Operations Research

Generalized Integrated Brownian Fields for Simulation Metamodeling

Peter Salemi, Jeremy Staum, Barry L. Nelson

Generalized Integrated Brownian Fields for Simulation Metamodeling

Peter Salemi,[a] Jeremy Staum,[b] Barry L. Nelson[b]

[a] Operations Research Department, The MITRE Corporation, Bedford, Massachusetts 01730; [b] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208
Contact: psalemi@mitre.org, http://orcid.org/0000-0001-9040-8082 (PS); j-staum@northwestern.edu,
http://orcid.org/0000-0001-6163-2624 (JS); nelsonb@northwestern.edu, http://orcid.org/0000-0002-1325-2624 (BLN)

Abstract. We introduce a novel class of Gaussian random fields (GRFs), called generalized integrated Brownian fields (GIBFs), focusing on the use of GIBFs for Gaussian process regression in deterministic and stochastic simulation metamodeling. We build GIBFs from the well-known Brownian motion and discuss several of their properties, including differentiability that can differ in each coordinate, no mean reversion, and the Markov property. We explain why we desire to use GRFs with these properties and provide formal definitions of mean reversion and the Markov property for real-valued, differentiable random fields. We show how to use GIBFs with stochastic kriging, covering trend modeling and parameter fitting, discuss their approximation capability, and show that the resulting metamodel also has differentiability that can differ in each coordinate. Last, we use several examples to demonstrate superior prediction capability as compared with the GRFs corresponding to the Gaussian and Matérn covariance functions.

1. Introduction

Stochastic simulations are often used to model complex systems in industrial engineering and operations research. Because simulation models are typically not limited by the complexity of the underlying system, simulation runs may be time-consuming to execute, especially when there are many scenarios that need to be evaluated. This limits the use of simulation models for supporting real-time decision making. When the simulation model can be run for a significant amount of time before decisions must be made, we can use the output from the simulation to build a statistical model of the response surface. We call this statistical model a simulation metamodel. Using the metamodel, we can predict the value of the response surface for any scenario, even if it has not been simulated.

A great deal of research has been directed toward fitting linear regression models to simulation output. However, we are particularly interested in general metamodeling approaches that assume less structure than linear models. In the deterministic computer experiments literature, the use of Gaussian process (GP) regression has been remarkably successful for global metamodeling (Santner et al. 2010). Following the introduction of GP regression into the design and analysis of deterministic computer experiments, Mitchell and Morris (1992) introduced GP regression for representing the response surface in stochastic simulation. Because the predictions are made by fitting a GP, we are able to obtain a measure of uncertainty in predictions, which gives rise to confidence intervals. Furthermore, the measure of uncertainty in predictions facilitates sequential, adaptive experiment designs and can provide statistical inference about the fitted model (Ankenman et al. 2010).

In simulation metamodeling using GP regression, the response surface is modeled as a sample path of a Gaussian random field (GRF). A critical choice in fitting GP regression models is specifying the GRF. To obtain better prediction capability, the GRF should have desirable properties and be flexible enough to capture the characteristics of the response surface, such as its level of differentiability. A GRF is completely specified by its mean function (often assumed to be identically zero) and covariance function. Thus, selecting the proper covariance function is crucial for determining the prediction capability of the resulting GP regression model. Indeed, much research has been done that discusses the choice of covariance functions for GP regression (Paciorek and Schervish 2004, Santner et al. 2010, Xie et al. 2010).

GP regression models were initially used in geostatistics to predict the amount of gold in underground deposits (Krige 1951). For these applications, if we were interested in predicting the amount of gold underneath a region, knowing the amount of gold underneath the boundary of the region would not be sufficient information for our prediction. For example, if we knew there was a lot of gold near the region, but none necessarily underneath its boundary, we would still expect gold to be underneath the region. However, we are mainly concerned with response surfaces in operations research, which are different than response surfaces in geostatistics. In operations research, if we are interested in predicting the value of the response surface in a region, then given sufficient information about the response surface on the boundary, information about the response surface outside of the region often would not assist in our predictions. By sufficient information, we mean the level of the response surface and perhaps some derivatives. For GRFs, this property is analogous to the Markov property: given sufficient information (level and derivatives) about the GRF on the boundary, the GRF on the inside of a region is independent of the GRF on the outside. *A contribution of this paper is establishing the Markov property as an important property for GRFs to have for GP regression in operations research, as well as providing a novel definition of the Markov property for real-valued, differentiable GRFs.*

The ability to control the differentiability of the GRF is a characteristic that has received considerable attention in the literature (Santner et al. 2010). A common class of GRFs that are used for metamodeling corresponds to the power exponential covariance function (Santner et al. 2010), for which the differentiability is controlled by a single parameter. However, these GRFs can only be nondifferentiable or infinitely differentiable, depending on the value of the parameter. Another class of covariance functions is the radial basis function form of the class of Matérn covariance functions (Santner et al. 2010), which also has a single parameter that controls the differentiability of the GRF. In contrast to the power exponential covariance function, the GRFs corresponding to the radial basis function form of the Matérn class can have differentiability of any order, although the differentiability cannot differ in each coordinate. Another form of the Matérn class, the product form, can have differentiability that differs in each coordinate. However, the radial basis function form of the Matérn class is more often used and studied in the literature. See the beginning of Section 3 for further discussion.

