorem 2). The convergence rate of the original (unmodified) likelihood calibration is given in Theorem 3. To achieve both calibration consistency and optimal convergence rate, we introduce a new method called least  $L_2$  distance calibration in Section 4.3 and prove such properties in Theorem 4 for the case of cheap code, i.e., when the computer code can be evaluated with no cost. Its extension to expensive code is given in Theorem 5 of Section 5. The convergence rate is slower than that in Theorem 4 because, in expensive code, there is cost in evaluating the code and an unknown function  $y^s$  associated with the code needs to be estimated from data. Concluding remarks are given in Section 6. Technical details are given in two appendices. Throughout the paper, mathematical tools and results in native space (Wendland, 2005) are extensively used.

1.1. *Notation and Terminology.* Suppose we have a physical system with a vector of control variables  $x$  as its input. Denote the response of this system by  $y^p(x)$  given  $x$ . We suppose  $y^p(\cdot)$  is a real-valued function. Denote the input domain of  $x$  by  $\Omega$ , which is a convex and compact subset of  $\mathbf{R}^d$ . We call the physical system *deterministic* if  $y^p(x)$  is a fixed value for each  $x \in \Omega$ , and *stochastic* if  $y^p(x)$  is random for some  $x$ . Each run of the physical system is costly. To study the response surface, we conduct experiments on a selected set of points  $\{x_1, \dots, x_n\}$ . The set  $\mathcal{D} = \{x_1, \dots, x_n\}$  is called the experimental design or *design* for brevity.

We also have a computer code to simulate the physical system. The input of this computer code consists of two types of variables: the control variable  $x$  and the *calibration variable*  $\theta$ . The latter represents inherent attributes of the physical system, which cannot be controlled in the physical experiment. Denote the input domain for  $\theta$  by  $\Theta$ , which is a compact subset of  $\mathbf{R}^q$ . The computer code gives a deterministic function of  $x$  and  $\theta$ , denoted as  $y^s(x, \theta)$ .

Computer experiments are much less costly than the corresponding physical experiments. An ideal case is that a computer run only takes a short time so that we can run the computer code as many times as we want. Math-

ematically speaking, we call a computer code *cheap* if the functional form for  $y^s$  is *known*. However, computer runs may be time-consuming so that we can only evaluate  $y^s$  on a set of design points. In this case, an estimate  $\hat{y}^s(\cdot)$  based on the observed  $y^s$  values (and the corresponding input values) is needed and we call the computer code *expensive*.

For any convex region  $\Omega \subset \mathbf{R}^d$ , let  $C(\Omega)$  be the set of continuous functions over  $\Omega$ . For a multiple index  $\alpha = (\alpha_1, \dots, \alpha_d)$ , define  $|\alpha| = \alpha_1 + \dots + \alpha_d$ . Given  $x = (x_1, \dots, x_d)$  and function  $f$ , we denote

$$D^\alpha f := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} f.$$

For integer  $k > 0$ , define  $C^k(\Omega) = \{f \in C(\Omega) : D^\alpha f \in C(\Omega) \text{ for } |\alpha| \leq k\}$ . The  $L_2$  norm of  $f$  over  $\Omega$  is defined by  $\|f\|_{L_2(\Omega)} = (\int_\Omega f^2)^{1/2}$ .

**2. Calibration Problem and Functional Approximation Methods.** In many cases, the true value of the calibration parameters cannot be measured physically. For instance, material properties like porosity and permeability are important computer inputs in computational material simulations, which cannot be measured directly in physical experiments. A standard approach to identify those parameters is to use physical responses to adjust the computer outputs. Use of the physical response and computer output to estimate the calibration parameters is referred to as *calibration*.

**2.1.  $L_2$  Distance Projection.** In this article, we suppose the physical system is *deterministic*. Then the calibration problem can be formulated as

$$(2.1) \quad y^p(x) = y^s(x, \theta_0) + \delta(x),$$

where  $\theta_0$  is the true value of the calibration parameter and  $\delta$  is the *discrepancy* between  $y^p$  and  $y^s(\cdot, \theta_0)$ . Usually  $\delta$  is a *nonzero* function because the computer code is built based on certain assumptions or simplifications which do not match the reality exactly.

One may want to estimate the true value  $\theta_0$  in (2.1). Unfortunately, because of the existence of the unknown bias term  $\delta(\cdot)$ ,  $\theta_0$  in (2.1) is *unidentifiable*. This identifiability issue is discussed in [Bayarri et al. \(2007a,b\)](#); [Han, Santner and Rawlinson \(2009\)](#) and other papers.

The main purpose of this section is to provide a rigorous theoretical study on the estimation of calibration parameters. Given the lack of identifiability for  $\theta$ , we need to find a well-defined parameter in order to study the estimation problem. A standard approach when the model parameters are

unidentifiable is to redefine the true parameter value as one that minimizes the “distance” between the model and the observed data. First define

$$(2.2) \quad \epsilon(x, \theta) := y^p(x) - y^s(x, \theta).$$

Here we adopt the  $L_2$  norm in defining the distance. If another distance measure is chosen, the proposed mathematical framework can be pursued in a similar manner.

DEFINITION 1. The  $L_2$  distance projection of  $\theta$  is given by

$$(2.3) \quad \theta^* = \operatorname{argmin}_{\theta \in \Theta} \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)},$$

where  $\epsilon$  is defined in (2.2). For brevity, we will also refer to it as the  $L_2$  projection.

In Definition 1 we find a  $\theta$  value that minimizes the  $L_2$  discrepancy between the physical and computer observations, because the true value of  $\theta$  is not estimable. The value  $\theta^*$  given by (2.3) minimizes the average predictive error given by the computer code. This is relevant and useful since our interest lies in the prediction of the physical response. One justification of choosing the  $L_2$  norm comes from the common use of the quadratic loss criterion. Suppose we want to predict the physical response at a new point  $x_0$  and  $x_0$  is uniformly distributed over  $\Omega$ . Then the expected quadratic loss given  $\theta$  is

$$(2.4) \quad \int_{\Omega} (y^p(x) - y^s(x, \theta))^2 dx = \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2.$$

Thus the  $\theta$  value minimizing (2.4) is also the  $L_2$  distance projection  $\theta^*$ .

Since the functional forms for  $y^p$  and  $y^s$  are unknown,  $\theta^*$  cannot be obtained by solving (2.3). Given a sequence of designs  $\{\mathcal{D}_n\}_{n=1}^{\infty}$ , an estimator  $\hat{\theta}(\mathcal{D}_n)$  is said to be *consistent* if  $\hat{\theta}(\mathcal{D}_n)$  tends to  $\theta^*$ .

2.2. *Kernel Interpolation and Gaussian Process Prediction.* To estimate the calibration parameter, it is important to reconstruct the unknown function  $y^p$  based on the observations  $y^p(x_1), \dots, y^p(x_n)$ . In this article, we consider the method of *kernel interpolation* (Fasshauer, 2011). This method provides a good functional approximation when the design  $\mathcal{D}$  consists of *scattered* points, i.e., the design points are not assumed to have any regular structure.

Suppose  $y$  is a smooth function over  $\Omega$  and  $y(x_1), \dots, y(x_n)$  are observed. A kernel interpolator  $\hat{y}$  is built as follows. First choose a symmetric positive

definite function  $\Phi(\cdot, \cdot)$  over  $\Omega \times \Omega$ . Two common choices for  $\Phi$  are the Gaussian family, with

$$(2.5) \quad \Phi(s, t; \phi) = \exp\{-\phi\|s - t\|\}$$

and the Matérn family (Stein, 1999), with

$$(2.6) \quad \Phi(s, t; \nu, \phi) = \frac{1}{\Gamma(\nu)2^{\nu-1}} (2\sqrt{\nu}\phi\|s - t\|)^{\nu} K_{\nu}(2\sqrt{\nu}\phi\|s - t\|),$$

where  $K_{\nu}$  is the modified Bessel function of the second kind. Let  $\Phi = (\Phi(x_i, x_j))_{ij}$ . Since the function  $\Phi$  is positive definite, the matrix  $\Phi$  is also positive definite. Thus the linear system about  $u = (u_1, \dots, u_n)^T$

$$(2.7) \quad Y = \Phi u$$

has a unique solution  $u = \Phi^{-1}Y$ , where  $Y = (y(x_1), \dots, y(x_n))^T$ . For  $x \in \Omega$ , let

$$(2.8) \quad \hat{y}(x) = \sum_{i=1}^n u_i \Phi(x, x_i).$$

We call the kernel  $\Phi$  *stationary* if  $\Phi(x_1, x_2)$  depends only on the difference  $x_1 - x_2$ . Another special case of the kernel interpolation is the *radial basis function interpolation* (Buhmann, 2003), in which the kernel function  $\Phi(x_1, x_2)$  depends only on the distance  $\|x_1 - x_2\|$  as in (2.5) or (2.6).

In computer experiments, *Gaussian process models*, also known as *kriging*, are widely used as surrogate models for unknown functions (Sacks et al., 1989). There is a known relationship between the kernel interpolation and the Gaussian process prediction (Anjyo and Lewis, 2011). Suppose  $z(x)$  is a Gaussian process on  $\Omega$  with mean 0 and covariance function  $\Phi$ . Then given  $Z = (z(x_1), \dots, z(x_n))^T$ , the predictive mean of  $z(x)$  for any  $x$  is (Santner, Williams and Notz, 2003)

$$(2.9) \quad E[z(x)|Z] = \Phi(x, \mathbf{x})^T \Phi^{-1}Z,$$

where  $\Phi(x, \mathbf{x}) = (\Phi(x, x_1), \dots, \Phi(x, x_n))^T$ . It can be seen that  $\hat{y}$  in (2.8) has the same form as the predictive mean in (2.9).

**3. Kennedy-O'Hagan Method.** Kennedy and O'Hagan (2001) was the first to propose a Bayesian framework for the estimation of the calibration parameters. The original version of the Kennedy-O'Hagan method works for stochastic systems with expensive computer codes. Denote the

physical response by  $y^p(x_i)$ , for  $i = 1, \dots, n$ . Suppose  $y^p(x_i) = \zeta(x_i) + e_i$ , where  $\zeta(\cdot)$  is an underlying true process and  $e_i$  are the measurement errors with 0 mean. The interest lies in estimating the true value of the calibration parameter  $\theta_0$  in (2.1). Kennedy and O'Hagan (2001) proposed the following link function for  $y^s(\cdot, \theta_0)$  and  $\zeta(\cdot)$ :

$$\zeta(\cdot) = \rho y^s(\cdot, \theta_0) + \delta(\cdot),$$

where  $\rho$  is an unknown regression coefficient,  $\delta(\cdot)$  is an unknown discrepancy function. Thus  $y^p$  and  $y^s$  are related via the model:

$$(3.1) \quad y^p(x) = \rho y^s(x, \theta_0) + \delta(x) + e.$$

As is typically done in the kriging literature, they assume that  $y^s(\cdot, \cdot)$  and  $\delta(\cdot)$  are independent realizations of Gaussian processes. In kriging modeling, we usually choose the Gaussian or Matérn correlation functions (see (2.5) and (2.6)) and regard the parameters like  $\phi$  and  $\nu$  as *unknown* model parameters. Then  $\theta_0$  can be estimated from (3.1) through a Bayesian approach.

By assuming a Gaussian process prior, i.e., by imposing constraints on some parameters in the model, Kennedy-O'Hagan's approach circumvents the identifiability problem. However, our later discussions (see Example 1) will show that the use of such priors may lead to unreasonable estimation of the calibration parameters.

