Rational Kriging

V. Roshan Joseph

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology, Atlanta, GA 30332, USA
roshan@gatech.edu

Abstract

This article proposes a new kriging that has a rational form. It is shown that the generalized least squares estimator of the mean from rational kriging is much more well behaved than that of ordinary kriging. Parameter estimation and uncertainty quantification for rational kriging are proposed using a Gaussian process framework. A generalized version of rational kriging is also proposed, which includes ordinary and rational kriging as special cases. Extensive simulations carried out over a wide class of functions show that the generalized rational kriging performs on par or better than both ordinary and rational kriging in terms of prediction and uncertainty quantification. The only extra step needed for generalized rational kriging over ordinary kriging is the computation of Perron eigenvector of an augmented correlation matrix which can be computed in near linear time and therefore, its overall computational complexity is no more than that of ordinary kriging. The potential applications of the new kriging methods in the emulation of computationally expensive models and model calibration problems are illustrated with real and simulated examples.

Keywords: Calibration; Computer experiments; Gaussian process; Rational radial basis functions; Surrogate model; Uncertainty quantification.

1 Introduction

Kriging is a technique for multivariate interpolation of arbitrarily scattered data. It is originated from some mining-related applications, which is developed into the field of geostatistics by the pioneering work of Matheron (1963). Kriging/Gaussian process has now become a prominent technique for computer experiments (Santner et al., 2003). Examples include calibration of ion channel models in cardiac cells (Plumlee et al., 2016), emulation of large eddy simulations in a rocket engine design (Mak et al., 2018), and optimization of 3D

printed aortic valves using finite element models (Chen et al., 2021), to name a few. Kriging is also widely used in spatial statistics (Cressie, 2015) and machine learning (Rasmussen and Williams, 2006), but computer experiments will be the main focus of this article.

Kriging can be briefly explained as follows. Suppose we have observed the data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x} \in \mathbb{R}^p$ is the p-dimensional input and y the output. The aim is to predict y for a future $\mathbf{x}$. To do this, construct a linear predictor $\hat{y}(\mathbf{x}) = \mathbf{a}(\mathbf{x})'\mathbf{y} = \sum_{i=1}^n a_i(\mathbf{x}) y_i$. Kriging gives the best linear unbiased predictor (BLUP) under some assumptions of the data generating process. Specifically, if the data are generated from a second-order stationary stochastic process with mean μ, variance τ^2, and correlation function $cor\{Y(\mathbf{u}), Y(\mathbf{v})\} = R(\mathbf{u} - \mathbf{v})$, then the kriging predictor can be obtained by minimizing the mean squared prediction error (Santner et al., 2003)

$$E\left\{Y(\mathbf{x}) - \hat{y}(\mathbf{x})\right\}^2$$

with respect to $\mathbf{a}(\mathbf{x})$ subject to the condition that $E\{\hat{y}(\mathbf{x})\} = \mu$ for all $\mathbf{x}$. The optimal solution is given by

$$\mathbf{a}(\mathbf{x})' = \{1 - \mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{1}\}\frac{\mathbf{1}'\mathbf{R}^{-1}}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}} + \mathbf{r}(\mathbf{x})'\mathbf{R}^{-1},$$

where $\mathbf{R} = \{R(\mathbf{x}_i - \mathbf{x}_j)\}_{n\times n}$ is the correlation matrix, $\mathbf{r}(\mathbf{x}) = (R(\mathbf{x} - \mathbf{x}_1), \ldots, R(\mathbf{x} - \mathbf{x}_n))'$, and $\mathbf{1}$ is a vector of n 1's. Substituting the solution in the linear predictor and simplifying, we obtain the (ordinary) kriging predictor as

$$\hat{y}_{OK}(\mathbf{x}) = \hat{\mu}_{OK} + \mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}(\mathbf{y} - \hat{\mu}_{OK}\mathbf{1}), \tag{1}$$

where

$$\hat{\mu}_{OK} = \frac{\mathbf{1}'\mathbf{R}^{-1}\mathbf{y}}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}}. \tag{2}$$

The expression in (2) shows that $\hat{\mu}_{OK}$ is the well-known Generalized Least Squares (GLS) estimate of μ.

Joseph (2006) noticed that the ordinary kriging predictor sometimes has a "mean

reversion" issue (Zhang and Apley, 2014) and proposed a modified predictor

$$\hat{y}(\mathbf{x}) = \frac{\mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{y}}{\mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{1}}, \tag{3}$$

whose predictions tend towards the nearest neighbor value when the correlations go to zero and thus, avoids the mean reversion issue. This predictor can be obtained as the limiting case of a simple kriging predictor with a recursive estimation of μ and hence it is called limit kriging. It is a linear predictor in $\mathbf{y}$, but it can also be viewed as a rational predictor in $\mathbf{x}$ because (3) can be written as

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i=1}^{n} b_i R(\mathbf{x} - \mathbf{x}_i)}{\sum_{i=1}^{n} c_i R(\mathbf{x} - \mathbf{x}_i)},$$

where $\mathbf{b} = \mathbf{R}^{-1}\mathbf{y}$ and $\mathbf{c} = \mathbf{R}^{-1}\mathbf{1}$. The purpose of this article is to examine optimal rational predictors in $\mathbf{x}$.

Although rational polynomials have a long history in interpolation, its extension to radial basis functions (RBFs) is very recent. Jakobsson et al. (2009) proposed to use rational RBFs for modeling resonance phenomena. Sarra and Bai (2018) also found that rational RBFs perform exceptionally well for modeling functions with discontinuities and steep gradients. In a more recent work, Buhmann et al. (2020) showed that rational RBFs have comparable approximation accuracy to the classical RBFs, but have more robust prediction performance. However, RBFs cannot provide any uncertainty quantification. In contrast, owing to its probabilistic formulation, kriging can automatically provide prediction intervals and can easily be integrated into Bayesian methods and non-normal data settings.

Different from the RBF literature, we will motivate the benefit of rational predictors using parameter estimation accuracy. As an example, consider the deflection of a simply supported beam with uniform load shown as an inset in the left panel of Figure 1. The

deflection at a distance x from the left end of the beam is given by

$$y = -\frac{P}{24EI}x(x^3 - 2Lx^2 + L^3),$$

where P is the uniform load density, E is the elastic modulus, I is the area moment of inertia, L is the length of the beam, and $x \in [0, L]$. Let $P/(24EI) = 1$ and $L = 1$. The function is plotted in the left panel of Figure 1 along with 11 equi-spaced x_i's from $[0, 1]$. An ordinary kriging was fitted to this data with a Gaussian correlation function $R(h) = e^{-(h/\theta)^2}$, where the unknown correlation parameter θ is estimated from the data using maximum likelihood. We used the R package `DiceKriging` (Roustant et al., 2012) for estimation, where a small nugget of 10^{-6} is applied for numerical stability. The predictions in $[0, 1]$ are plotted in the left panel of Figure 1, which are almost indistinguishable with the true function values showing excellent prediction performance. The GLS estimate of μ from (2) is obtained as $\hat{\mu}_{OK} = 0.224$. Interestingly, this value is outside the range of the observed function values, which are from $[-0.3125, 0]$.

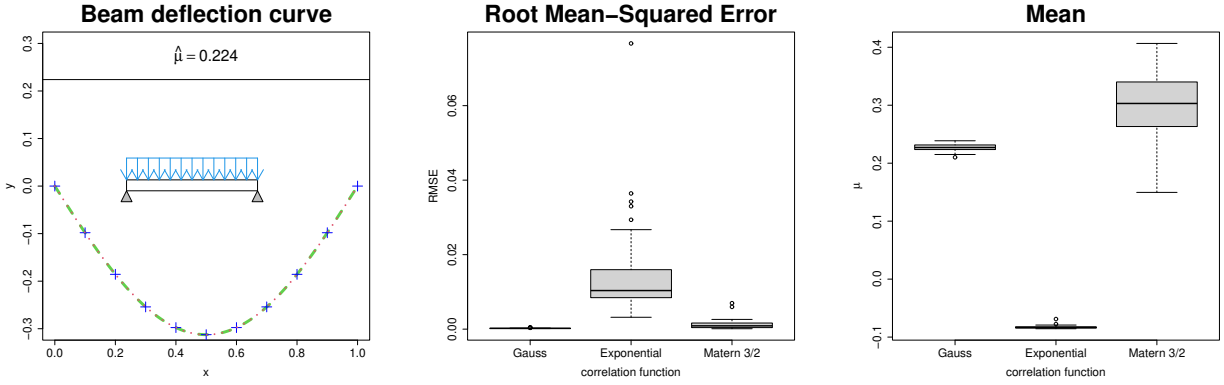

Figure 1: *(left) Plot of the beam deflection curve (dashed-green), data (blue-pluses), and ordinary kriging predictor (dotted-red). A simply supported beam with uniform load is shown as an inset of this plot. (middle) Boxplot of $\hat{\mu}$'s from 50 simulations using three correlation functions where the x_i's are randomly sampled in $[0, 1]$. (right) Root mean squared errors from the 50 simulations.*

We repeated this exercise 50 times by uniformly sampling x_i's from $[0, 1]$ and using

two more correlation functions: exponential and Matérn 3/2 (Rasmussen and Williams, 2006, p.84). We can see from the middle panel of Figure 1 that the estimates of μ using exponential correlation function are around -0.1, which are in the range of the observed values, but the estimates of μ from the Matérn 3/2 correlation function are generally higher than those obtained from the Gaussian correlation function. For each simulation, the root mean squared error (RMSE) is calculated over a grid of 1,001 values and is shown in the right panel of Figure 1. We can see that the Gaussian correlation function gives the best prediction in this example. The prediction from the exponential correlation function is the worst in spite of having the mean in the observed range of y_i's.