**3.1. Simplified KO calibration.** In order to conduct a rigorous mathematical analysis, we make the following simplifications to the Kennedy-O'Hagan method. We refer to this simplified version as the simplified KO calibration.

- (i) The physical system is deterministic and the computer code is cheap.
- (ii) Suppose  $\rho = 1$ . The discrepancy function  $\delta$  is a realization of a Gaussian process with mean 0 and the covariance function  $\sigma^2\Phi$ , where  $\sigma^2$  is an unknown parameter and the function  $\Phi$  is *known*.
- (iii) Maximum likelihood estimation (MLE) is used to estimate  $(\theta, \sigma^2)$  instead of Bayesian analysis.

The assumption (i) on cheap code will be relaxed for the  $L_2$  calibration in Section 5. In assumption (iii), we only consider the likelihood portion of the Bayesian formulation in order to have a clean and workable mathematical analysis. As will be discussed later (after (3.5)), this simplification should not affect the general message one may draw regarding the original Bayesian formulation. Not to break the flow, further comments and justifications for the assumptions will be deferred to the concluding section.

Under these assumptions, the functions  $\epsilon(x_i, \cdot)$  are known for  $i = 1, \dots, n$ . Then it can be shown that the log-likelihood function for  $(\theta, \sigma^2)$  is given by

$$(3.2) \quad l(\theta, \sigma^2; Y) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |\mathbf{\Phi}| - \frac{1}{2\sigma^2} \epsilon(\mathbf{x}, \theta)^T \mathbf{\Phi}^{-1} \epsilon(\mathbf{x}, \theta),$$

where  $\epsilon(\mathbf{x}, \theta) = (\epsilon(x_1, \theta), \dots, \epsilon(x_n, \theta))^T$  and  $\mathbf{\Phi} = (\Phi(x_i, x_j))_{ij}$ .

Our study will employ the reproducing kernel Hilbert spaces (also called the native spaces) as the mathematical tool (Wendland, 2005). Given symmetric and positive definite function  $\Phi$ , define the linear space

$$F_{\Phi}(\Omega) = \left\{ \sum_{i=1}^N \beta_i \Phi(\cdot, x_i) : N \in \mathbb{N}, \beta_i \in \mathbf{R}, x_i \in \Omega \right\}$$

and equip this space with the bilinear form

$$\left\langle \sum_{i=1}^N \beta_i \Phi(\cdot, x_i), \sum_{j=1}^M \gamma_j \Phi(\cdot, y_j) \right\rangle_{\Phi} := \sum_{i=1}^N \sum_{j=1}^M \beta_i \gamma_j \Phi(x_i, y_j).$$

Define the native space  $\mathcal{N}_{\Phi}(\Omega)$  as the *closure* of  $F_{\Phi}(\Omega)$  under the inner product  $\langle \cdot, \cdot \rangle_{\Phi}$ . The inner product of  $\mathcal{N}_{\Phi}(\Omega)$ , denoted as  $\langle \cdot, \cdot \rangle_{\mathcal{N}_{\Phi}(\Omega)}$ , is induced by  $\langle \cdot, \cdot \rangle_{\Phi}$ . Define the native norm as  $\|f\|_{\mathcal{N}_{\Phi}(\Omega)} = \sqrt{\langle f, f \rangle_{\mathcal{N}_{\Phi}(\Omega)}}$ . Some required properties of reproducing kernel Hilbert spaces are given in Appendix A under Propositions 1-5. Their equations are numbered as (A.1) to (A.7).

Direct calculation shows that the maximum likelihood estimation (MLE) for  $\theta$  is

$$(3.3) \quad \hat{\theta}_{Lik} = \underset{\theta \in \Theta}{\operatorname{argmin}} \epsilon(\mathbf{x}, \theta)^T \mathbf{\Phi}^{-1} \epsilon(\mathbf{x}, \theta).$$

For fixed  $\theta$ , let  $\hat{\epsilon}(\cdot, \theta)$  be the kernel interpolator for  $\epsilon(\cdot, \theta)$  given by (2.8), i.e.,

$$(3.4) \quad \hat{\epsilon}(\cdot, \theta) = \Phi(\cdot, \mathbf{x})^T \mathbf{\Phi}^{-1} \epsilon(\mathbf{x}, \theta).$$

From the definition of the native norm, we have

$$\|\hat{\epsilon}(\cdot, \theta)\|_{\mathcal{N}_{\Phi}(\Omega)}^2 = \epsilon(\mathbf{x}, \theta)^T \mathbf{\Phi}^{-1} \mathbf{\Phi} \mathbf{\Phi}^{-1} \epsilon(\mathbf{x}, \theta) = \epsilon(\mathbf{x}, \theta)^T \mathbf{\Phi}^{-1} \epsilon(\mathbf{x}, \theta),$$

which, together with (3.3), gives

$$(3.5) \quad \hat{\theta}_{Lik} = \underset{\theta \in \Theta}{\operatorname{argmin}} \|\hat{\epsilon}(\cdot, \theta)\|_{\mathcal{N}_{\Phi}(\Omega)}^2.$$

Noting that  $\hat{\theta}_{Lik}$  is based on maximizing the likelihood function, we should refer to calibration based on  $\hat{\theta}_{Lik}$  as the likelihood-based calibration or *likelihood calibration* for brevity. Its main difference from the Bayesian formulation is the presence of the prior, whose information will be overwhelmed by that in the likelihood as the data get larger. Therefore, we believe the general message one can draw from our theoretical results on  $\hat{\theta}_{Lik}$  should carry to the original Kennedy-O'Hagan formulation. Further discussions will be given in Section 6.

Now we study the asymptotic property for the likelihood calibration method. To this end, we require the design points to become dense over  $\Omega$ . This property is measured by the fill distance.

DEFINITION 2. For a design  $\mathcal{D} \in \Omega^n$ , define its *fill distance* as

$$(3.6) \quad h(\mathcal{D}) := \max_{x \in \Omega} \min_{x_i \in \mathcal{D}} \|x_i - x\|.$$

We use  $\hat{\theta}_{Lik}(\mathcal{D})$  to denote the estimator  $\hat{\theta}_{Lik}$  under  $\mathcal{D}$ . Theorem 1 gives the limiting value of  $\hat{\theta}_{Lik}(\mathcal{D})$ .

THEOREM 1. Suppose there exists  $v_\theta \in L_2(\Omega)$ , such that  $\epsilon(x, \theta) = \int_\Omega \Phi(x, t)v_\theta(t)dt$  for any  $\theta \in \Theta$ . Moreover, suppose  $\sup_{\theta \in \Theta} \|v_\theta\|_{L_2} < +\infty$ ,  $\Phi$  has continuous second order derivatives, and there exists a unique  $\theta' \in \Theta$  such that

$$(3.7) \quad \|\epsilon(\cdot, \theta')\|_{\mathcal{N}_\Phi(\Omega)} = \inf_{\theta \in \Theta} \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_\Phi(\Omega)}.$$

Then  $\hat{\theta}_{Lik}(\mathcal{D}_n) \rightarrow \theta'$  as  $h(\mathcal{D}_n) \rightarrow 0$ .

PROOF. Let  $\hat{\epsilon}_n$  be the kernel interpolator for  $\epsilon$  under design  $\mathcal{D}_n$ . From (3.5) and (3.7), it suffices to prove that  $\|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_\Phi(\Omega)}^2$  converges to  $\|\epsilon(\cdot, \theta)\|_{\mathcal{N}_\Phi(\Omega)}^2$  uniformly with respect to  $\theta \in \Theta$ . From Proposition 1, we have

$$(3.8) \quad \begin{aligned} \|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_\Phi(\Omega)}^2 - \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_\Phi(\Omega)}^2 &= \|\hat{\epsilon}_n(\cdot, \theta) - \epsilon(\cdot, \theta)\|_{\mathcal{N}_\Phi(\Omega)}^2 \\ &\leq C_\Phi^2 h^2(\mathcal{D}_n) \|v_\theta\|_{L_2(\Omega)}^2 \\ &\leq C_\Phi^2 h^2(\mathcal{D}_n) \sup_{\theta \in \Theta} \|v_\theta\|_{L_2(\Omega)}^2, \end{aligned}$$

where the first equality follows from the identity  $\langle \epsilon(\cdot, \theta), \hat{\epsilon}_n(\cdot, \theta) - \epsilon(\cdot, \theta) \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$ . The right side of (3.8) goes to 0 as  $n \rightarrow \infty$  and is independent of  $\theta$ . This completes the proof.  $\square$

Because in general  $\theta' \neq \theta^*$ , according to Definition 1, the likelihood calibration is *not* consistent. In view of the convergence result in Theorem 1, it is important to find out *how different the limiting value  $\theta'$  of the likelihood calibration is from  $\theta^*$* . To address this question, we consider the difference between the two norms  $\|\cdot\|_{L_2(\Omega)}$  and  $\|\cdot\|_{\mathcal{N}_\Phi(\Omega)}$ . This difference is related to the eigenvalues of the integral operator defined as

$$(3.9) \quad \kappa(f) = \int_{\Omega} \Phi(\cdot, x) f(x) dx,$$

for  $f \in L_2(\Omega)$ . Denote the eigenvalues of  $\kappa$  by  $\lambda_1 \geq \lambda_2 \geq \dots$ . Let  $f_i$  be the eigenfunction associated with  $\lambda_i$  with  $\|f_i\|_{L_2(\Omega)} = 1$ . Then it can be shown that

$$(3.10) \quad \|f_i\|_{\mathcal{N}_\Phi(\Omega)}^2 = \langle f_i, \lambda_i^{-1} f_i \rangle_{L_2(\Omega)}^2 = \lambda_i^{-1},$$

where the first equality follows from (A.2) and the fact that  $\kappa(\lambda_i^{-1} f_i) = f_i$ . It is known in functional analysis that  $\kappa$  is a compact operator and therefore  $\lim_{k \rightarrow \infty} \lambda_k = 0$  (Conway, 1990). Thus (3.10) yields that  $\|f_i\|_{\mathcal{N}_\Phi(\Omega)}^2 / \|f_i\|_{L_2(\Omega)}^2 = \lambda_i^{-1} \rightarrow \infty$  as  $i \rightarrow \infty$ . This leads to

$$(3.11) \quad \sup_{f \in \mathcal{N}_\Phi^1(\Omega)} \frac{\|f\|_{\mathcal{N}_\Phi(\Omega)}}{\|f\|_{L_2(\Omega)}} = \infty,$$

which implies that there are functions  $f$  with arbitrarily small  $L_2$  norm while their  $\mathcal{N}_\Phi$  norm is bounded away from zero. Therefore, by Definition 1, the likelihood calibration can give results that are *far from* the  $L_2$  distance projection. The following example shows that this effect can be dramatic.

EXAMPLE 1. Consider a calibration problem with a three-level calibration parameter. Let  $\Omega = [-1, 1]$ ,  $\Phi(x_1, x_2) = \exp\{-(x_1 - x_2)^2\}$ . By using a numerical method, we obtain the eigenvalue and eigenfunction of  $\kappa$  defined in (3.9). The first and second eigenvalues are  $\lambda_1 = 1.546$  and  $\lambda_2 = 0.398$ . For a better visual effect, we use the eigenfunctions whose  $L_2$  norms are  $\sqrt{20}$ . We plot the first and second eigenfunctions of  $\kappa$  in Figure 1. We also plot  $\sin 2\pi x$  for later comparison. Suppose we have three different computer codes. Denote the discrepancy between the physical response and each of the computer output by  $\epsilon_1$ ,  $\epsilon_2$  and  $\epsilon_3$  respectively. Suppose  $\epsilon_1, \epsilon_2, \epsilon_3$  are the three functions given in Figure 1, i.e.,  $\epsilon_1$  and  $\epsilon_2$  are the first and second eigenfunction of  $\kappa$ , and  $\epsilon_3$  is  $\sin 2\pi x$ . Then which code is the best? From  $\|\epsilon_1\|_{L_2(\Omega)} = \|\epsilon_2\|_{L_2(\Omega)} = \sqrt{20}$  and  $\|\epsilon_3\|_{L_2(\Omega)} = 1$ , the third computer code is the best according to Definition 1.

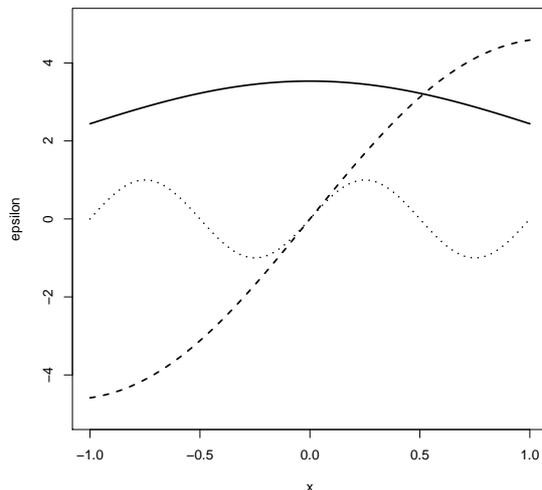


FIG 1. Three functions considered in Example 1. The solid and dashed lines are the first and second eigenfunction of  $\kappa$  with  $L_2$  norm  $\sqrt{20}$  respectively. The dotted line shows function  $\sin 2\pi x$ .

However, by using a Gaussian process model with the same correlation function  $\Phi$ , we get a different result. By (3.3), maximizing the likelihood function is equivalent to minimizing the pivoted sum of square (PSS):  $\varepsilon_i^T \Phi^{-1} \varepsilon_i$ , where  $\varepsilon_i = (\varepsilon_i(x_1), \dots, \varepsilon_i(x_n))^T$  for  $i = 1, 2, 3$ . We choose a space-filling design of 11 points, given by  $x_j = -1 + (j - 1)/5$  for  $j = 1, \dots, 11$ . The PSSs are 12.594 for  $i = 1$ , 57.908 for  $i = 2$ , and 17978.65 for  $i = 3$ . Thus the likelihood calibration will choose the first code because it has the smallest PSS value. This demonstrates that the likelihood-based method can give very different rankings of the competing codes from the  $L_2$  projection. From Figure 1 we can see that  $|\varepsilon_3(x)|$  is smaller than  $|\varepsilon_1(x)|$  and  $|\varepsilon_2(x)|$  for every  $x$ , i.e., the point-wise prediction error for the third code is uniformly smaller than the first two. Therefore, the likelihood calibration made a wrong choice. This also gives a good justification for choosing the  $L_2$  norm rather than the native norm in Definition 1.

To give a more general explanation for the phenomenon in Example 1, we quote a result in Wendland (2005, p. 170, p. 201). It states that for the Matérn correlation functions defined by (2.6), the native space  $\mathcal{N}_\Phi(\Omega)$  equals to the (fractional) Sobolev space  $H^{v+d/2}(\Omega)$  and the two norms are

equivalent if  $v \geq 1$ , where the Sobolev norm is defined as

$$\|f\|_{H^k(\Omega)}^2 = \|f\|_{L_2(\Omega)}^2 + \sum_{i=1}^k \left\| \frac{\partial^i f}{\partial x^i} \right\|_{L_2(\Omega)}^2$$

for positive integer  $k$  and  $d = 1$ . Note that the Sobolev norm can be big for a function with wild oscillation even when its  $L_2$  norm is small (the same as that shown by (3.11)). Thus, the Sobolev norm, which is equivalent to the native norm in this context, is not a good measure of discrepancy, because we only care about the magnitude of the discrepancy, not its oscillation. Therefore it is not suitable to use the likelihood calibration in most practical problems.

*3.2. Numerical Study on Kennedy-O'Hagan Method with Estimated Correlation Function.* To give a more realistic comparison, we extend the study in Example 1 by considering the frequentist version of the original Kennedy-O'Hagan method, in which the function  $\Phi$  is estimated as well. Specifically, we suppose that  $\Phi$  depends on a model parameter  $\phi$ , denoted as  $\Phi_\phi(\cdot, \cdot)$ . Then the log-likelihood function is

$$(3.12) \quad l(\theta, \sigma^2, \phi; Y) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |\Phi_\phi| - \frac{1}{2\sigma^2} \epsilon(\mathbf{x}, \theta)^T \Phi_\phi^{-1} \epsilon(\mathbf{x}, \theta),$$

where  $\Phi_\phi = (\Phi_\phi(x_i, x_j))_{ij}$ . By substituting the analytical form of  $\hat{\sigma}^2$  into (3.12), we obtain the log-likelihood function with respect to  $(\theta, \phi)$ :

$$(3.13) \quad l(\theta, \phi; Y) = -\frac{n}{2} \log \epsilon(\mathbf{x}, \theta)^T \Phi_\phi^{-1} \epsilon(\mathbf{x}, \theta) - \frac{1}{2} \log |\Phi_\phi|.$$

As in Section 3.1, we estimate  $(\theta, \phi)$  by using the maximum likelihood. In this section we present some numerical results, which show that even when  $\Phi$  is estimated, the frequentist KO method still suffers from the problem previously demonstrated in Example 1.

EXAMPLE 1 (continued). We use the same true functions  $\epsilon_1, \epsilon_2, \epsilon_3$  and design points as in Example 1. We compute the log-likelihood functions given in (3.13) with  $\Phi_\phi(x_1, x_2) = \exp\{-\phi(x_1 - x_2)^2\}$ . Denote the log-likelihood function in (3.13) by  $l(\theta, \phi)$ , where  $\theta = 1, 2, 3$  correspond to the candidate functions  $\epsilon_1, \epsilon_2, \epsilon_3$ . The functions  $l(\theta, \cdot)$  are plotted in Figure 2 for  $\theta = 1, 2, 3$ . From the figure we can see that  $\sup_{\theta \in \{1, 2, 3\}, \phi \in [1, 6]} l(\theta, \phi) = l(1, 1)$ . Therefore the frequentist KO method will pick  $\epsilon_1$ , which gives the same (incorrect) result as in Example 1.

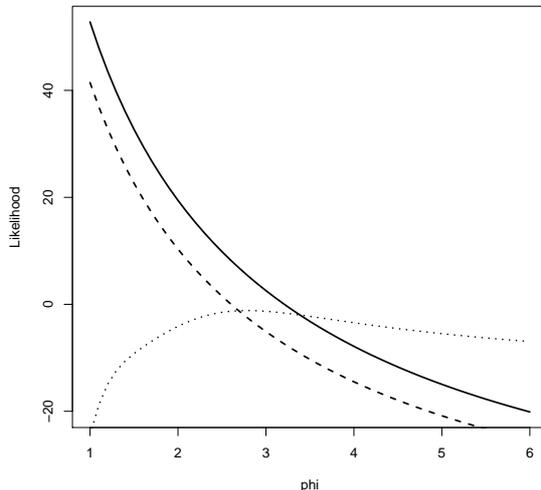


FIG 2. Log-likelihood functions for the three functions  $\epsilon_1, \epsilon_2, \epsilon_3$  given in Example 1 are plotted in solid, dashed and dotted lines respectively.

**4. Asymptotic Results: Cheap Code.** Theorem 1 is the first asymptotic result we present in this article. In this section we shall study other convergence properties, assuming that the computer code is cheap as in Section 3.

4.1. *Modified Likelihood Calibration.* Given the wide-spread use of the Gaussian process modeling in calibration problems (as in the KO method), a fundamental question is whether we can modify it to rectify its inconsistency problem. For convenience we assume a stationary Gaussian process model  $Y(x)$ . The correlation of  $Y$  is given by

$$R(x) = \text{Corr}(Y(\cdot + x), Y(\cdot)),$$

where  $R$  is a positive definite kernel over  $\mathbf{R}^d$ . The *Fourier transform* (Stein and Weiss, 1971) is a useful tool for studying stationary kernels. We will use the notation  $\mathcal{N}_R(\Omega)$  instead of  $\mathcal{N}_{R(\cdot, \cdot)}(\Omega)$  for simplicity.

DEFINITION 3. For  $f \in L_1(\mathbf{R}^d)$  define the Fourier transform by

$$\tilde{f}(w) := (2\pi)^{-d/2} \int_{\mathbf{R}^d} f(x) e^{-iw^T x} dx.$$

From Theorem 1, it can be seen that the likelihood calibration is not consistent if the correlation function  $R$  is fixed. In order to construct consistent calibration, we should use a sequence of  $R$  functions indexed by  $n$ , denoted by  $R_n$ . From (A.2), the  $\|\cdot\|_{\mathcal{N}_\Phi(\Omega)}$  norm becomes  $\|\cdot\|_{L_2}$  only when  $\Phi(x_1, x_2) = \delta(x_1 - x_2)$ , where  $\delta$  denotes the Dirac delta function. We need the convergence  $R_n(x) \rightarrow \delta(x)$  in order to obtain consistency. An easy way to achieve this convergence is to introduce a scale parameter. Suppose  $R(\cdot; \phi)$  is a family of correlation functions on  $\mathbf{R}^d$  with  $\phi > 0$ . Call  $\phi$  a *scale parameter* if  $R(x; \phi) = R(\phi x; 1)$  for any  $\phi > 0$  and any  $x \in \mathbf{R}^d$ . Most correlation families like Gaussian or Matérn family satisfy these conditions.

Write  $R_n(x) = R(x; \phi_n)$ . Let  $\hat{\theta}(R_n, \mathcal{D}_n)$  be the estimate of  $\theta$  given by the likelihood calibration using correlation function  $R_n$  under design  $\mathcal{D}_n$ , referred to as the *modified likelihood calibration*. The calibration consistency requires  $\phi_n \rightarrow \infty$ . But to ensure the convergence of the interpolation,  $\phi_n$  cannot diverge too fast. The next theorem suggests that the modified likelihood calibration is consistent if we choose a suitable increasing rate for  $\phi_n$ . Define the convolution  $f * f(x) = \int_{\mathbf{R}^d} f(x-t)f(t)dt$  for any  $f \in L_2(\mathbf{R}^d)$ . We list the required regularity conditions before stating the theorem. The lengthy proof is deferred to Appendix B.

- A1:  $\sup_{x \in \Omega, \theta \in \Theta} \|\nabla_x \epsilon(x, \theta)\| < +\infty$ , where  $\nabla_x \epsilon(x, \theta)$  is the gradient of  $\epsilon(x, \theta)$  with respect to  $x$ .
- A2: There exists  $\phi_0 > 0$  such that  $\sup_{\theta \in \Theta} \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{R(\cdot; \phi_0) * R(\cdot; \phi_0)}(D)} < +\infty$ .
- A3:  $R(\cdot; 1)$  is integrable and  $\sup_{w \neq 0, \alpha \geq 1} \tilde{R}(\alpha w) / \tilde{R}(w) < +\infty$ , where  $\tilde{R}$  is the Fourier transform of  $R(\cdot; 1)$ .