Although better prediction is obtained when $\hat{\mu}$ is outside the observed range of function values, the interpretation of those estimates becomes questionable. One could argue that μ is the mean of a stochastic process in which the beam deflection curve is just a realization and thus a value around 0.2 is an admissible estimate. However, if μ has a physical interpretation, then this estimate is meaningless. For example, a positive value of mean would imply that the beam will deflect in the opposite direction of the force, which is against the law of nature! This is a common dilemma in model calibration problems (Kennedy and O'Hagan, 2001). We will show that the use of rational kriging can surprisingly avoid this issue without sacrificing the prediction performance.

A quick fix to the foregoing issue is to estimate the mean using ordinary least squares (Pronzato and Zhigljavsky, 2023). However, it leads to inconsistencies in the modeling framework– an uncorrelated process for estimation and a correlated process for prediction. Plumlee and Joseph (2018) argued that the estimation problems are caused by identifiability issues between the stochastic process and the mean function (a constant function in the case of ordinary kriging). They proposed to overcome the identifiability issue by making the stochastic process orthogonal to the mean function. Although their approach is very general, it leads to a nonstationary correlation function that involves high-dimensional integrals making the estimation computationally challenging and numerically unstable. In contrast, rational kriging requires only a rescaling of the original predictor, which is very

easy to implement in practice.

The article is organized as follows. Section 2 develops the rational kriging and a Gaussian process framework for parameter estimation and uncertainty quantification. A generalization of the rational kriging is also proposed in which ordinary and rational kriging are special cases. Simulations with several test functions are provided in Section 3. Potential applications of rational and generalized rational kriging in emulation and calibration of computer models are illustrated with some examples in Section 4. Some concluding remarks are given in Section 5.

2 Methodology

We will first derive the optimal rational predictor by minimizing the mean squared prediction error and then investigate its estimation properties by assuming Gaussianity for the stochastic process. A generalization of the rational predictor will also be discussed.

2.1 Rational Kriging

Let $\mathcal{X}$ be the input region for data collection. Most of the time, it can be scaled in $[0,1]^p$. Notice that this region does not come into the formulation or derivation of the ordinary kriging predictor because we assume the stationary stochastic process has a constant mean μ and variance τ^2 for all $\mathbf{x} \in \mathbb{R}^p$. This could be the reason why the estimate of μ went outside the observed range of y values in the example that we saw earlier. So, we can possibly overcome the issue by assuming a nonstationary variance $\tau^2(\mathbf{x})$, where it should increase as $\mathbf{x}$ goes outside of $\mathcal{X}$.

Now consider a predictor of the form

$$\hat{y}(\mathbf{x}) = \frac{\mathbf{a}(\mathbf{x})'\mathbf{y}}{b(\mathbf{x})}, \tag{4}$$

where $\mathbf{a}(\mathbf{x}) = (a_1(\mathbf{x}), \ldots, a_n(\mathbf{x}))'$ and $b(\mathbf{x})$ are functions of the input variables $\mathbf{x} =$

$(x_1, \ldots, x_p)'$. Assume that the data $\mathbf{y}$ is a realization from a second-order stationary stochastic process with mean μ, variance $\tau^2(\mathbf{x})$, and correlation function $R(\cdot)$. Then, the predictor in (4) will be unbiased if

$$E\{\hat{y}(\mathbf{x})\} = \frac{\mathbf{a}(\mathbf{x})' E\{\mathbf{y}\}}{b(\mathbf{x})} = \mu \frac{\mathbf{a}(\mathbf{x})'\mathbf{1}}{b(\mathbf{x})} = \mu$$

for all $\mathbf{x}$, which implies $b(\mathbf{x}) = \mathbf{a}(\mathbf{x})'\mathbf{1}$. Now we can find the best unbiased predictor by minimizing

$$MSPE = E\left\{ Y(\mathbf{x}) - \frac{\mathbf{a}(\mathbf{x})'\mathbf{y}}{\mathbf{a}(\mathbf{x})'\mathbf{1}} \right\}^2$$

with respect to $\mathbf{a}(\mathbf{x})$. Note that unlike ordinary kriging, we do not need to impose any constraints on $\mathbf{a}(\mathbf{x})$. It is easy to show that $cov(Y(\mathbf{x}), \mathbf{y}) = \tau(\mathbf{x}) diag(\boldsymbol{\tau}) \mathbf{r}(\mathbf{x})$ and $var\{\mathbf{y}\} = diag(\boldsymbol{\tau}) \mathbf{R} diag(\boldsymbol{\tau})$, where $\boldsymbol{\tau} = (\tau(\mathbf{x}_1), \ldots, \tau(\mathbf{x}_n))'$ and $diag(\boldsymbol{\tau})$ is a diagonal matrix with diagonal elements $\boldsymbol{\tau}$. Thus,

$$MSPE = \tau^2(\mathbf{x}) - 2\frac{\mathbf{a}(\mathbf{x})'}{\mathbf{a}(\mathbf{x})'\mathbf{1}} \tau(\mathbf{x}) diag(\boldsymbol{\tau}) \mathbf{r}(\mathbf{x}) + \frac{\mathbf{a}(\mathbf{x})'}{\mathbf{a}(\mathbf{x})'\mathbf{1}} diag(\boldsymbol{\tau}) \mathbf{R} diag(\boldsymbol{\tau}) \frac{\mathbf{a}(\mathbf{x})}{\mathbf{a}(\mathbf{x})'\mathbf{1}}.$$

Differentiating with respect to $\mathbf{a}(\mathbf{x})$ and equating to zero, we obtain

$$diag(\boldsymbol{\tau}) \left\{ -2\tau(\mathbf{x})\mathbf{r}(\mathbf{x}) + 2\mathbf{R} diag(\boldsymbol{\tau}) \frac{\mathbf{a}(\mathbf{x})}{\mathbf{a}(\mathbf{x})'\mathbf{1}} \right\} \frac{\partial}{\partial \mathbf{a}(\mathbf{x})} \left(\frac{\mathbf{a}(\mathbf{x})}{\mathbf{a}(\mathbf{x})'\mathbf{1}} \right) = 0.$$

Thus,

$$\frac{\mathbf{a}(\mathbf{x})}{\mathbf{a}(\mathbf{x})'\mathbf{1}} = \tau(\mathbf{x}) diag(\boldsymbol{\tau}^{-1}) \mathbf{R}^{-1} \mathbf{r}(\mathbf{x})$$

is a solution, provided $\tau(\mathbf{x})\mathbf{r}(\mathbf{x})'\mathbf{R}^{-1} diag(\boldsymbol{\tau}^{-1})\mathbf{1} = 1$. Therefore, we can let

$$\tau(\mathbf{x}) = \frac{1}{\mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}(\mathbf{1}/\boldsymbol{\tau})}, \tag{5}$$

where $\mathbf{1}/\boldsymbol{\tau} = (1/\tau_1, \ldots, 1/\tau_n)'$ and $\tau_i = \tau(\mathbf{x}_i)$ for $i = 1, \ldots, n$. Since $\mathbf{r}(\mathbf{x}_i)'\mathbf{R}^{-1} = (0, \ldots, 1, \ldots, 0)'$ with 1 at the ith position, (5) holds for $i = 1, \ldots, n$. However, (5) is

meaningful only if $\mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}(\mathbf{1}/\boldsymbol{\tau}) > 0$ for all $\mathbf{x}$. This can be ensured by putting a constraint on $\boldsymbol{\tau} \geq 0$ such that $\mathbf{R}^{-1}(\mathbf{1}/\boldsymbol{\tau}) \geq 0$ and choosing a correlation function that does not vanish. Interestingly, $\tau(\mathbf{x}) \to \infty$ as $||\mathbf{x} - \mathbf{x}_i|| \to \infty$ for all i, which agrees with our intuition.

Thus, we obtain the optimal rational kriging predictor as

$$\hat{y}(\mathbf{x}) = \frac{\mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}(\mathbf{y}/\boldsymbol{\tau})}{\mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}(\mathbf{1}/\boldsymbol{\tau})}. \tag{6}$$

We can see that the limit kriging predictor in (3) is a special case of this predictor with $\boldsymbol{\tau} = \mathbf{1}$. However, limit kriging is not an admissible predictor in the new formulation because $\mathbf{R}^{-1}\mathbf{1}$ is not guaranteed to be nonnegative. The new predictor has n additional unknown parameters $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)'$, which can be chosen to ensure that $\mathbf{R}^{-1}(\mathbf{1}/\boldsymbol{\tau}) \geq 0$ and $\boldsymbol{\tau} \geq 0$.

Let $\mathbf{R}^{-1}(\mathbf{1}/\boldsymbol{\tau}) = \mathbf{c}/\nu$, where $\mathbf{c} \geq 0$, $\|\mathbf{c}\|_2 = 1$, and ν is a positive constant. Since $\mathbf{R}$ is a positive matrix, $\mathbf{c} \geq 0$ implies $\mathbf{1}/\boldsymbol{\tau} = \mathbf{R}\mathbf{c}/\nu \geq 0$. Thus, the rational kriging predictor can be written as

$$\hat{y}(\mathbf{x}) = \frac{\mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}diag(\mathbf{R}\mathbf{c})\mathbf{y}}{\mathbf{r}(\mathbf{x})'\mathbf{c}}, \tag{7}$$

where $\mathbf{c} \geq 0$. This is the same predictor obtained by Kang and Joseph (2016) as the limiting case of an iterated kernel regression. The choice of $\mathbf{c}$ will be discussed in the next section.

The derivation of rational kriging predictor does not give any estimate of μ. However, since $\mathbf{y}$ is a random vector with mean $\mu\mathbf{1}$ and variance $diag(\boldsymbol{\tau})\mathbf{R}diag(\boldsymbol{\tau})$, we can use the GLS estimate for μ:

$$\begin{aligned} \hat{\mu} &= \frac{\mathbf{1}'diag(\boldsymbol{\tau}^{-1})\mathbf{R}^{-1}diag(\boldsymbol{\tau}^{-1})\mathbf{y}}{\mathbf{1}'diag(\boldsymbol{\tau}^{-1})\mathbf{R}^{-1}diag(\boldsymbol{\tau}^{-1})\mathbf{1}} \\ &= \frac{\mathbf{c}'diag(\mathbf{R}\mathbf{c})\mathbf{y}}{\mathbf{c}'\mathbf{R}\mathbf{c}}. \end{aligned} \tag{8}$$

Since $\mathbf{c} \geq 0$, we have the following result, which is in stark contrast to the GLS estimate of μ in ordinary kriging, where it can go outside the range of the data as we have observed in an example in the introduction.

Theorem 1. *In rational kriging, the GLS estimate of μ is a convex combination of $\{y_i\}_{i=1}^n$ and therefore, it will always be in the range $[\min_i y_i, \max_i y_i]$ for any positive definite correlation function.*

In order to understand if the GLS estimate from rational kriging is good or not, we need to define the notion of a "true value" for μ. Define the true value as the L_2-projection of the underlying function as in Tuo and Wu (2015):

$$\mu^* = \underset{\mu}{\operatorname{argmin}} \int_{\mathcal{X}} \{y(\mathbf{x}) - \mu\}^2 dF(\mathbf{x}) = \int_{\mathcal{X}} y(\mathbf{x}) dF(\mathbf{x}),$$

where $F(\cdot)$ is the distribution function of $\mathbf{x}$ with support $\mathcal{X}$ from which the input points are generated. Since μ^* is a convex combination of the $y(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, we can expect the rational kriging estimate $\hat{\mu}$ to be closer to μ^* than $\hat{\mu}_{OK}$ to μ^*. We will investigate this more in Section 3 using simulations.

The mean squared prediction error for the optimal rational kriging predictor is given by

$$MSPE = \frac{\nu^2}{\{\mathbf{r}(\mathbf{x})'\mathbf{c}\}^2}\{1 - \mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{r}(\mathbf{x})\}, \tag{9}$$

which can be used for uncertainty quantification. It can be computed only after specifying the parameter ν and the coefficients $\mathbf{c}$. Moreover, there are unknown parameters in the correlation function that need to be specified. We will develop their estimation procedure after introducing Gaussian Process (GP) in the next section.

2.2 Rational Gaussian Process

It is well known that the ordinary kriging predictor can be obtained as the posterior mean if we assume a GP prior for the true function that generated the data (Currin et al., 1991; Rasmussen and Williams, 2006). A similar framework can be developed for rational kriging.

Following Kang and Joseph (2016), assume

$$y(\mathbf{x}) = \mu + \frac{\nu}{\mathbf{r}(\mathbf{x})'\mathbf{c}} Z(\mathbf{x}), \quad Z(\mathbf{x}) \sim GP(0, R(\cdot)). \tag{10}$$

It is easy to show that

$$y(\mathbf{x})|\mathbf{y} \sim N\left(\hat{y}(\mathbf{x}), \frac{\nu^2}{\{\mathbf{r}(\mathbf{x})'\mathbf{c}\}^2}\{1 - \mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{r}(\mathbf{x})\}\right), \tag{11}$$

where $\hat{y}(\mathbf{x})$ is the rational kriging predictor given in (7) and the posterior variance is the same as the MSPE in (9). As alluded to in the introduction, (11) can be used for constructing the prediction intervals, which is a major advantage of GPs over RBFs.

There are several unknown parameters in (11): μ, ν, and $\mathbf{c}$. The correlation function also has unknown parameters; denote them by $\boldsymbol{\theta}$. Among all these parameters, we will give a fully Bayesian treatment only for μ. All the other parameters will be estimated or specified as follows.

The likelihood is given by

$$\mathbf{y}|\mu, \nu, \mathbf{c}, \boldsymbol{\theta} \sim N(\mu\mathbf{1}, \nu^2 diag(\mathbf{1}/\mathbf{Rc})\ \mathbf{R}\ diag(\mathbf{1}/\mathbf{Rc})).$$

Assuming a non-informative prior for μ: $p(\mu) \propto 1$, we obtain

$$\mu|\mathbf{y}, \nu, \mathbf{c}, \boldsymbol{\theta} \sim N\left(\hat{\mu}, \frac{\nu^2}{\mathbf{c}'\mathbf{R}\mathbf{c}}\right),$$

where $\hat{\mu}$ is the GLS estimate of μ given in (8). Looking at the posterior variance of μ, it is tempting to choose a $\mathbf{c} \geq 0$ to maximize $\mathbf{c}'\mathbf{R}\mathbf{c}$. In fact, an elegant solution to this optimization problem exists. Under the constraint $\|\mathbf{c}\|_2 = 1$, the quadratic form $\mathbf{c}'\mathbf{R}\mathbf{c}$ is maximized by the eigenvector corresponding to the largest eigenvalue of $\mathbf{R}$. Since $\mathbf{R}$ is a symmetric positive matrix, this eigenvector can be chosen to be positive by Perron's theorem (Perron, 1907). This unique eigenvector is also known as Perron eigenvector. Thus,

we have the following result.

Theorem 2. *In rational kriging, the posterior variance of μ can be minimized by taking* $\mathbf{c}$ *to be the Perron eigenvector of* $\mathbf{R}$.

Buhmann et al. (2020) also suggests using this estimate for $\mathbf{c}$. Their suggestion is based on minimizing the native space norm of functions with kernel $K(\mathbf{u}, \mathbf{v}) = R(\mathbf{u}-\mathbf{v})$. With this choice of $\mathbf{c}$, $\mathbf{r}(\mathbf{x})'\mathbf{c}$ can be viewed as the Nyström approximation of the first eigenfunction of $R(\cdot)$ (Rasmussen and Williams, 2006, Sec. 4.3.2). In our trials, we found this estimate of $\mathbf{c}$ to work well when the functions are smooth, but poorly when the functions are non-smooth. This is because $\mathbf{r}(\mathbf{x})'\mathbf{c}$ can become very small for some value of $\mathbf{x}$, which can inflate the variance term in (11).

Another possibility is to let $\mathbf{c} = \mathbf{R}^{-1}\mathbf{1}$ as in limit kriging (Joseph, 2006), but this does not ensure nonnegativity of $\mathbf{c}$. We can overcome the nonnegativity issue as follows. Let $\hat{\gamma}$ be the smallest $\gamma \in [0, 1]$ such that $[(1-\gamma)\mathbf{R} + \gamma\mathbf{I}]^{-1}\mathbf{1} \geq \Delta\mathbf{1}$ component-wise, where $\Delta \in [0, 1]$. Such a $\hat{\gamma}$ always exists because $\gamma = 1$ trivially satisfies the inequality. Therefore, let

$$\hat{\mathbf{c}} = [(1-\hat{\gamma})\mathbf{R} + \hat{\gamma}\mathbf{I}]^{-1}\mathbf{1}. \tag{12}$$

Empirically, we found that $\Delta = \lambda_1/n$ works well, where λ_1 is the largest eigenvalue of $\mathbf{R}$.