**THEOREM 2.** *Suppose conditions (A1-A3) are satisfied and  $R(\cdot; 1)$  has continuous derivatives. Then  $\hat{\theta}(R_n, \mathcal{D}_n) \rightarrow \theta^*$  if  $\phi_n \rightarrow +\infty$  and  $\phi_n h(\mathcal{D}_n) \rightarrow 0$ .*

Although the modified likelihood calibration is consistent, its implementation relies on a prespecified sequence  $\phi_n$ . For a calibration problem with a fixed sample size, there is no theoretical guidelines on choosing the  $\phi$  value. In Section 4.3, we will propose a novel calibration method that is consistent and does not rely on the choice of the kernel function.

**4.2. Convergence Rate.** Under the assumption of Theorem 1, we can employ (A.6) in Proposition 1 to show that the interpolation error given by a kernel  $\Phi$  with  $2k$  derivatives is equivalent to  $O(h^{2k}(\mathcal{D}_n))$ . Given this rate, Theorem 3 shows that the likelihood calibration, which converges to  $\theta'$  in Theorem 1, reaches the same rate. Let  $\theta = (\theta_1, \dots, \theta_q)$ .

**THEOREM 3.** *Under the conditions of Theorem 1, suppose  $\Phi$  has  $2k$  continuous derivatives. We assume that  $\theta'$  is an interior point of  $\Theta$  and there exist a neighborhood  $U \subset \Theta$  of  $\theta'$ , and functions  $D_i v_\theta, D_{ij} v_\theta \in C(\Omega)$  such that*

$$\begin{aligned}\frac{\partial \epsilon}{\partial \theta_i}(x, \theta) &= \int_{\Omega} \Phi(x, t) D_i v_\theta(t) dt, \\ \frac{\partial^2 \epsilon}{\partial \theta_i \partial \theta_j}(x, \theta) &= \int_{\Omega} \Phi(x, t) D_{ij} v_\theta(t) dt,\end{aligned}$$

for  $x \in \Omega, \theta \in U$  and  $1 \leq i, j \leq q$ . Moreover,

$$(4.1) \quad \sup_{\theta \in U, 1 \leq i, j \leq q} \{ \|D_i v_\theta\|_{L_2(\Omega)}, \|D_{ij} v_\theta\|_{L_2(\Omega)} \} < \infty, \text{ and}$$

$$\frac{\partial^2}{\partial \theta \partial \theta^T} \|\epsilon(\cdot, \theta')\|_{\mathcal{N}_\Phi(\Omega)}^2 \text{ is invertible.}$$

Then  $\|\hat{\theta}_{Lik}(\mathcal{D}_n) - \theta'\| = O(h^{2k}(\mathcal{D}_n))$ .

**PROOF.** For the proof we need some inequalities. The first one follows immediately from (A.7) in Proposition 1:

$$(4.2) \quad \|\hat{\epsilon}_n(\cdot, \theta) - \epsilon(\cdot, \theta)\|_{\mathcal{N}_\Phi(\Omega)} \leq C_\Phi h^k(\mathcal{D}_n) \|v_\theta\|_{L_2(\Omega)}.$$

From the definition of  $\hat{\epsilon}_n$  in (3.4), we have  $\frac{\partial \hat{\epsilon}_n}{\partial \theta_j}(\cdot, \theta) = \Phi(\cdot, \mathbf{x}) \Phi^{-1} \frac{\partial \epsilon}{\partial \theta_j}(\mathbf{x}, \theta)$ . Because  $\frac{\partial \hat{\epsilon}_n}{\partial \theta_j}(\cdot, \theta)$  is also spanned by the functions  $\{\Phi(\cdot, x_i)\}$ ,  $\frac{\partial \hat{\epsilon}_n}{\partial \theta_j}(\cdot, \theta)$  is equal to the kernel interpolator for the pairs  $(x_i, \frac{\partial \epsilon}{\partial \theta_j}(x_i, \theta_j))$ . As a result of (A.7) in Proposition 1, we obtain

$$(4.3) \quad \left\| \frac{\partial(\hat{\epsilon}_n - \epsilon)}{\partial \theta_i}(\cdot, \theta) \right\|_{\mathcal{N}_\Phi(\Omega)} \leq C_\Phi h^k(\mathcal{D}_n) \|D_i v_\theta\|_{L_2(\Omega)}.$$

Similarly we have

$$(4.4) \quad \left\| \frac{\partial^2(\hat{\epsilon}_n - \epsilon)}{\partial \theta_i \partial \theta_j}(\cdot, \theta) \right\|_{\mathcal{N}_\Phi(\Omega)} \leq C_\Phi h^k(\mathcal{D}_n) \|D_{ij} v_\theta\|_{L_2(\Omega)}.$$

As  $\|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_\Phi(\Omega)}^2$  is minimized at  $\hat{\theta}_{Lik}(\mathcal{D}_n)$ , the Taylor expansion of  $\frac{\partial}{\partial \theta} \|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_\Phi(\Omega)}^2$  with respect to  $\theta$  gives

$$\begin{aligned}0 &= \frac{\partial}{\partial \theta} \|\hat{\epsilon}_n(\cdot, \hat{\theta}_n)\|_{\mathcal{N}_\Phi(\Omega)}^2 \\ &= \frac{\partial}{\partial \theta} \|\hat{\epsilon}_n(\cdot, \theta')\|_{\mathcal{N}_\Phi(\Omega)}^2 + \left( \frac{\partial^2}{\partial \theta \partial \theta^T} \|\hat{\epsilon}_n(\cdot, \hat{\theta}_n)\|_{\mathcal{N}_\Phi(\Omega)}^2 \right) (\hat{\theta}_{Lik}(\mathcal{D}_n) - \theta'),\end{aligned}$$

where  $\tilde{\theta}_n$  is located between  $\theta'$  and  $\hat{\theta}_{Lik}(\mathcal{D}_n)$ . By Theorem 1,  $\tilde{\theta}_n \rightarrow \theta'$ . Thus

$$(4.5) \quad \begin{aligned} & \hat{\theta}_{Lik}(\mathcal{D}_n) - \theta' \\ &= - \left( \frac{\partial^2}{\partial \theta \partial \theta^T} \|\hat{\epsilon}_n(\cdot, \tilde{\theta}_n)\|_{\mathcal{N}_\Phi(\Omega)}^2 \right)^{-1} \frac{\partial}{\partial \theta} \|\hat{\epsilon}_n(\cdot, \theta')\|_{\mathcal{N}_\Phi(\Omega)}^2, \end{aligned}$$

where  $\frac{\partial^2}{\partial \theta \partial \theta^T} \|\hat{\epsilon}_n(\cdot, \tilde{\theta}_n)\|_{\mathcal{N}_\Phi(\Omega)}^2$  is invertible because of assumption (4.2) and the fact that  $\tilde{\theta}_n \rightarrow \theta'$ . Furthermore,

$$(4.6) \quad \begin{aligned} & \frac{\partial^2}{\partial \theta_i \partial \theta_j} \|\hat{\epsilon}_n(\cdot, \tilde{\theta}_n)\|_{\mathcal{N}_\Phi(\Omega)}^2 - \frac{\partial^2}{\partial \theta_i \partial \theta_j} \|\epsilon(\cdot, \tilde{\theta}_n)\|_{\mathcal{N}_\Phi(\Omega)}^2 \\ &= \frac{\partial^2}{\partial \theta_i \partial \theta_j} \|\hat{\epsilon}_n(\cdot, \tilde{\theta}_n) - \epsilon(\cdot, \tilde{\theta}_n)\|_{\mathcal{N}_\Phi(\Omega)}^2 \\ &= 2 \left\{ \left\langle \frac{\partial^2(\hat{\epsilon}_n - \epsilon)}{\partial \theta_i \partial \theta_j}(\cdot, \tilde{\theta}_n), (\hat{\epsilon}_n - \epsilon)(\cdot, \tilde{\theta}_n) \right\rangle \right. \\ & \quad \left. + \left\langle \frac{\partial(\hat{\epsilon}_n - \epsilon)}{\partial \theta_i}(\cdot, \tilde{\theta}_n), \frac{\partial(\hat{\epsilon}_n - \epsilon)}{\partial \theta_j}(\cdot, \tilde{\theta}_n) \right\rangle \right\} \\ &\leq 2 \left\{ \|\hat{\epsilon}_n(\cdot, \tilde{\theta}_n) - \epsilon(\cdot, \tilde{\theta}_n)\|_{\mathcal{N}_\Phi(\Omega)} \left\| \frac{\partial^2(\hat{\epsilon}_n - \epsilon)}{\partial \theta_i \partial \theta_j}(\cdot, \tilde{\theta}_n) \right\|_{\mathcal{N}_\Phi(\Omega)} \right. \\ & \quad \left. + \left\| \frac{\partial(\hat{\epsilon}_n - \epsilon)}{\partial \theta_i}(\cdot, \tilde{\theta}_n) \right\|_{\mathcal{N}_\Phi(\Omega)} \left\| \frac{\partial(\hat{\epsilon}_n - \epsilon)}{\partial \theta_j}(\cdot, \tilde{\theta}_n) \right\|_{\mathcal{N}_\Phi(\Omega)} \right\}, \end{aligned}$$

where the first equality follows from (A.4). Invoking (4.2)-(4.4), and the condition  $h(\mathcal{D}_n) \rightarrow 0$  in Theorem 1, (4.6) tends to 0. This results in

$$(4.7) \quad \frac{\partial^2}{\partial \theta \partial \theta^T} \|\hat{\epsilon}_n(\cdot, \tilde{\theta}_n)\|_{\mathcal{N}_\Phi(\Omega)}^2 \rightarrow \frac{\partial^2}{\partial \theta \partial \theta^T} \|\epsilon(\cdot, \theta')\|_{\mathcal{N}_\Phi(\Omega)}^2,$$

because  $\tilde{\theta}_n$  tends to  $\theta'$ . By the definition of  $\theta'$  in (3.7) and the assumption that  $\theta'$  is an interior point of  $\Theta$ ,  $\frac{\partial}{\partial \theta_i} \|\epsilon(\cdot, \theta')\|_{\mathcal{N}_\Phi(\Omega)}^2 = 0$ . Therefore, we have

$$(4.8) \quad \begin{aligned} \frac{\partial}{\partial \theta_i} \|\hat{\epsilon}_n(\cdot, \theta')\|_{\mathcal{N}_\Phi(\Omega)}^2 &= \frac{\partial}{\partial \theta_i} \|\hat{\epsilon}_n(\cdot, \theta') - \epsilon(\cdot, \theta')\|_{\mathcal{N}_\Phi(\Omega)}^2 \\ &= 2 \left\langle \frac{\partial(\hat{\epsilon}_n - \epsilon)}{\partial \theta_i}(\cdot, \theta'), \hat{\epsilon}_n(\cdot, \theta') - \epsilon(\cdot, \theta') \right\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &\leq 2 \left\| \frac{\partial(\hat{\epsilon}_n - \epsilon)}{\partial \theta_i}(\cdot, \theta') \right\|_{\mathcal{N}_\Phi(\Omega)} \|\hat{\epsilon}_n(\cdot, \theta') - \epsilon(\cdot, \theta')\|_{\mathcal{N}_\Phi(\Omega)} \\ &\leq 2C_\Phi^2 h^{2k}(\mathcal{D}_n) \|v_{\theta'}\|_{L_2(\Omega)} \|D_i v_{\theta'}\|_{L_2(\Omega)} \rightarrow 0, \end{aligned}$$

where the first equality follows from (A.4); the last inequality follows from (4.2) and (4.3); and the limiting relationship follows from (4.1) and the fact that  $h(\mathcal{D}_n) \rightarrow 0$ . Then we obtain the desired result by combining (4.5), (4.7) and (4.8).  $\square$

Noting that consistency is a necessary requirement for an estimator, we would like to find an estimator that is consistent and attains the same convergence rate as in Theorem 3. In the following subsection we find an estimator that guarantees consistency and full efficiency.