When correlations are high, $\mathbf{R} \approx \lambda_1\mathbf{E}_1\mathbf{E}_1'$, where $\mathbf{E}_1$ is the eigenvector corresponding to λ_1. Then,

$$\mathbf{R}\hat{\mathbf{c}} \approx \frac{\lambda_1}{(1-\hat{\gamma})\lambda_1 + \hat{\gamma}}\mathbf{E}_1\mathbf{E}_1'\mathbf{1} \propto \lambda_1\mathbf{E}_1 = \mathbf{R}\mathbf{E}_1.$$

That is, the solution given in (12) behaves exactly like the eigenvector solution of Buhmann et al. (2020) when correlations are high (smooth functions). On the other hand, when correlations are small (non-smooth functions), $\hat{\mathbf{c}} \approx [(1-\hat{\gamma})\mathbf{I} + \hat{\gamma}\mathbf{I}]^{-1}\mathbf{1} \propto \mathbf{1}$, whereas $\mathbf{E}_1$ will be approximately the unit vector $(1, 0, \ldots, 0)'$. When this happens, the eigenvector solution will make $\mathbf{r}(\mathbf{x})'\mathbf{c} \approx 0$ for a large portion of $\mathcal{X}$, whereas $\mathbf{r}(\mathbf{x})'\hat{\mathbf{c}} \approx 1$ for $\mathbf{x}$ in the neighborhood of the observed data points. Thus, the solution given in (12) will be better behaved in all

correlations regimes and therefore, will be adopted in this article. We also note that this solution is quite different from that of Kang and Joseph (2016), where they estimate $\mathbf{c}$ by maximizing the unnormalized posterior, which is computationally prohibitive.

Thus,

$$\begin{aligned} p(\nu, \boldsymbol{\theta}|\mathbf{y}, \hat{\mathbf{c}}) &\propto \int p(\mathbf{y}|\mu, \nu, \hat{\mathbf{c}}, \boldsymbol{\theta})\, d\mu \\ &\propto \frac{|diag(\mathbf{R}\hat{\mathbf{c}})|}{\nu^{n-1}|\mathbf{R}|^{1/2}(\hat{\mathbf{c}}\mathbf{R}\hat{\mathbf{c}})^{1/2}} \exp\left\{-\frac{1}{2\nu^2}(\mathbf{y}-\hat{\mu}\mathbf{1})' diag(\mathbf{R}\hat{\mathbf{c}})\mathbf{R}^{-1} diag(\mathbf{R}\hat{\mathbf{c}})(\mathbf{y}-\hat{\mu}\mathbf{1})\right\}, \end{aligned}$$

where $\hat{\mathbf{c}}$ is given in (12). Maximizing this with respect to ν and $\boldsymbol{\theta}$, we obtain

$$\hat{\nu}^2 = \frac{1}{n-1}(\mathbf{y}-\hat{\mu}\mathbf{1})' diag(\mathbf{R}\hat{\mathbf{c}})\mathbf{R}^{-1} diag(\mathbf{R}\hat{\mathbf{c}})(\mathbf{y}-\hat{\mu}\mathbf{1}), \tag{13}$$

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\{ (n-1)\log\hat{\nu}^2 + \log|\mathbf{R}| - 2\sum_{i=1}^{n}\log(\mathbf{R}_i\hat{\mathbf{c}}) + \log(\hat{\mathbf{c}}'\mathbf{R}\hat{\mathbf{c}}) \right\}, \tag{14}$$

where $\mathbf{R}_i$ is the ith row of $\mathbf{R}$.

Rational kriging/GP can be used with any positive definite correlation function. One of the most commonly used correlation function in computer experiments is the Gaussian correlation function given by $R(\mathbf{h}) = \exp\{-\sum_{i=1}^{p}(h_i/\theta_i)^2\}$. Let $\theta_i^2 = \theta^2/w_i$, where $\sum_{i=1}^{p} w_i = 1$ and $w_i \geq 0$ for $i = 1, \ldots, n$. Then the Gaussian correlation function can be written as

$$R(\mathbf{h}) = \exp\left\{-\|\mathbf{h}\|_w^2/\theta^2\right\}, \tag{15}$$

where $\|\mathbf{h}\|_w^2 = \sum_{i=1}^{p} w_i h_i^2$. It is interesting to study the properties of the rational kriging predictor when the length-scale parameter (θ) becomes small. Using a result in Kang and Joseph (2016), it is easy to show that the rational kriging tends to the nearest neighbor predictor defined by the norm $\|\cdot\|_w$ as $\theta \to 0$. This property helps rational kriging to overcome the "mean reversion" problem commonly observed with ordinary kriging.

There is another correlation function that makes the foregoing limiting case even more

interesting. Consider the rational quadratic function (Rasmussen and Williams, 2006) (also known as Cauchy function) given by

$$R(\mathbf{h}) = \left(1 + \|\mathbf{h}\|_w^2/\theta^2\right)^{-1}. \tag{16}$$

When the length-scale parameter $\theta \to 0$, we have $\mathbf{R} \to \mathbf{I}$ and therefore $\hat{\mathbf{c}} \to \mathbf{1}$. Moreover, $R(\mathbf{x} - \mathbf{x}_i)/\mathbf{r}(\mathbf{x})'\hat{\mathbf{c}} \to \|\mathbf{x} - \mathbf{x}_i\|_w^{-2} / \sum_{j=1}^n \|\mathbf{x} - \mathbf{x}_j\|_w^{-2}$. The predictor

$$\hat{y}_{IDW}(\mathbf{x}) = \frac{\sum_{i=1}^n \|\mathbf{x} - \mathbf{x}_i\|_w^{-2} y_i}{\sum_{j=1}^n \|\mathbf{x} - \mathbf{x}_j\|_w^{-2}}$$

is the well-known inverse distance weighting (IDW) predictor (Shepard, 1968; Joseph and Kang, 2011). Thus, we have the following result.

Theorem 3. *Under rational quadratic correlation function in (16), the rational kriging predictor converges to the inverse distance weighting predictor as the length-scale parameter goes to 0.*

2.3 Generalized Rational Kriging

The mean squared prediction error of rational kriging (9) can go to ∞ as $\mathbf{r}(\mathbf{x})'\mathbf{c} \to 0$. Therefore, the confidence intervals can blow up when predicting in data sparse regions of $\mathcal{X}$, especially with non-smooth functions. To see this, consider a non-smooth function from Gramacy and Lee (2012):

$$y = \frac{\sin 10\pi x}{2x} + (x-1)^4, \quad x \in [0.5, 2.5].$$

The predictions and 95% confidence intervals from ordinary and rational using a rational quadratic correlation function are shown in Figure 2. The predictions from the two methods are quite similar, but confidence intervals differ a lot. We can see that the confidence intervals from rational kriging are unnecessarily wide in at least two regions.

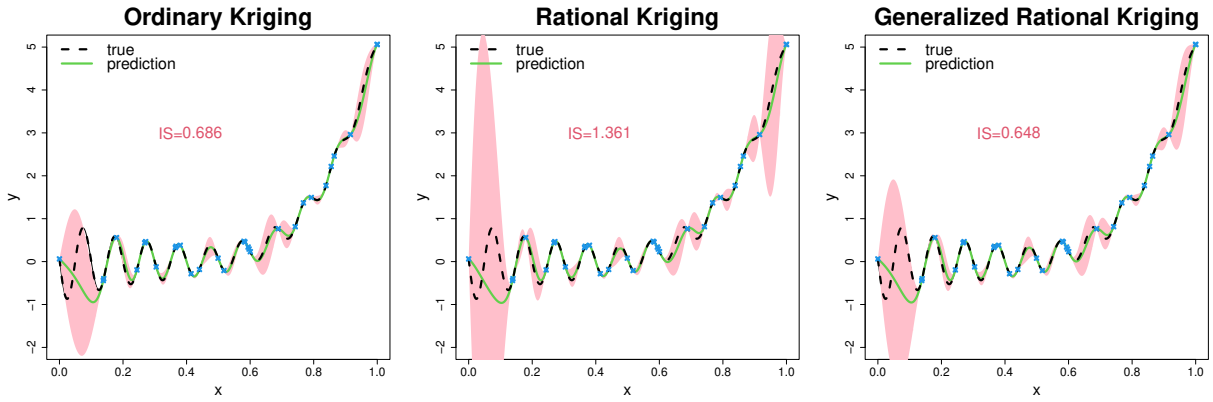

Figure 2: *Ordinary (left), rational (middle), and the generalized rational (right) kriging fitted on* $n = 30$ *data points (blue crosses). The 95% confidence intervals for the predictions are shown as a shaded region.*

We can avoid the foregoing issue by adding a positive constant to the denominator term $\mathbf{r}(\mathbf{x})'\mathbf{c}$. Therefore, consider a new Gaussian process model

$$y(\mathbf{x}) = \mu + \frac{\nu}{c_0 + \mathbf{r}(\mathbf{x})'\mathbf{c}} Z(\mathbf{x}), \quad Z(\mathbf{x}) \sim GP(0, R(\cdot)), \tag{17}$$

where $c_0 \geq 0$,$\mathbf{c} \geq 0$, and $c_0^2 + \|\mathbf{c}\|_2^2 = 1$. We can see that rational kriging and ordinary kriging can be obtained as special cases of this general model by setting $c_0 = 0$ and $c_0 = 1$, respectively. Furthermore, for $c_0 > 0$, the variance of $y(\mathbf{x})$ in (17) is bounded by ν^2/c_0^2.