4.3. *Least  $L_2$  Distance Calibration.* Let  $\hat{y}^p$  be the kernel interpolator defined in (2.8) for  $y^p$  under design  $\mathcal{D}$ . Define the *least  $L_2$  distance calibration* by

$$(4.9) \quad \hat{\theta}_{L_2}(\mathcal{D}) = \operatorname{argmin}_{\theta \in \Theta} \|\hat{y}^p(\cdot) - y^s(\cdot, \theta)\|_{L_2(\Omega)}.$$

For brevity, we will also refer to it as the  *$L_2$  calibration*. Han, Santner and Rawlinson (2009) used the  $L_2$  norm in a different context, i.e., choosing an optimal tuning parameter value to minimize the  $L_2$  norm of the observed discrepancy  $\hat{y}^p - y^s$ . Theorem 4 shows that  $\hat{\theta}_{L_2}(\mathcal{D}_n)$  converges to the  $L_2$  projection  $\theta^*$  at the optimal rate.

**THEOREM 4.** *Suppose  $\theta^*$  is the unique solution to (2.3) and an interior point of  $\Theta$ ;  $y^p \in \mathcal{N}_\Phi(\Omega)$ ;  $\frac{\partial^2}{\partial\theta\partial\theta^T} \|\epsilon(\cdot, \theta^*)\|_{L_2(\Omega)}^2$  is invertible;  $\Phi$  has  $2k$  continuous derivatives; and there exists a neighborhood  $U \subset \Theta$  of  $\theta^*$  such that  $y^s \in L_2(\Omega \times U)$  and  $y^s(x, \cdot) \in C^2(U)$  for  $x \in \Omega$ . Then as  $h(\mathcal{D}_n) \rightarrow 0$ ,*

$$(4.10) \quad \|\hat{\theta}_{L_2}(\mathcal{D}_n) - \theta^*\| = O(h^k(\mathcal{D}_n)).$$

*Furthermore, if there exists  $v \in L_2(\Omega)$  such that  $y^p(x) = \int_\Omega \Phi(x, t)v(t)dt$  for all  $x \in \Omega$ , then the convergence rate can be improved to*

$$(4.11) \quad \|\hat{\theta}_{L_2}(\mathcal{D}_n) - \theta^*\| = O(h^{2k}(\mathcal{D}_n)).$$

**PROOF.** First we prove the consistency, i.e., the convergence of  $\hat{\theta}_{L_2}(\mathcal{D}_n)$  to  $\theta^*$ . Because  $\hat{\theta}_{L_2}(\mathcal{D}_n)$  minimizes  $\|\hat{y}_n^p - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2$  and  $\theta^*$  is the unique minimizer of  $\|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2$ , it suffices to prove that  $\|\hat{y}^p - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2$  converges to  $\|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2$  uniformly with respect to  $\theta$ . Note

$$\begin{aligned} & \|\hat{y}^p - y^s(\cdot, \theta)\|_{L_2(\Omega)}^2 - \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2 \\ &= (\|\hat{y}^p - y^s(\cdot, \theta)\|_{L_2(\Omega)} - \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}) \cdot \end{aligned}$$

$$\begin{aligned}
& (\|\hat{y}^p - y^s(\cdot, \theta)\|_{L_2(\Omega)} + \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}) \\
& \leq \|\hat{y}^p - y^s(\cdot, \theta) - \epsilon(\cdot, \theta)\|_{L_2(\Omega)} (\|\hat{y}^p - y^s(\cdot, \theta)\|_{L_2(\Omega)} + \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}) \\
(4.12) \leq & \|\hat{y}^p - y^p\|_{L_2(\Omega)} (\|\hat{y}^p\|_{L_2(\Omega)} + \|y^s(\cdot, \theta)\|_{L_2(\Omega)} + \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}).
\end{aligned}$$

Because  $h(\mathcal{D}_n) \rightarrow 0$ , it can be seen from (A.5) that  $\sup |\hat{y}^p(x) - y^p(x)| \rightarrow 0$ . Together with the compactness of  $\Omega$ , we have  $\|\hat{y}^p - y^p\|_{L_2(\Omega)} = o(1)$  and  $\|\hat{y}^p\|_{L_2(\Omega)} = O(1)$ . Therefore, the uniform convergence is obtained from (4.12), and this leads to the consistency of  $\hat{\theta}_{L_2}(\mathcal{D}_n)$ .

The convergence rate can be derived by following a similar argument as in Theorem 3. Note that for any  $f \in L_2(\Omega)$ ,

$$(4.13) \quad \|f\|_{L_2} \leq \sqrt{\text{Vol}(\Omega)} \sup_{x \in \Omega} |f(x)|,$$

where  $\text{Vol}(\Omega)$  is the volume of  $\Omega$ . Because  $\|\hat{\epsilon}_n(\cdot, \theta)\|_{L_2(\Omega)}^2$  is minimized at  $\hat{\theta}_{L_2}(\mathcal{D}_n)$ ,  $\hat{\theta}_{L_2}(\mathcal{D}_n)$  tends to  $\theta^*$  and  $\theta^*$  is an interior point, the Taylor expansion gives

$$\begin{aligned}
0 &= \frac{\partial}{\partial \theta} \|\hat{y}^p - y^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2 \\
(4.14) \quad &+ \left( \frac{\partial^2}{\partial \theta \partial \theta^T} \|\hat{y}^p - y^s(\cdot, \tilde{\theta}_n)\|_{L_2(\Omega)}^2 \right) (\hat{\theta}_{L_2}(\mathcal{D}_n) - \theta^*),
\end{aligned}$$

where  $\tilde{\theta}_n$  is located between  $\theta^*$  and  $\hat{\theta}_{L_2}(\mathcal{D}_n)$ . Since  $\hat{y}^p$  is independent of  $\theta$ , it is easy to see that

$$\frac{\partial^2}{\partial \theta \partial \theta^T} \|\hat{y}^p - y^s(\cdot, \tilde{\theta}_n)\|_{L_2(\Omega)}^2 \rightarrow \frac{\partial^2}{\partial \theta \partial \theta^T} \|\epsilon(\cdot, \theta^*)\|_{L_2(\Omega)}^2,$$

which is invertible by the assumption. Therefore the convergence rate is given by

$$\begin{aligned}
& \frac{\partial}{\partial \theta_i} \|\hat{y}^p - y^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2 \\
&= \frac{\partial}{\partial \theta_i} \|\hat{y}^p - y^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2 - \frac{\partial}{\partial \theta_i} \|y^p - y^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2 \\
(4.15) \quad &= 2 \left\langle y^p - \hat{y}^p, \frac{\partial y^s}{\partial \theta_i}(\cdot, \theta^*) \right\rangle_{L_2(\Omega)}
\end{aligned}$$

$$(4.16) \quad \leq 2\sqrt{\text{Vol}(\Omega)} C_{\Phi} h^k(\mathcal{D}_n) \|y^p\|_{\mathcal{N}_{\Phi}(\Omega)} \left\| \frac{\partial y^s}{\partial \theta_i}(\cdot, \theta^*) \right\|_{L_2(\Omega)},$$

where the first equality follows from the definition of  $\theta^*$  in (2.3) and the fact that  $\theta^*$  is an interior point; the inequality follows from Schwarz's inequality,

(4.13) and (A.5) in Proposition 1. Hence (4.10) is obtained by combining (4.14), (4.15) and (4.16).

If there exists  $v \in L_2(\Omega)$  such that  $y^p(x) = \int_{\Omega} \Phi(x, t)v(t)dt$ , (A.6) in Proposition 1 can be applied to (4.15), which proves (4.11).  $\square$

By comparing the results and conditions in Theorems 3 and 4, we can make the following observations. First,  $\|\hat{\theta}_{L_{ik}}(\mathcal{D}_n) - \theta^*\|$  and  $\|\hat{\theta}_{L_2}(\mathcal{D}_n) - \theta^*\|$  enjoy the convergence rate  $O(h^{2k}(\mathcal{D}_n))$  under the same conditions. Second, the  $L_2$  calibration has the additional property that, even under much less restrictive conditions, it still enjoys convergence, though at a slower rate. But this slower rate is optimal under these conditions because the interpolator  $\hat{y}^p$  has the same convergence rate given by (A.5).

**5. Least  $L_2$  Distance Calibration for Expensive Code.** Now we turn to the case of expensive computer code for which  $y^s$  cannot be evaluated for infinitely many times. In this situation we need another surrogate model for  $y^s$ . Note that the input space for a computer run is  $\Omega \times \Theta \subset \mathbf{R}^{d+q}$ . Let  $\mathcal{G}$  be the set of design points for the computer experiment with its fill distance  $h(\mathcal{G})$ . Suppose  $\Theta$  is convex. Choose a positive definite function  $\Psi$  over  $(\Omega \times \Theta) \times (\Omega \times \Theta)$ . For kernel  $\Psi$  and design  $\mathcal{G}$ , let  $\hat{y}^s$  be the interpolate for  $y^s$  defined by (2.8). Then we can define the  $L_2$  calibration in a similar way:

$$(5.1) \quad \hat{\theta}_{L_2}(\mathcal{D}, \mathcal{G}) := \operatorname{argmin}_{\theta \in \Theta} \|\hat{y}^p(\cdot) - \hat{y}^s(\cdot, \theta)\|_{L_2(\Omega)}.$$

Note that the only difference from the definition in (4.9) is the replacement of  $y^s$  by its interpolate  $\hat{y}^s$  in (5.1).

We want to study the asymptotic behavior of the  $L_2$  calibration for expensive computer code. First we need to extend the definition of  $\theta^*$  in (2.3) to

$$\theta^*(\mathcal{G}) := \operatorname{argmin}_{\theta \in \Theta} \|y^p(\cdot) - \hat{y}^s(\cdot, \theta)\|_{L_2(\Omega)}.$$

Then we have

$$(5.2) \quad \|\hat{\theta}_{L_2}(\mathcal{D}, \mathcal{G}) - \theta^*\| \leq \|\hat{\theta}_{L_2}(\mathcal{D}, \mathcal{G}) - \theta^*(\mathcal{G})\| + \|\theta^*(\mathcal{G}) - \theta^*\|.$$

If we regard the interpolate  $\hat{y}^s$  as the true computer output,  $\theta^*(\mathcal{G})$  can be viewed as an “ $L_2$  projection”. Following similar steps as in the proof of Theorem 4, we can prove that

$$(5.3) \quad \|\hat{\theta}_{L_2}(\mathcal{D}_n, \mathcal{G}_n) - \theta^*(\mathcal{G}_n)\| = O(h^k(\mathcal{D}_n))$$

under the same conditions in Theorem 4. It remains to find a bound for  $\|\theta^*(\mathcal{G}) - \theta^*\|$ . The following theorem shows that its rate is slower than that in Theorem 4.

**THEOREM 5.** *Suppose  $\theta^*$  is the unique solution to (2.3);  $\Psi$  has  $2k'$  continuous derivatives with  $k' \geq 3$ ;  $\theta^*$  is an interior point of  $\Theta$ ;  $y^p \in L_2(\Omega)$ ;  $y^s \in \mathcal{N}_\Psi(\Omega \times \Theta)$  and  $\frac{\partial^2}{\partial\theta\partial\theta^T}\|\epsilon(\cdot, \theta^*)\|_{L_2(\Omega)}^2$  is invertible. Then  $\|\theta^*(\mathcal{G}_n) - \theta^*\| = O(h^{k'-1}(\mathcal{G}_n))$  as  $h(\mathcal{G}_n) \rightarrow 0$ .*

**PROOF.** As is in the proof of Theorem 4, the convergence of  $\theta^*(\mathcal{G}_n)$  to  $\theta^*$  is a direct consequence of (A.5). As  $\|y^p(\cdot) - \hat{y}_n^s(\cdot, \theta)\|_{L_2(\Omega)}^2$  is minimized at  $\theta^*(\mathcal{G}_n)$ , the Taylor expansion gives

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta} \|y^p(\cdot) - \hat{y}_n^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2 \\ &\quad + \left( \frac{\partial^2}{\partial\theta\partial\theta^T} \|y^p(\cdot) - \hat{y}_n^s(\cdot, \tilde{\theta}_n)\|_{L_2(\Omega)}^2 \right) (\theta^*(\mathcal{G}_n) - \theta^*), \end{aligned}$$

where  $\tilde{\theta}_n$  is located between  $\theta^*$  and  $\hat{\theta}_{L_2}(\mathcal{D}_n)$ . It follows from (4.13) and (A.5) in Proposition 1 with  $|\alpha| = 0, 1, 2$  that for all  $\theta \in \Theta$ ,

$$(5.4) \quad \|(y^s - \hat{y}_n^s)(\cdot, \theta)\|_{L_2(\Omega)} \leq \sqrt{\text{Vol}(\Omega)} C_\Psi h^{k'}(\mathcal{G}_n) \|y^s\|_{\mathcal{N}_\Psi(\Omega \times \Theta)},$$

$$(5.5) \quad \left\| \frac{\partial(y^s - \hat{y}_n^s)}{\partial\theta_i}(\cdot, \theta) \right\|_{L_2(\Omega)} \leq \sqrt{\text{Vol}(\Omega)} C_\Psi h^{k'-1}(\mathcal{G}_n) \|y^s\|_{\mathcal{N}_\Psi(\Omega \times \Theta)},$$

$$\left\| \frac{\partial^2(y^s - \hat{y}_n^s)}{\partial\theta_i\partial\theta_j}(\cdot, \theta) \right\|_{L_2(\Omega)} \leq \sqrt{\text{Vol}(\Omega)} C_\Psi h^{k'-2}(\mathcal{G}_n) \|y^s\|_{\mathcal{N}_\Psi(\Omega \times \Theta)},$$

for  $1 \leq i, j \leq q$ , which implies

$$\frac{\partial^2}{\partial\theta\partial\theta^T} \|y^p(\cdot) - \hat{y}_n^s(\cdot, \tilde{\theta}_n)\|_{L_2(\Omega)}^2 \rightarrow \frac{\partial^2}{\partial\theta\partial\theta^T} \|y^p(\cdot) - y^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2,$$

which is invertible by the assumption. Now as in Theorems 3 and 4, we have

$$\begin{aligned} &\frac{\partial}{\partial\theta_i} \|y^p(\cdot) - \hat{y}_n^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2 \\ &= \frac{\partial}{\partial\theta_i} \|y^p(\cdot) - \hat{y}_n^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2 - \frac{\partial}{\partial\theta_i} \|y^p(\cdot) - y^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2 \\ &= -2 \left\langle y^p(\cdot) - \hat{y}_n^s(\cdot, \theta^*), \frac{\partial\hat{y}_n^s}{\partial\theta_i}(\cdot, \theta^*) \right\rangle_{L_2(\Omega)} \\ &\quad + 2 \left\langle y^p(\cdot) - y^s(\cdot, \theta^*), \frac{\partial y^s}{\partial\theta_i}(\cdot, \theta^*) \right\rangle_{L_2(\Omega)} \end{aligned}$$

$$\begin{aligned}
(5.6) \quad &= 2 \left\langle y^p(\cdot) - y^s(\cdot, \theta^*), \frac{\partial(y^s - \hat{y}_n^s)}{\partial \theta_i}(\cdot, \theta^*) \right\rangle_{L_2(\Omega)} \\
&\quad + 2 \left\langle (\hat{y}_n^s - y^s)(\cdot, \theta^*), \frac{\partial \hat{y}_n^s}{\partial \theta_i}(\cdot, \theta^*) \right\rangle_{L_2(\Omega)}.
\end{aligned}$$

Because  $h(\mathcal{G}_n) \rightarrow 0$ , (5.4) and (5.5) implies that for sufficiently large  $n$ ,

$$(5.7) \quad \left\| \frac{\partial \hat{y}_n^s}{\partial \theta_i}(\cdot, \theta^*) \right\|_{L_2(\Omega)} \leq 2 \left\| \frac{\partial y^s}{\partial \theta_i}(\cdot, \theta^*) \right\|_{L_2(\Omega)}.$$

By applying Schwarz's inequality to (5.6) and using the bounds in (5.4), (5.5), (5.7) we obtain  $\frac{\partial}{\partial \theta_i} \|y^p(\cdot) - \hat{y}_n^s(\cdot, \theta^*)\|_{L_2(\Omega)}^2 = O(h^{k'-1}(\mathcal{G}_n))$ . This leads to the desired result.  $\square$

Theorem 5, together with (5.2)-(5.3), yields the following result on the convergence rate of the  $L_2$  calibration for expensive computer code.

**THEOREM 6.** *Under the assumptions of Theorems 4 and 5,  $\|\hat{\theta}_{L_2}(\mathcal{D}_n, \mathcal{G}_n) - \theta^*\| = O(\max(h^k(\mathcal{D}_n), h^{k'-1}(\mathcal{G}_n)))$ .*

**6. Further Discussions and Remarks.** This paper provides the first theoretical framework for studying modeling and estimation of the calibration parameters in statistical models that are motivated by and closely related to the original Kennedy-O'Hagan (2001) approach. Being the first piece of such work and because of the space limitations, it has left some issues to be further considered. The definition of consistency in this paper is based on using  $\theta^*$  in Definition 1 as the "true" calibration parameters. The same mathematical results should hold if a different positive definite metric is employed in defining  $\theta^*$ . The technical details may be more involved but the same lines of arguments can be used to obtain similar results. A bigger issue is whether  $\theta^*$  is an appropriate definition for the calibration parameters. For example, suppose we adopt  $\theta'$  in (3.7) as the "true" calibration parameters. Then the likelihood calibration  $\hat{\theta}_{Lik}$  can be declared consistent according to Theorem 1. The question is whether  $\theta'$  can be used as a legitimate definition for the calibration parameters. We would argue no because  $\theta'$  depends on the covariance function  $\Phi$ , which is part of the modeling assumption. A legitimate definition of calibration parameters should not depend on a further assumption imposed on the model. What if a different  $\Phi$  function is used in the modeling assumption? Would or should the definition of consistency be changed accordingly?

Next we turn to the discussions on the simplifications (i)-(iii) of assumptions made in Section 3.1. We have relaxed (i) in Section 5 for the  $L_2$  calibration. What about a similar extension to the likelihood calibration for the expensive code? As the situation becomes more messy (i.e., the need of estimating the function  $y^s$ ), there are reasons to believe that the procedure will remain inconsistent but the mathematical details can be more daunting. However, until further analysis is done, we are not sure if the likelihood calibration would converge to  $\theta'$ . Regarding (ii), the first part on assuming  $\rho = 1$  can be easily relaxed to an unknown  $\rho$  because  $\rho$  can be included as part of the calibration parameters in the theoretical analysis. Its second part on assuming the  $\Phi$  function is known can also be relaxed but a rigorous analysis will require more work. If we assume that the parameters for the  $\Phi$  function (such as those in (2.5)-(2.6)) lie in a compact set not involving zero, our analysis should still be applicable because the sequence of estimated values of these parameters should have a nonzero limit point. For a related discussion, see Bull (2011). Regarding assumption (iii), we can deal with the original Bayesian formulation in the KO paper by bringing back the prior information. Some heuristic calculations suggest that the prior for  $\theta$  should have no effect on the asymptotic results given in Theorem 1, provided that  $\theta'$  lies on the support of the prior distribution. Because of the space limitations, such extensions are deferred to future work.

Our work can be extended in another direction. We assume in (2.1) that the physical system is a deterministic function  $y^p(x)$  for each  $x$ . To what extent can the present work be extended to a stochastic physical system, where  $y^p(x)$  is random for some or all  $x$ ? Since this extension would require additional formulation and mathematical tools, it will be taken up in a future investigation.

## APPENDIX A: REPRODUCING KERNEL HILBERT SPACES

We give a brief summary of properties regarding reproducing kernel Hilbert space (RKHS). See Wendland (2005) for details. First, if  $\Omega$  is compact and there exists  $v \in L^2(\Omega)$ , such that

$$(A.1) \quad f(x) = \int_{\Omega} \Phi(x, t)v(t)dt,$$

then for any  $g \in \mathcal{N}_{\Phi}(\Omega)$ ,

$$(A.2) \quad \langle f, g \rangle_{\mathcal{N}_{\Phi}(\Omega)} = \int_{\Omega} v(x)g(x)dx.$$

The existence of  $v$  can be guaranteed if  $f \in \mathcal{N}_{\Phi * \Phi}(\Omega)$  (Haaland and Qian, 2011), where  $\Phi * \Phi(x) = \int_{\mathbf{R}^d} \Phi(x - t)\Phi(t)dt$  is the convolution. Further-

more, they show that in this situation there exists a continuous function  $v$  satisfying (A.1) and

$$(A.3) \quad \|v\|_{L_2(\Omega)} \leq \|f\|_{\mathcal{N}_{\Phi^* \Phi}(\Omega)}.$$

Wendland (2005) discusses the error estimates of the kernel interpolation. First, the following equality

$$(A.4) \quad \|\hat{y}\|_{\mathcal{N}_{\Phi}(\Omega)}^2 + \|y - \hat{y}\|_{\mathcal{N}_{\Phi}(\Omega)}^2 = \|y\|_{\mathcal{N}_{\Phi}(\Omega)}^2,$$

follows from the projective property of RKHS. Proposition 1 gives the error estimates for the interpolation and the native norm.

PROPOSITION 1 (Wendland, 2005, p. 181). *Suppose that  $\Phi$  has  $2k$  continuous derivatives. Then there exists a constant  $C_{\Phi}$  such that*

$$(A.5) \quad \sup_{x \in \Omega} |D^{\alpha} y(x) - D^{\alpha} \hat{y}(x)| \leq C_{\Phi} h_{X, \Omega}^{k-|\alpha|} \|y\|_{\mathcal{N}_{\Phi}(\Omega)},$$

if  $y \in \mathcal{N}_{\Phi}$  and  $|\alpha| \leq k$ , where  $\hat{y}$  is defined by (2.8);  $C_{\Phi}$  is independent of  $X$  and  $y$ ; and  $x_i$  is any component of  $x$ . Furthermore, if there exists  $v \in L_2(\Omega)$ , such that  $y(x) = \int_{\Omega} \Phi(x, t)v(t)dt$ . Then the error bounds can be improved as follows:

$$(A.6) \quad \sup_{x \in \Omega} |y(x) - \hat{y}(x)| \leq C_{\Phi} h_{X, \Omega}^{2k} \|y\|_{\mathcal{N}_{\Phi}(\Omega)},$$

$$(A.7) \quad \|y - \hat{y}\|_{\mathcal{N}_{\Phi}(\Omega)} \leq C_{\Phi} h_{X, \Omega}^k \|v\|_{L_2(\Omega)}.$$

We now turn to the extension or restriction of native spaces to another region. Assume that we have two convex regions  $\Omega_1 \subset \Omega_2 \subset \mathbf{R}^d$  and  $\Phi$  is a positive definite kernel over  $\Omega_2 \times \Omega_2$ .