Let $\tilde{\mathbf{c}} = (c_0, \mathbf{c}')'$ and $\tilde{\mathbf{R}} = [\mathbf{1}\ \mathbf{R}]$. As before, assuming $p(\mu) \propto 1$, it can be shown that

$$y(\mathbf{x})|\mathbf{y} \sim N\left(\tilde{y}(\mathbf{x}), s^2(\mathbf{x})\right),$$

where

$$\tilde{y}(\mathbf{x}) = \tilde{\mu} + \frac{\mathbf{r}(\mathbf{x})'}{c_0 + \mathbf{r}(\mathbf{x})'\mathbf{c}} \mathbf{R}^{-1} diag(\tilde{\mathbf{R}}\tilde{\mathbf{c}})(\mathbf{y} - \tilde{\mu}\mathbf{1}), \tag{18}$$

$$s^2(\mathbf{x}) = \frac{\nu^2}{\{c_0 + \mathbf{r}(\mathbf{x})'\mathbf{c}\}^2} \left\{ 1 - \mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}) + c_0^2 \frac{(1 - \mathbf{r}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{1})^2}{\tilde{\mathbf{c}}'\tilde{\mathbf{R}}'\mathbf{R}^{-1}\tilde{\mathbf{R}}\tilde{\mathbf{c}}} \right\}, \tag{19}$$

and

$$\tilde{\mu} = \frac{\tilde{\mathbf{c}}'\tilde{\mathbf{R}}'\mathbf{R}^{-1}diag(\tilde{\mathbf{R}}\tilde{\mathbf{c}})\mathbf{y}}{\tilde{\mathbf{c}}'\tilde{\mathbf{R}}'\mathbf{R}^{-1}\tilde{\mathbf{R}}\tilde{\mathbf{c}}}. \tag{20}$$

Furthermore, the posterior distribution of μ is given by

$$\mu|\mathbf{y} \sim N\left(\tilde{\mu}, \frac{\nu^2}{\tilde{\mathbf{c}}'\tilde{\mathbf{R}}'\mathbf{R}^{-1}\tilde{\mathbf{R}}\tilde{\mathbf{c}}}\right).$$

Since $\mathbf{R}$ is a positive definite matrix, $\mathbf{1}'\mathbf{R}^{-1}\mathbf{1} > 0$. Thus,

$$\tilde{\mathbf{R}}'\mathbf{R}^{-1}\tilde{\mathbf{R}} = \begin{bmatrix} \mathbf{1}'\mathbf{R}^{-1}\mathbf{1} & \mathbf{1}' \\ \mathbf{1} & \mathbf{R} \end{bmatrix}$$

is a symmetric positive matrix. Therefore, by Perron's theorem (Perron, 1907), the eigenvector of $\tilde{\mathbf{R}}'\mathbf{R}^{-1}\tilde{\mathbf{R}}$ corresponding to its largest eigenvalue can be chosen to be positive, which gives the following result.

Theorem 4. *For generalized rational kriging, the posterior variance of μ can be minimized by taking $(c_0, \mathbf{c}')'$ to be the Perron eigenvector of $\tilde{\mathbf{R}}'\mathbf{R}^{-1}\tilde{\mathbf{R}}$.*

For non-smooth functions, the length-scale parameters in the correlation function will be small. So let $\theta \to 0$ in (15) or (16). It is easy to show that

$$(c_0, \mathbf{c}') \to \frac{1}{\sqrt{n(n+1)}}(n, 1, 1, \ldots, 1) \text{ as } \theta \to 0.$$

Therefore, $c_0 + \mathbf{r}(\mathbf{x})'\mathbf{c}$ will not go to 0, which will avoid the inflation of the variance in (19) in data sparse regions of $\mathcal{X}$. The right panel of Figure 2 shows the generalized rational kriging predictor and its 95% confidence intervals, which look very reasonable compared to those of rational and ordinary kriging.

3 Simulations

Consider again the beam deflection function used in the introduction: $y = -x(1 - 2x^2 + x^3)$ for $x \in [0, 1]$. Let $x_i \overset{iid}{\sim} U(0, 1)$ for $i = 1, \ldots, 11$. These points are re-scaled such that $x_1 = 0$ and $x_{11} = 1$. Ordinary kriging (OK), rational kriging (RK), and generalized rational kriging (GRK) are fitted to the data using the Gaussian correlation function $R(h) = \exp\{-(h/\theta)^2\}$ and Rational Quadratic function $R(h) = \{1 + (h/\theta)^2\}^{-1}$. This simulation is repeated for 50 times. The left panel of Figure 3 shows the Root Mean-Squared Errors (RMSEs) computed over a grid of 1,001 points in $[0, 1]$. It shows that, on the average, RK is more accurate than OK, and GRK is more accurate than RK. Interval Score (Gneiting and Raftery, 2007)

$$IS = \frac{1}{N} \sum_{i=1}^{N} \left[(u - l) + \frac{2}{\alpha} \{ (l - t_i)_+ + (t_i - u)_+ \} \right]$$

is computed for assessing the accuracy of $(1 - \alpha)$ confidence intervals $[l, u]$, where $(x)_+ = x$ if $x > 0$ and 0 otherwise, and $\{t_i\}_{i=1}^N$ are the $N = 1001$ testing locations. This is shown in the middle panel of Figure 3 for 95% confidence intervals. A small IS value indicates better confidence intervals (small width at prescribed coverage). In this example, IS shows comparable performance for the three methods. The most striking result is the plot on the right panel of Figure 3. While OK produces $\hat{\mu}$'s much larger than the maximum value of y_i's, the estimate from RK is around the true value $\mu^* = \int_0^1 -x(1 - 2x^2 + x^3)\, dx = -0.20$ (shown as a red line in the same figure). Although the result of Theorem 1 does not hold true for GRK, the estimate of mean from GRK is only slightly worse than that of RK.

For OK, a priori, 95% of the function is believed to lie in $\mu \pm 2\tau$, whereas for RK/GRK the 95% prior confidence interval is $\mu \pm 2\nu/(c_0 + \mathbf{r}(x)'\mathbf{c})$. They are plotted in Figure 4 by setting $\mu = 0$, $\tau = 1$, and choosing ν to be the harmonic mean of $c_0 + \mathbf{r}(x_i)'\mathbf{c}$, $i = 1, \ldots, n$, where $x_i = (i - 1)/(n - 1)$. We can see that they pretty much agree within the input region $[0, 1]$. Outside $[0, 1]$, the confidence intervals for OK remain constant, whereas they increase for RK/GRK. In other words, OK assigns equal "weight" to the whole of $\mathbb{R}$, whereas

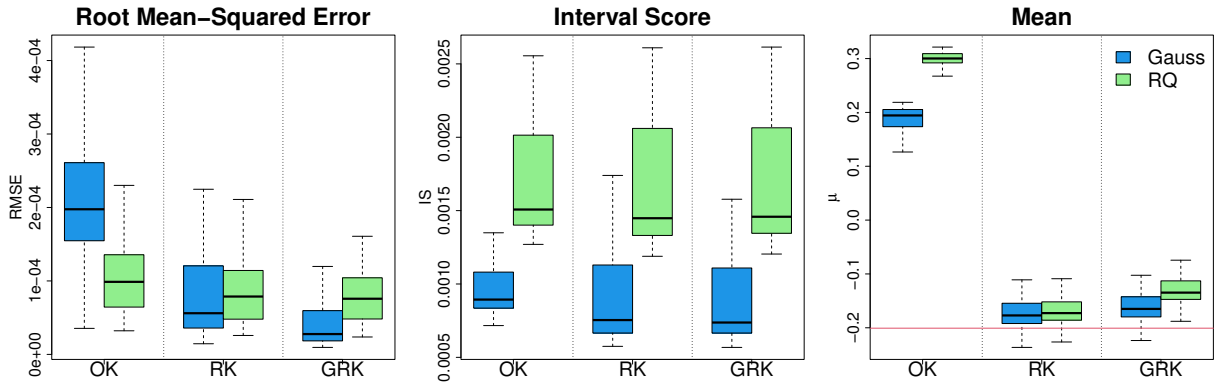

Figure 3: *Boxplots of RMSE (left), IS (middle), and estimate of μ (right) from the simulation using the beam deflection function. The simulation is done by randomly sampling $\{x_i\}_{i=1}^{11}$ from $[0, 1]$. Outliers are not shown. The true value μ^* is plotted as a red line in the right plot.*

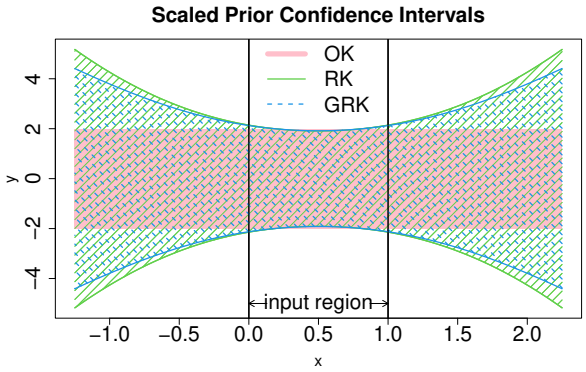

Figure 4: *Scaled 95% prior confidence regions for the function with mean centered at 0 are shown as shaded regions for OK, RK, and GRK. Eleven equally-spaced points in $[0, 1]$ is used as the design.*

RK/GRK assigns more "weight" in the input region and less "weight" outside the input region. This could be the reason why the estimates from RK and GRK are well behaved.