PROPOSITION 2 (Wendland, 2005, p. 169). *Each function  $f \in \mathcal{N}_{\Phi}(\Omega_1)$  has a natural extension to a function  $Ef \in \mathcal{N}_{\Phi}(\Omega_2)$ . Furthermore,  $\|Ef\|_{\mathcal{N}_{\Phi}(\Omega_2)} = \|f\|_{\mathcal{N}_{\Phi}(\Omega_1)}$ .*

PROPOSITION 3 (Wendland, 2005, p. 170). *The restriction  $f|_{\Omega_1}$  of any function  $f \in \mathcal{N}_{\Phi}(\Omega_2)$  is contained in  $\mathcal{N}(\Omega_1)$  with  $\|f|_{\Omega_1}\|_{\mathcal{N}_{\Phi}(\Omega_1)} \leq \|f\|_{\mathcal{N}_{\Phi}(\Omega_2)}$ .*

Usually we assume the kernel function has the form  $\Phi(x, y) = R(x - y)$ , where  $R$  is continuous and integrable over  $\mathbf{R}^d$ . Denote the Fourier transform of  $R$  by  $\tilde{R}$ . Since  $R$  is symmetric,  $\tilde{R}$  is real and  $R$  can be recovered from  $\tilde{R}$ . Proposition 4 shows that the RKHS  $\mathcal{N}_R(\mathbf{R}^d)$  can be represented by using Fourier transforms.

PROPOSITION 4 (Wendland, 2005, p. 139). *Suppose that  $R \in C(\mathbf{R}^d) \cap L_1(\mathbf{R}^d)$  is a real-valued positive definite function. Then*

$$\mathcal{N}_R(\mathbf{R}^d) = \{f \in L_2(\mathbf{R}^d) \cap C(\mathbf{R}^d) : \tilde{f}/\sqrt{\tilde{R}} \in L_2(\mathbf{R}^d)\},$$

with the inner product given by

$$\langle f, g \rangle_{\mathcal{N}_R(\mathbf{R}^d)} = (2\pi)^{-d/2} \int_{\mathbf{R}^d} \frac{\tilde{f}(w)\overline{\tilde{g}(w)}}{\tilde{R}(w)} dw.$$

Let  $R_\theta(x) = R(\theta x)$  for  $\theta > 0$ . The next result, given by Haaland and Qian (2011), shows that under certain conditions (A.7) can be expressed in a more direct manner.

PROPOSITION 5. *Suppose  $y \in \mathcal{N}_{R_\theta * R_\theta}(\Omega)$ , and  $\hat{y}$  is the kernel interpolator given by  $R_\theta$ . If  $R$  has  $k$  continuous derivatives, then*

$$\|y - \hat{y}\|_{\mathcal{N}_{R_\theta}(\Omega)} \leq C_R \theta^{k/2} h_{X,\Omega}^{k/2} \|y\|_{\mathcal{N}_{R_\theta * R_\theta}(\Omega)},$$

where  $C_R$  is independent of  $X$ ,  $y$  and  $\theta$ .

## APPENDIX B: PROOF OF THEOREM 2

Without loss of generality, we assume  $\phi_0 = 1$ . Let  $R_\phi(x) = R(x; \phi)$  and  $Q_\phi = (\phi^d R(\cdot; \phi)) * (\phi^d R(\cdot; \phi))$ . The Fourier transform of  $Q_\phi$  is

$$(B.1) \quad \widetilde{Q}_\phi = (2\pi)^{d/2} \phi^{2d} \widetilde{R}_\phi^2.$$

We first study the relationship between  $\|\cdot\|_{\mathcal{N}_{Q_1}(\Omega)}$  and  $\|\cdot\|_{\mathcal{N}_{Q_\phi}(\Omega)}$ , for  $\phi > 1$ . For any  $f \in \mathcal{N}_{Q_1}(\Omega)$ , by Proposition 2, there exists an extension  $Ef \in \mathcal{N}_{Q_1}(\mathbf{R}^d)$ , such that  $\|f\|_{\mathcal{N}_{Q_1}(\Omega)} = \|Ef\|_{\mathcal{N}_{Q_1}(\mathbf{R}^d)}$ . Then we have

$$\begin{aligned} \|f\|_{\mathcal{N}_{Q_\phi}(\Omega)}^2 &\leq \|Ef\|_{\mathcal{N}_{Q_\phi}(\mathbf{R}^d)}^2 \\ &= (2\pi)^{-d/2} \int_{\mathbf{R}^d} \frac{|\widetilde{Ef}(w)|^2}{\widetilde{Q}_\phi(w)} dw \\ &= (2\pi)^{-d/2} \int_{\mathbf{R}^d} \frac{|\widetilde{Ef}(w)|^2}{(2\pi)^{d/2} \phi^{2d} \widetilde{R}_\phi^2(w)} dw \\ &= (2\pi)^{-d/2} \int_{\mathbf{R}^d} \frac{|\widetilde{Ef}(w)|^2}{(2\pi)^{d/2} \widetilde{R}^2(w/\phi)} dw \end{aligned}$$

$$\begin{aligned}
&\leq (2\pi)^{-d/2} \sup_{w \neq 0} \{\tilde{R}(\phi w)/\tilde{R}(w)\}^2 \int_{\mathbf{R}^d} \frac{|\widetilde{E}f(w)|^2}{(2\pi)^{d/2} \tilde{R}^2(w)} dw \\
&= (2\pi)^{-d/2} \sup_{w \neq 0} \{\tilde{R}(\phi w)/\tilde{R}(w)\}^2 \int_{\mathbf{R}^d} \frac{|\widetilde{E}f(w)|^2}{\widetilde{Q}_1(w)} dw \\
&= \sup_{w \neq 0} \{\tilde{R}(\phi w)/\tilde{R}(w)\}^2 \|Ef\|_{\mathcal{N}_{Q_1}(\mathbf{R}^d)}^2 \\
\text{(B.2)} \quad &\leq \sup_{w \neq 0, \alpha \leq 1} \{\tilde{R}(\alpha w)/\tilde{R}(w)\}^2 \|f\|_{\mathcal{N}_{Q_1}(\Omega)}^2,
\end{aligned}$$

where the first inequality follows from Proposition 3; the first equality follows from Proposition 4; the second equality follows from (B.1); the third equality follows from the fact that  $\tilde{R}_\phi(w) = \phi^{-d} \tilde{R}(w/\phi)$ ; the second inequality follows from factoring out  $\{\tilde{R}(\phi w)/\tilde{R}(w)\}^2$ ; the fourth equality follows from (B.1); the fifth equality follows from Proposition 4; and the last inequality follows from condition A3 and Proposition 2.

Let  $C_0 = \int_{\mathbf{R}^d} R(x; 1) dx$ . Define integral operator  $\kappa_\phi : L_2(\Omega) \mapsto L_2(\Omega)$  by

$$\kappa_\phi(f)(x) = C_0^{-1} \phi^d \int_{\Omega} R(x-y; \phi) f(y) dy, \quad x \in \Omega,$$

for any  $\phi > 0$ . Obviously  $\kappa_\phi$  is a self-adjoint operator, i.e.,  $\langle f, \kappa_\phi(g) \rangle = \langle \kappa_\phi(f), g \rangle$  for any  $f, g \in L_2(\Omega)$ . For any  $x \in \Omega$ , let  $x - \Omega = \{x - x_0 : x_0 \in \Omega\}$  and  $\phi(x - \Omega) = \{\phi(x - x_0) : x_0 \in \Omega\}$ . First we show that for any interior point  $x$  in  $\Omega$ ,

$$\begin{aligned}
&\lim_{\phi \rightarrow +\infty} C_0^{-1} \phi^d \int_{\Omega} R(x-y; \phi) dy \\
&= \lim_{\phi \rightarrow +\infty} C_0^{-1} \phi^d \int_{\mathbf{R}^d} I(y \in x - \Omega) R(y; \phi) dy \\
&= \lim_{\phi \rightarrow +\infty} C_0^{-1} \phi^d \int_{\mathbf{R}^d} I(y \in x - \Omega) R(\phi y; 1) dy \\
&= \lim_{\phi \rightarrow +\infty} C_0^{-1} \int_{\mathbf{R}^d} I(y \in \phi(x - \Omega)) R(y; 1) dy \\
\text{(B.3)} \quad &= C_0^{-1} \int_{\mathbf{R}^d} R(y; 1) dy = 1,
\end{aligned}$$

where the fourth equality follows from the dominated convergence theorem and the fact that  $x - \Omega$  contains a neighborhood of 0.

Let  $r_n = C_0^{-1} \phi^d \int_{\Omega} R(x-y; \phi) dy - 1$ . Then for any interior point  $x$  in  $\Omega$ ,

$$\sup_{\theta} |\kappa_\phi(\epsilon(\cdot, \theta))(x) - \epsilon(x, \theta)|$$

$$\begin{aligned}
&= \sup_{\theta} \left| C_0^{-1} \phi^d \int_{\Omega} R(x-y; \phi) \epsilon(y, \theta) dy \right. \\
&\quad \left. - C_0^{-1} \phi^d \int_{\Omega} R(x-y; \phi) \epsilon(x, \theta) dy + r_n \epsilon(x; \theta) \right| \\
&\leq \sup_{\theta} \left| C_0^{-1} \int_{\mathbf{R}^d} I(y \in \phi(x-\Omega)) R(y; 1) \right. \\
&\quad \left. \{ \epsilon(x-y/\phi, \theta) - \epsilon(x, \theta) \} dy \right| + \sup_{\theta} |r_n \epsilon(x, \theta)| \\
\text{(B.4)} \quad &\leq \phi^{-1} \sup_{\theta} \|\nabla_x \epsilon(x, \theta)\| \left| \int_{\mathbf{R}^d} I(y \in \phi(x-\Omega)) R(y; 1) \|y\| dy \right| \\
&\quad + |r_n| \sup_{\theta} |\epsilon(x, \theta)|,
\end{aligned}$$

where the last inequality follows from the mean value theorem and condition A1. Using the dominated convergence theorem, we have  $\int_{\mathbf{R}^d} I(y \in \phi(x-\Omega)) R(y; 1) (\|y\|/\phi) dy \rightarrow 0$ , since  $y/\phi$  lies in the bounded set  $x-\Omega$ . This shows that the first term in (B.4) tends to 0. The second term in (B.4) also tends to 0 because of (B.3). Thus (B.4) shows that  $\kappa_{\phi}(\epsilon(\cdot, \theta))(x)$  converges to  $\epsilon(x, \theta)$  uniformly with respect to  $\theta$ .