Simulations with three other one-dimensional test functions under a similar setup are given in Figure 5 except that we used $n = 30$. The details of the test functions are given in the supplementary materials. The first two are non-smooth functions and we can see that OK does better than RK on prediction (RMSE) and uncertainty quantification (IS),

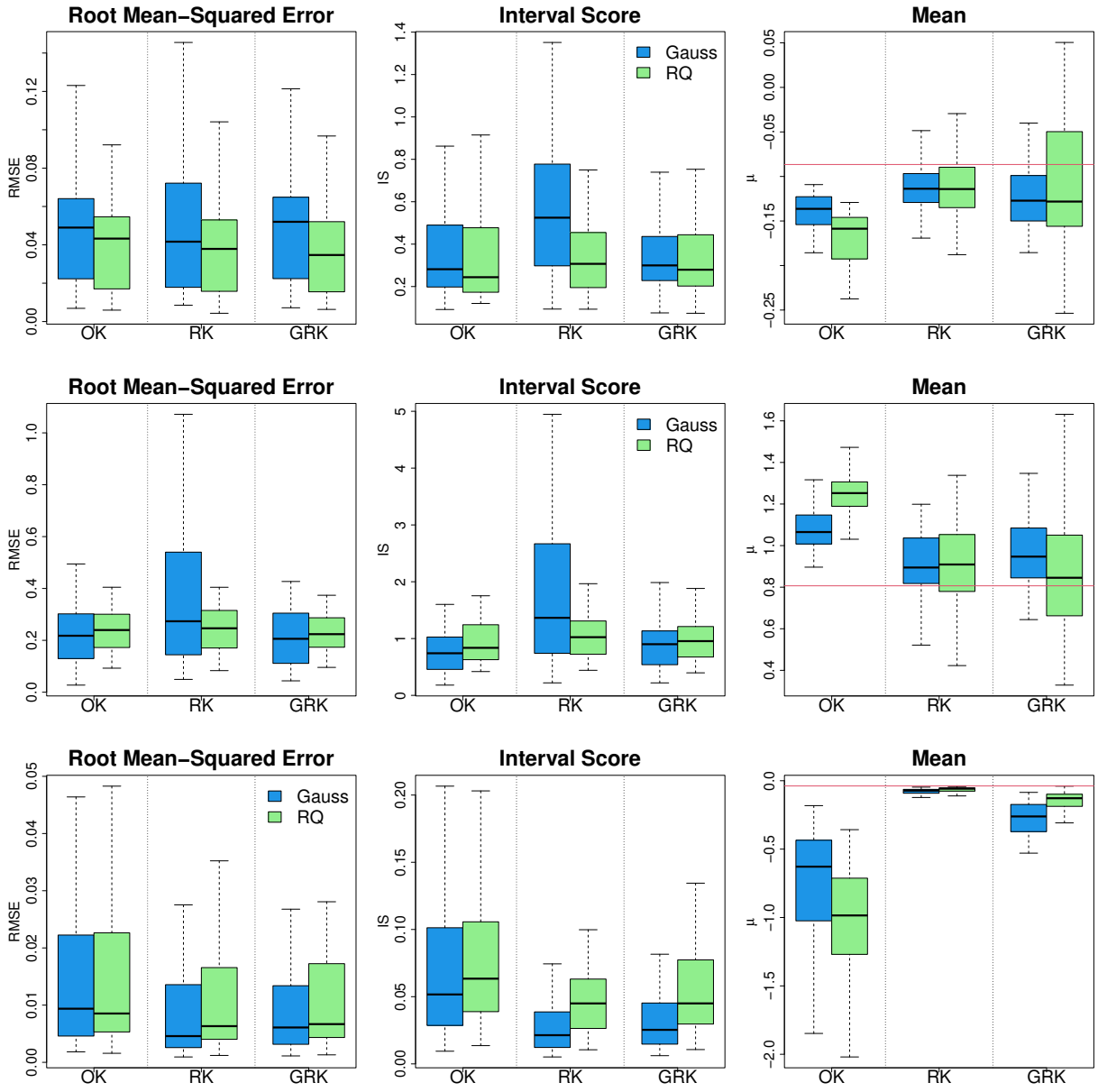

Figure 5: *Boxplots of RMSE (left column), IS (middle column), and estimate of μ (right column) from the simulation using Xiong et al. function (top row), Gramacy-Lee function (middle row), and Buhmann et al. function (bottom row). The simulation is done by randomly sampling $\{x_i\}_{i=1}^{30}$ from $[0, 1]$. Outliers are not shown. The true value μ^* is plotted as a red line in the right panels.*

whereas the last function is a smooth function in which RK performs better than OK. In all the three functions GRK's performance is as good as or even better than both OK and RK. Clearly, RK does the best in terms of estimating the mean, followed by GRK. Although the Gaussian correlation function performed better for RK and GRK for the third function, the performance with rational quadratic correlation function is found to be more stable.

The simulation of the third function is repeated by varying n from 100 to 1000 using an equally-spaced design and rational quadratic correlation function. Figure 6 shows the RMSE and the computational time for fitting the three methods. The simulation is done on a 2.40 GHz desktop with 16.0 GB memory. The RMSE shows that all the three methods seem to have the same rate of convergence (we used a fixed nugget of 10^{-6} for numerical stability). The computational time for RK is high because of the full eigen decomposition required for computing the $\mathbf{c}$ in (12), whereas GRK is as fast as OK because the Perron eigenvector of $\tilde{\mathbf{R}}'\mathbf{R}^{-1}\tilde{\mathbf{R}}$ can be computed in nearly linear time. In our implementation, we have used the R package `RSpectra` (Qiu et al., 2022) for computing the Perron eigenvector.

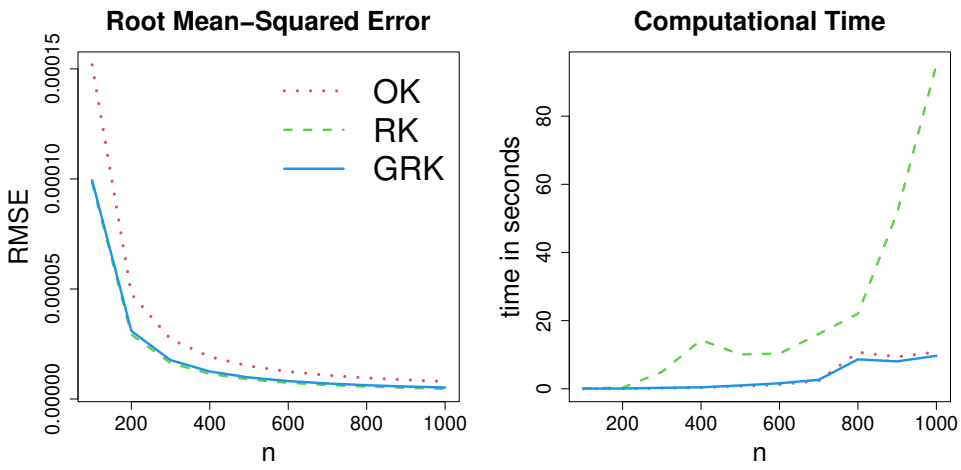

Figure 6: *The left plot shows the RMSE of OK, RK, and GRK on the Bhumann et al. function by varying n from 100 to 1,000, and the right plot shows the corresponding computational time.*

4 Applications

In this section, we use rational kriging and generalized rational kriging in two important applications of computer experiments: emulation and calibration.

4.1 Emulation

Joseph (2016) reported a case study on the robust parameter design optimization of a solid end milling process for which an accurate emulator is needed. A 60-run MaxPro design (Joseph et al., 2015) was generated for the six variables: hardness of the workpiece material (x_1), rake angle (x_2), helix angle (x_3), relief angle (x_4), corner radius (x_5), and flute length (x_6) of the cutting tool. The simulations were carried out using the commercially available Production Module software (Third Wave Systems, Minneapolis, MN). We will use the force on the tool (averaged over time) as the response.

To evaluate the different kriging methods the dataset is split into training and testing sets in the ratio 4:1. We used the SPlit method (Joseph and Vakayil, 2022) to get a reliable result. For a given training set, OK, RK, and GRK were fitted using the Gaussian and rational quadratic correlation functions and the RMSE, IS, and mean were computed on the test set. This process is repeated 50 times and the results are summarized in Figure 7. The average of the force values over the 60 runs is used as the "true value" of μ and is plotted as a red line in the right plot of the figure. We can see that both RK and GRK give slightly better prediction and uncertainty quantification than those of OK and at the same time give substantially improved estimates of μ. Thus, in this case study, RK and GRK are able to provide more accurate emulators than that of OK, which will help in obtaining more reliable results for the robust settings of the process variables.

For further testing, three test functions that are widely used for emulation in computer experiments are chosen: 8-dimensional borehole function (Morris et al., 1993), 7-dimensional piston simulation function (Kenett and Zacks, 2021), and 6-dimensional OTL circuit function (Ben-Ari and Steinberg, 2007). The details of these functions are given in the supplementary

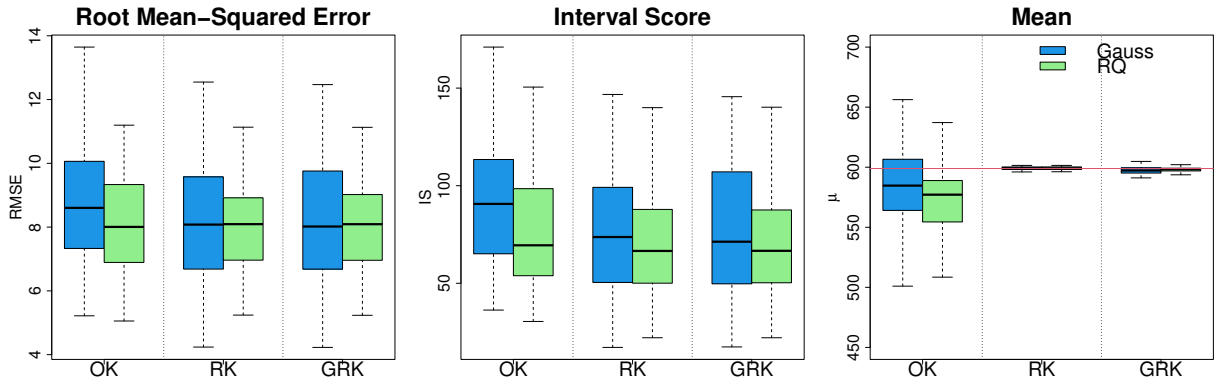

Figure 7: *Boxplots of RMSE (left), IS (middle), and estimate of μ (right) from the milling process dataset. The simulation is done by splitting the dataset into training (48 runs) and testing sets (12 runs) 50 times. The average value of the force from the 60 runs is plotted as a red line in the right plot.*

materials. The simulations are carried out by randomly sampling $n = 10p$ points from $[0, 1]^p$ and the results are evaluated on a test set by randomly sampling 1,001 points from $[0, 1]^p$. The results are summarized in Figure 8. In all the three test functions, RK's and GRK's prediction and uncertainty quantification performance are found to be superior to OK along with better and more consistent estimates for the mean.

4.2 Calibration

Consider a physics-based model $y = f(\mathbf{x}; \boldsymbol{\eta})$, where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_q)'$ are the unknown calibration parameters that need to be estimated from the real data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$. Since the physics-based model could be biased, Kennedy and O'Hagan (2001) proposed to use a Gaussian process model to capture the discrepancy between the physics-based model and the data. Their model can be written as

$$y = f(\mathbf{x}; \boldsymbol{\eta}) + \tau\delta(\mathbf{x}) + \epsilon,\ \delta(\mathbf{x}) \sim GP(0, R(\cdot)) \text{ and } \epsilon \overset{iid}{\sim} N(0, \sigma^2). \tag{21}$$

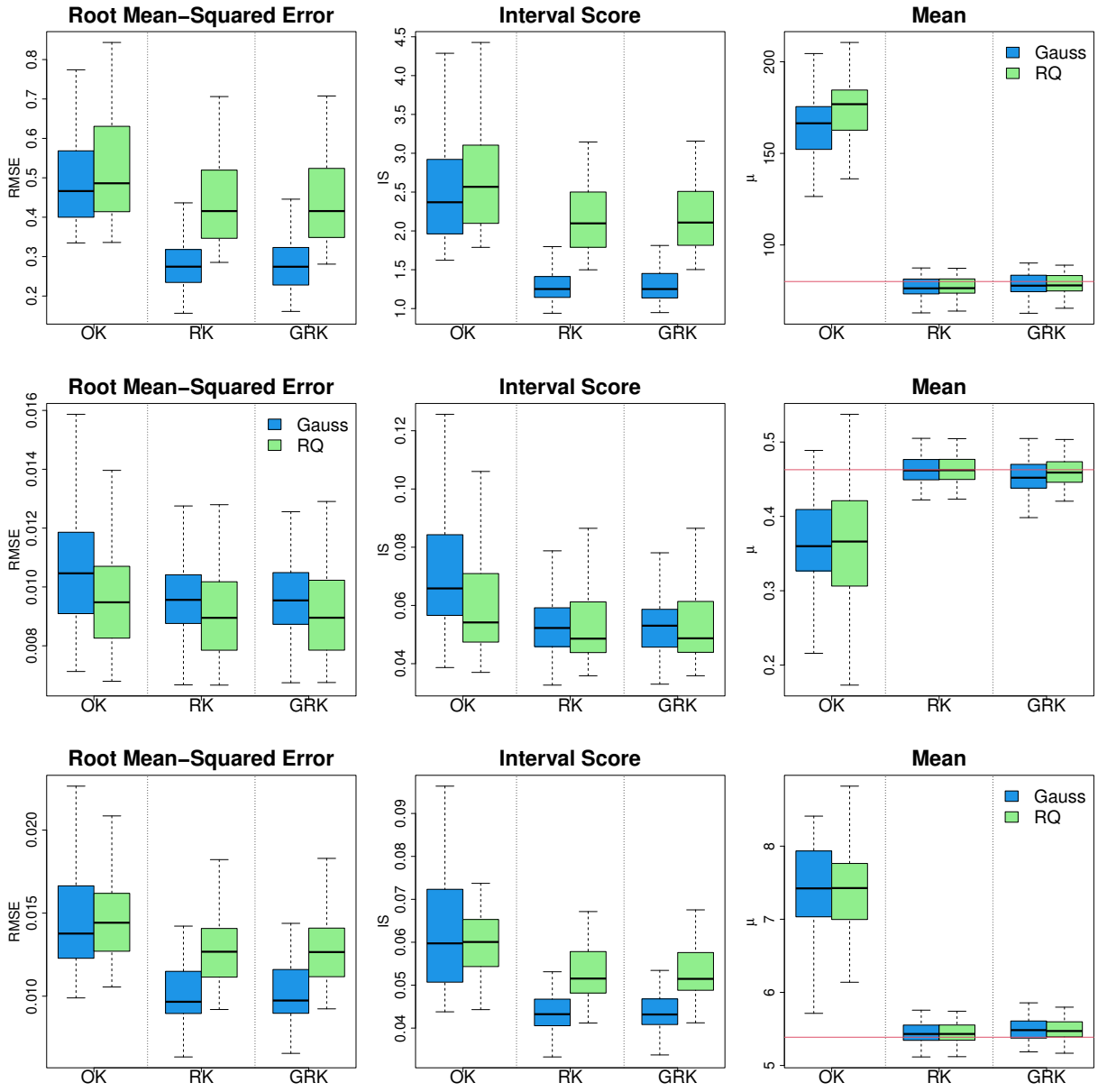

Figure 8: *Boxplots of RMSE (left column), IS (middle column), and estimate of μ (right column) from the simulation using borehole function (top row), piston simulation function (middle row), and OTL circuit function (bottom row). The simulation is done by randomly sampling $n = 10p$ points from $[0, 1]^p$. The true value μ^* is plotted as a red line in the right panels.*

Tuo and Wu (2015, 2016) have shown that this model could produce poor estimates of $\boldsymbol{\eta}$ because of the non-identifiability between $\boldsymbol{\eta}$ and $\delta(\cdot)$. Since then, several proposals have appeared in the literature aimed at tackling the identifiability issue (Wong et al., 2017; Plumlee, 2017; Gu and Wang, 2018; Tuo, 2019). However, Kennedy-O'Hagan (KOH) method continue to maintain its popularity among the practitioners, which could be due to its simplicity and ease of implementation.

Encouraged by the results of previous sections, we could consider using rational or generalized rational GP in the KOH model:

$$y = f(\mathbf{x}; \boldsymbol{\eta}) + \frac{\nu}{c_0 + \mathbf{r}(\mathbf{x})'\mathbf{c}}\delta(\mathbf{x}) + \epsilon,\ \delta(\mathbf{x}) \sim GP(0, R(\cdot)) \text{ and } \epsilon \overset{iid}{\sim} N(0, \sigma^2). \tag{22}$$

We make no claims about overcoming the identifiability issue with this new model. Our hope is that this model would produce better estimates of $\boldsymbol{\eta}$ than with the original KOH model without sacrificing the quality of prediction and uncertainty quantification. Moreover, this new model requires only a rescaling of the discrepancy term, which is easy to implement in practice.