From the definition of native spaces, it is easily seen that

$$\text{(B.5)} \quad \|f\|_{\mathcal{N}_K(\Omega)}^2 = c \|f\|_{\mathcal{N}_{cK}(\Omega)}^2$$

for any  $f$ ,  $K$  and  $c > 0$ . Since condition A2 holds, by applying (B.2) to  $\epsilon(\cdot, \theta)$ , we see that for any  $\phi > 1$  and  $\theta \in \Theta$ ,  $\epsilon(\cdot, \theta) \in \mathcal{N}_{Q_{\phi}}(\Omega)$ . Consequently, from (A.1)-(A.3), for any  $\theta \in \Theta$  and  $\phi > 1$ , there exists  $v_{\phi, \theta} \in L_2(\Omega)$ , such that  $\epsilon(x, \theta) = \int_{\Omega} R(x-t; \phi) v_{\phi, \theta}(t) dt$ . Thus

$$\begin{aligned}
\epsilon(x, \theta) &= \int_{\Omega} (C_0^{-1} \phi^d) R(x-t; 1) (C_0 \phi^{-d}) v_{\phi, \theta}(t) dt \\
\text{(B.6)} \quad &= \kappa((C_0 \phi^{-d}) v_{\phi, \theta})(x).
\end{aligned}$$

Applying (A.3) and (B.2), we have

$$\begin{aligned}
&\|C_0 \phi^{-d} v_{\phi, \theta}\|_{L_2(\Omega)} \leq \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{Q_{\phi}}(\Omega)} \\
\text{(B.7)} \quad &\leq \sup_{w \neq 0, \alpha \leq 1} \{ \tilde{R}(\alpha w) / \tilde{R}(w) \}^2 \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{Q_1}(\Omega)}^2 < +\infty.
\end{aligned}$$

Then

$$\begin{aligned}
&C_0 \phi^{-d} \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{R_{\phi}}(\Omega)}^2 - \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2 \\
&= \left\langle \epsilon(\cdot, \theta), C_0 \phi^{-d} v_{\phi, \theta} \right\rangle_{L_2(\Omega)} - \left\langle \epsilon(\cdot, \theta), \epsilon(\cdot, \theta) \right\rangle_{L_2(\Omega)}
\end{aligned}$$

$$\begin{aligned}
&= \left\langle \epsilon(\cdot, \theta), C_0 \phi^{-d} v_{\phi, \theta} - \epsilon(\cdot, \theta) \right\rangle_{L_2(\Omega)} \\
&= \left\langle \kappa_\phi(C_0 \phi^{-d} v_{\phi, \theta}), C_0 \phi^{-d} v_{\phi, \theta} - \epsilon(\cdot, \theta) \right\rangle_{L_2(\Omega)} \\
&= \left\langle C_0 \phi^{-d} v_{\phi, \theta}, \kappa_\phi(C_0 \phi^{-d} v_{\phi, \theta} - \epsilon(\cdot, \theta)) \right\rangle_{L_2(\Omega)} \\
&= \left\langle C_0 \phi^{-d} v_{\phi, \theta}, \epsilon(\cdot, \theta) - \kappa_\phi(\epsilon(\cdot, \theta)) \right\rangle_{L_2(\Omega)} \\
&\leq \|C_0 \phi^{-d} v_{\phi, \theta}\|_{L_2(\Omega)} \|\epsilon(\cdot, \theta) - \kappa_\phi(\epsilon(\cdot, \theta))\|_{L_2(\Omega)} \\
\text{(B.8)} \quad &\leq \|C_0 \phi^{-d} v_{\phi, \theta}\|_{L_2(\Omega)} \left\| \sup_{\theta \in \Theta} |\epsilon(x, \theta) - \kappa_\phi(\epsilon(\cdot, \theta))(x)| \right\|_{L_2(\Omega)},
\end{aligned}$$

where the first equality follows from (A.2); the third equality follows from the definition of  $\kappa$ ; the fourth equality follows from the self-adjoint property of  $\kappa$ ; the fifth equality follows from (B.6); the first inequality follows from Schwarz's inequality. Note that (B.7) gives the uniform upper bound of  $\|C_0 \phi^{-d} v_{\phi, \theta}\|_{L_2}$  with respect to  $\theta$ . Using the dominated convergence theorem and (B.4),  $\|\epsilon(x, \theta) - \kappa_\phi(\epsilon(\cdot, \theta))(x)\|_{L_2} \rightarrow 0$ , as  $\phi \rightarrow \infty$ .

Now return to the settings of Theorem 2. Since  $\phi_n \rightarrow +\infty$  as  $n \rightarrow \infty$ , applying (B.5), we have

$$\text{(B.9)} \quad \sup_{\theta \in \Theta} \left| \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{C_0^{-1} \phi_n^d R_{\phi_n}}(\Omega)}^2 - \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2 \right| \rightarrow 0,$$

as  $n \rightarrow \infty$ . On the other hand, we have

$$\begin{aligned}
&\sup_{\theta \in \Theta} \left| \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{C_0^{-1} \phi_n^d R_{\phi_n}}(\Omega)}^2 - \|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_{C_0^{-1} \phi_n^d R_{\phi_n}}(\Omega)}^2 \right| \\
&= \sup_{\theta \in \Theta} \left| \|\epsilon(\cdot, \theta) - \hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_{C_0^{-1} \phi_n^d R_{\phi_n}}(\Omega)}^2 \right| \\
\text{(B.10)} \quad &\leq C_R^2 \sqrt{\phi_n h(\mathcal{D}_n)} \sup_{\theta \in \Theta} \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{C_0^{-2} Q_{\phi_n}}(\Omega)}^2 \\
&\leq \left( C_R^2 C_0^2 \sup_{w \neq 0, \alpha \leq 1} \{\tilde{R}(\alpha w) / \tilde{R}(w)\}^2 \sup_{\theta \in \Theta} \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{Q_1}(\Omega)}^2 \right) \sqrt{\phi_n h(\mathcal{D}_n)}
\end{aligned}$$

$$\text{(B.11)} \rightarrow 0,$$

where the equality follows from (A.4); the first inequality follows from Proposition 5; the second inequality follows from (B.2); and the limiting relationship follows from conditions A2, A3 and the fact that  $\phi_n h(\mathcal{D}_n) \rightarrow 0$ . One may notice that (B.10) does not follow immediately from Proposition 5 because in Proposition 5  $\hat{y}$  is built using the kernel  $R_\theta$  but here  $\hat{\epsilon}_n(\cdot, \theta)$  is built using  $R_{\phi_n}$  instead of  $C_0^{-1} \phi_n^d R_{\phi_n}$ . However, this is not a serious problem because using (2.7) and (2.8), and after some simple calculations, we

can verify that the two interpolators given by  $R_{\phi_n}$  and  $C_0^{-1}\phi_n^d R_{\phi_n}$  respectively are equal to each other. Based on this equivalence, Proposition 5 is still applicable.

$$\begin{aligned} & \text{Now we can bound } \left| C_0\phi_n^{-d} \|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_{R_{\phi_n}}(\Omega)}^2 - \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2 \right| \text{ with} \\ & \sup_{\theta \in \Theta} \left| C_0\phi_n^{-d} \|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_{R_{\phi_n}}(\Omega)}^2 - \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2 \right| \\ &= \sup_{\theta \in \Theta} \left| \|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_{C_0\phi_n^{-d}R_{\phi_n}}(\Omega)}^2 - \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2 \right| \\ &\leq \sup_{\theta \in \Theta} \left| \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{C_0^{-1}\phi_n^d R_{\phi_n}}(\Omega)}^2 - \|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_{C_0^{-1}\phi_n^d R_{\phi_n}}(\Omega)}^2 \right| \\ & \quad + \sup_{\theta \in \Theta} \left| \|\epsilon(\cdot, \theta)\|_{\mathcal{N}_{C_0^{-1}\phi_n^d R_{\phi_n}}(\Omega)}^2 - \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2 \right| \rightarrow 0, \end{aligned}$$

where the first equality follows from (B.5); the inequality follows from the triangle inequality; and the limiting relationship follows from (B.9) and (B.11). Therefore, we have established the following result

$$\hat{\theta}(R_n, \mathcal{D}_n) = \underset{\theta \in \Theta}{\operatorname{argmin}} \|\hat{\epsilon}_n(\cdot, \theta)\|_{\mathcal{N}_{R_{\phi_n}}(\Omega)}^2 \rightarrow \underset{\theta \in \Theta}{\operatorname{argmin}} \|\epsilon(\cdot, \theta)\|_{L_2(\Omega)}^2 = \theta^*,$$

as  $n \rightarrow \infty$ . This completes the proof.  $\square$

## REFERENCES

- ANJYO, K. and LEWIS, J. (2011). RBF interpolation and Gaussian process regression through an RKHS formulation. *Journal of Mathematics for Industry* **3** 63–71.
- BAYARRI, M. J., BERGER, J. O., CAFEO, J., GARCIA-DONATO, G., LIU, F., PALOMO, J., PARTHASARATHY, R. J., PAULO, R., SACKS, J. and WALSH, D. (2007a). Computer model validation with functional output. *The Annals of Statistics* **35** 1874–1906.
- BAYARRI, M. J., BERGER, J. O., PAULO, R., SACKS, J., CAFEO, J. A., CAVENDISH, J., LIN, C. H. and TU, J. (2007b). A framework for validation of computer models. *Technometrics* **49** 138–154.
- BUHMANN, M. D. (2003). *Radial Basis Functions: Theory and Implementations* **12**. Cambridge University Press.
- BULL, A. D. (2011). Convergence rates of efficient global optimization algorithms. *The Journal of Machine Learning Research* **12** 2879–2904.
- CONWAY, J. B. (1990). *A Course in Functional Analysis* **96**. Springer.
- FANG, K. T., LI, R. and SUDJANTO, A. (2005). *Design and Modeling for Computer Experiments* **6**. Chapman & Hall/CRC.
- FASSHAUER, G. E. (2011). Positive definite kernels: Past, present and future. *Dolomite Research Notes on Approximation* **4** 21–63.
- GOLDSTEIN, M. and ROUGIER, J. (2004). Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM Journal on Scientific Computing* **26** 467–487.
- HAALAND, B. and QIAN, P. Z. G. (2011). Accurate emulators for large-scale computer experiments. *The Annals of Statistics* **39** 2974–3002.

- HAN, G., SANTNER, T. J. and RAWLINSON, J. J. (2009). Simultaneous determination of tuning and calibration parameters for computer experiments. *Technometrics* **51** 464–474.
- 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 of Scientific Computing* **26** 448–466.
- HIGDON, D., GATTIKER, J., WILLIAMS, B. and RIGHTLEY, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association* **103** 570–583.
- JOSEPH, V. R. and MELKOTE, S. N. (2009). Statistical adjustments to engineering models. *Journal of Quality Technology* **41** 362–375.
- KENNEDY, M. C. and O’HAGAN, A. (2001). Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society: Series B* **63** 425–464.
- MURPHY, J. M., BOOTH, B. B. B., COLLINS, M., HARRIS, G. R., SEXTON, D. M. H. and WEBB, M. J. (2007). A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **365** 1993–2028.
- SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments. *Statistical Science* **4** 409–423.
- SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. (2003). *The Design and Analysis of Computer Experiments*. Springer Verlag.
- STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Verlag.
- STEIN, E. M. and WEISS, G. L. (1971). *Introduction to Fourier Analysis on Euclidean Spaces* **32**. Princeton University Press.
- WANG, S., CHEN, W. and TSUI, K. L. (2009). Bayesian validation of computer models. *Technometrics* **51** 439–451.
- WENDLAND, H. (2005). *Scattered Data Approximation*. Cambridge University Press.

ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE	SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING
CHINESE ACADEMY OF SCIENCES	GEORGIA INSTITUTE OF TECHNOLOGY
BEIJING, CHINA 100190	ATLANTA, GEORGIA 30332-0205
E-MAIL: <a href="mailto:tuorui@amss.ac.cn">tuorui@amss.ac.cn</a>	E-MAIL: <a href="mailto:jeffwu@isye.gatech.edu">jeffwu@isye.gatech.edu</a>