Consider a simple example from Plumlee (2017). Suppose $f(x; \eta) = \eta x$, but the data is generated from $y = 4x + x\sin(5x) + \epsilon$ with $\epsilon \overset{iid}{\sim} N(0, 0.02^2)$. Input values are generated by taking 17 equally spaced points in $[0, 0.8]$. We will illustrate the results using rational kriging, i.e., $c_0 = 0$ in (22). Since $f(\cdot)$ is linear in η, we can use the results of the rational version of the universal kriging model given in the online supplementary file with $\boldsymbol{\Sigma} = diag(\mathbf{1}/\mathbf{R}\hat{\mathbf{c}})\mathbf{R}diag(\mathbf{1}/\mathbf{R}\hat{\mathbf{c}}) + \sigma^2/\nu^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Figure 9 shows the plot of $\hat{\eta}$ for various values of θ using Gaussian and rational quadratic correlation functions. The least squares estimate of η is around 4.0 and is plotted in the same figure as a red dotted line. We can see that the estimates of η from the rational version of the Kennedy-O'Hagan (RK-KOH) model are much closer to the least squares estimate than those from the original KOH model for both the correlation functions. Clearly there is bias from the RK-KOH, but at least the use of rational kriging seems to stabilize the parameter estimates making $\hat{\eta}$ more

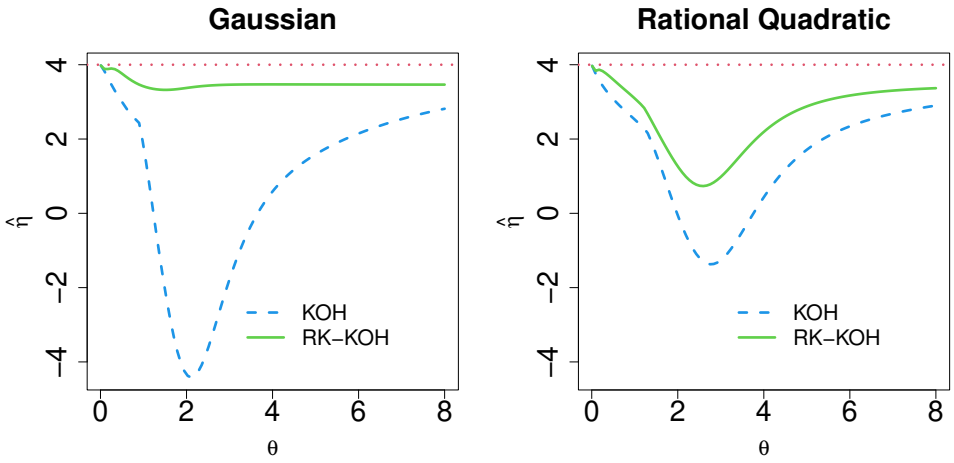

Figure 9: *Plot of $\hat{\eta}$ over various values of the lengthscale parameter θ using original Kennedy-O'Hagan (KOH) model and rational version of the Kennedy-O'Hagan (RK-KOH) model. The least squares estimate of η is shown as a red dotted line.*

robust to the misspecification of the correlation parameters. In fact, the theoretical results of Tuo and Wu (2015, 2016) show that the dependency of $\hat{\eta}$ with the Gaussian process prior specification is the main drawback of the KOH method. We can see that rational kriging can at least help in mitigating this dependency, which can be a major step towards overcoming the identifiability issue.

5 Conclusions

Although ordinary kriging has been widely used in statistics, the generalized least squares estimate of the mean parameter can sometimes be nonsensical. This issue has been largely ignored in the literature because prediction and uncertainty quantification can still be good if the correlation parameters are carefully tuned. Therefore, many practitioners replace the generalized least squares estimate of the mean with ordinary least squares estimate. However, this leads to inconsistencies in the modeling framework, especially when Bayesian modeling is applied. Furthermore, there are situations such as in model calibration problems, where the

parameters in the mean function have physical interpretation and thus meaningful estimates of them are desired. The rational kriging proposed in this article seems to overcome these issues. It gives comparable prediction and uncertainty quantification to those of ordinary kriging, but with substantially improved estimates for the mean parameters. This is achieved by simply scaling the stochastic part of the kriging/Gaussian process by a scaling function. Therefore, the proposed method can be easily implemented in complex statistical models.

The rational kriging predictor is found to be the same as the limiting case of an iterated kernel regression proposed in Kang and Joseph (2016), but its optimality in terms of minimum mean squared prediction error was not known earlier. We also proposed an efficient estimator of the scaling function used in rational kriging which turned out to be closely related to the first eigenfunction of the correlation function. A similar estimation procedure was also proposed in Buhmann et al. (2020) for the rational radial basis function interpolation. However, unlike rational kriging, the rational radial basis function techniques do not provide any uncertainty quantification. Our simulations discovered some of their pitfalls with the uncertainty quantification of non-smooth functions. Therefore, we have also proposed a generalized version of rational kriging in which rational and ordinary kriging are special cases. We found the performance of generalized rational kriging to be comparable or better than both ordinary and rational kriging for all types of functions. Moreover, the computational complexity of fitting a generalized rational kriging is no more than that of an ordinary kriging.

The rational kriging provides a new perspective for kriging with a nonstationary variance function. From the inception of the kriging technique, constant variance has been widely used for the stochastic component of the statistical model. This is under the assumption of stationarity that the true function is expected to lie within a constant band throughout the region of interest. This approach works well when the true function is indeed stationary. However, in practice, we never know if it is stationary or not. Thus, it makes sense to place a prior that has narrow confidence interval in the region of data collection and that becomes wider as the prediction point deviates from the input region of data (see Figure 4).

This introduces a fundamental shift in the way we deal with kriging and Gaussian process models.

Acknowledgments

This research is supported by a U.S. National Science Foundation grant DMS-2310637.

Supplementary Materials

`supp.pdf`: contains: (i) details of the test functions used in the simulations, (ii) universal kriging, and (iii) comparisons to fractional Brownian field.

`codedata.zip`: contains the R codes and data.

References

Ben-Ari, E. N. and Steinberg, D. M. (2007). Modeling data from computer experiments: an empirical comparison of kriging with mars and projection pursuit regression. *Quality Engineering*, 19(4):327–338.

Buhmann, M. D., De Marchi, S., and Perracchione, E. (2020). Analysis of a new class of rational RBF expansions. *IMA Journal of Numerical Analysis*, 40(3):1972–1993.

Chen, J., Mak, S., Joseph, V. R., and Zhang, C. (2021). Function-on-function kriging, with applications to three-dimensional printing of aortic tissues. *Technometrics*, 63(3):384–395.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.

Gramacy, R. B. and Lee, H. K. (2012). Cases for the nugget in modeling computer experiments. *Statistics and Computing*, 22:713–722.

Gu, M. and Wang, L. (2018). Scaled gaussian stochastic process for computer model calibration and prediction. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1555–1583.

Jakobsson, S., Andersson, B., and Edelvik, F. (2009). Rational radial basis function interpolation with applications to antenna design. *Journal of computational and applied mathematics*, 233(4):889–904.

Joseph, V. R. (2006). Limit kriging. *Technometrics*, 48(4):458–466.

Joseph, V. R. (2016). Space-filling designs for computer experiments: A review. *Quality Engineering*, 28(1):28–35.

Joseph, V. R., Gul, E., and Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*, 102(2):371–380.

Joseph, V. R. and Kang, L. (2011). Regression-based inverse distance weighting with applications to computer experiments. *Technometrics*, 53(3):254–265.

Joseph, V. R. and Vakayil, A. (2022). SPlit: An optimal method for data splitting. *Technometrics*, 64(2):166–176.

Kang, L. and Joseph, V. R. (2016). Kernel approximation: From regression to interpolation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):112–129.

Kenett, R. S. and Zacks, S. (2021). *Modern industrial statistics: With applications in R, MINITAB, and JMP*. John Wiley & Sons.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

Mak, S., Sung, C.-L., Wang, X., Yeh, S.-T., Chang, Y.-H., Joseph, V. R., Yang, V., and Wu, C. F. J. (2018). An efficient surrogate model for emulation and physics extraction of large eddy simulations. *Journal of the American Statistical Association*, 113(524):1443–1456.

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266.

Morris, M. D., Mitchell, T. J., and Ylvisaker, D. (1993). Bayesian design and analysis of

computer experiments: use of derivatives in surface prediction. *Technometrics*, 35(3):243–255.

Perron, O. (1907). Zur theorie der matrices. *Mathematische Annalen*, 64(2):248–263.

Plumlee, M. (2017). Bayesian calibration of inexact computer models. *Journal of the American Statistical Association*, 112(519):1274–1285.

Plumlee, M. and Joseph, V. R. (2018). Orthogonal gaussian process models. *Statistica Sinica*, pages 601–619.

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(514):500–509.

Pronzato, L. and Zhigljavsky, A. (2023). BLUE against OLSE in the location model: energy minimization and asymptotic considerations. *Statistical Papers*, 64:1187–1208.

Qiu, Y., Mei, J., Guennebaud, G., and Niesen, J. (2022). *RSpectra: Solvers for Large-Scale Eigenvalue and SVD Problems*. R package version 0.16-1.

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. The MIT Press, Cambridge, MA.

Roustant, O., Ginsbourger, D., and Deville, Y. (2012). Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of statistical software*, 51:1–55.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.

Sarra, S. A. and Bai, Y. (2018). A rational radial basis function method for accurately resolving discontinuities and steep gradients. *Applied Numerical Mathematics*, 130:131–142.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, pages 517–524.

Tuo, R. (2019). Adjustments to computer models via projected kernel calibration. *SIAM/ASA Journal on Uncertainty Quantification*, 7(2):553–578.

Tuo, R. and Wu, C. F. J. (2015). Efficient calibration for imperfect computer models. *The Annals of Statistics*, 43(6):2331–2352.

Tuo, R. and Wu, C. F. J. (2016). A theoretical framework for calibration in computer models: Parametrization, estimation and convergence properties. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):767–795.

Wong, R. K., Storlie, C. B., and Lee, T. C. (2017). A frequentist approach to computer model calibration. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(2):635–648.

Zhang, N. and Apley, D. W. (2014). Fractional brownian fields for response surface metamodeling. *Journal of Quality Technology*, 46(4):285–301.