Mean reversion can be an undesirable characteristic of a metamodel that arises from using a mean-reverting GRF. Mean reversion results because the covariances among the design points (the scenarios at which we run the simulation) and prediction points (the scenarios at which we want to make a prediction) gets weaker as the distance between them becomes greater. Thus, the prediction reverts to the prior mean of the GRF for prediction points that are sufficiently far from design points. Mean reversion can be undesirable when the rate of reversion is too fast, causing predictions to revert to the prior mean of the GRF even for prediction points not too far from any design points. Any GRF in which the covariance between two points monotonically decreases to zero as the distance between the points increases is mean-reverting. Owing to the poor predictions that can result from mean reversion, many methods have been proposed to reduce it in GP regression (see, for example, Li and Sudjianto 2005, Joseph 2006, and Joseph et al. 2008). Furthermore, for these covariance functions, extrapolation causes severe mean reversion because the design points will not contain the prediction point. As in the procedure in Liu and Staum (2010), it can sometimes be very expensive to avoid extrapolation, especially in high dimension, because an extremely large number of design points would be needed to cover a high-dimensional design space, the space of all possible design points. Thus, we would prefer to use a GRF that has no mean reversion. *Another contribution of this paper is providing a novel definition of mean reversion, allowing us to discuss it in a rigorous setting.*

The central contribution of this paper is the introduction of a novel class of GRFs, called generalized integrated Brownian fields (GIBFs), which have all of the desired properties mentioned above. As will be shown in Section 6, these GRFs lead to better predictions and avoid mean reversion simply by changing the covariance function used with the desired GP regression method. We consider the use of GIBFs for GP regression in deterministic and stochastic simulation metamodeling. The two ways to construct GIBFs are using a probabilistic approach and the theory of reproducing kernels. In the latter, the covariance functions of GIBFs can be constructed using a novel parametrization of the reproducing kernels corresponding to tensor products of the appropriate Sobolev spaces (Berlinet and Thomas-Agnan 2004). By placing the construction of GIBFs in the probabilistic setting, we can discuss their properties, such as differentiability that can vary in each coordinate, no mean reversion, and the Markov property. We also show how to implement stochastic kriging with GIBFs, and we use several examples to compare the prediction ability of GIBFs with the GRFs corresponding to the Gaussian and Matérn covariance functions. In the following, we call the GRFs corresponding to the Gaussian and Matérn covariance functions simply the Gaussian and Matérn GRFs, respectively.

GP regression with GIBFs is a generalization of using smoothing splines with integrated Brownian motion in one dimension (Berlinet and Thomas-Agnan 2004).

Berlinet and Thomas-Agnan (2004) provide general guidelines for creating smoothing splines in a tensor product reproducing kernel Hilbert space (RKHS). These guidelines assume the user has decomposed the tensor product RKHS into all of its orthogonal subspaces and has chosen which subspaces to penalize and which subspaces to disregard altogether by performing a model selection. Furthermore, the user must specify how each subspace is penalized. Instead of decomposing the tensor product of the appropriate Sobolev spaces into all of its orthogonal subspaces, we use the entire space by parameterizing its reproducing kernel, which also automatically handles how much each subspace is penalized. Thus, the method presented in this paper is much easier to implement; once the trend is chosen, the reproducing kernel (covariance function) follows automatically, and the parameters of the reproducing kernel are chosen from the simulation output using maximum likelihood estimation and cross-validation.

Brownian motion and fractional Brownian field have recently been proposed for GP regression (Sun et al. 2014, Zhang and Apley 2014). Although both processes have no mean reversion and Brownian motion has the Markov property, neither process has controllable differentiability, and they are nondifferentiable almost everywhere. Furthermore, fractional Brownian field does not have the Markov property. To create smoother GRFs based on fractional Brownian field, Zhang and Apley (2016) use the white noise integral representation of fractional Brownian field and replace the white noise process by any stationary GRF. The differentiability of the resulting processes, called BI GRFs, is not clear, and they still do not have the Markov property. Furthermore, the covariance functions of BI GRFs are difficult to compute and need to be numerically approximated. In contrast, the covariance functions of GIBFs are intuitive and simple to compute, with convenient closed-form expressions.

Although the use of GP regression in simulation metamodeling has led to several different metamodeling methods (see, for example, van Beers and Kleijnen 2003, Kleijnen and van Beers 2005, and Yin et al. 2011), we will focus on the simulation metamodeling method called stochastic kriging, which we discuss in Section 2. We then present GIBFs using a probabilistic approach in Section 3, discuss their properties in Section 4, and provide a guide to using these GRFs with stochastic kriging, as well as discuss their approximation capability, in Section 5. We conclude the paper with numerical experiments in Section 6, which show the improved prediction accuracy as compared with the well-known and highly used Gaussian and Matérn GRFs. The e-companion to this paper provides additional numerical experiments and discussion concerning the properties of GIBFs, as well as recommendations for using GIBFs and the proofs of all theorems.

2. Stochastic Kriging

GP regression has been used for approximating the output of deterministic computer experiments following the work of Sacks et al. (1989), which introduced kriging into the design and analysis of deterministic computer experiments. In kriging, the response surface $y(\cdot)$ at $\mathbf{x}$ is modeled as a realization of the random variable

$$Y_{\mathsf{M}}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^{\top}\boldsymbol{\beta} + \mathsf{M}(\mathbf{x}), \qquad (1)$$

where $\mathbf{x}$ is a point in the design space $\mathscr{X}$ (the space of all possible design points), $\mathbf{f}(\cdot)$ is a $p \times 1$ vector of known functions, that is, $\mathbf{f}(\cdot) = (f_1(\cdot), f_2(\cdot), \ldots, f_p(\cdot))^{\top}$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\mathsf{M}(\cdot)$ is a zero mean GRF (Santner et al. 2010). The GRF $\mathsf{M}(\cdot)$ is assumed to exhibit spatial covariance, which is determined by the covariance function $\Sigma_{\mathsf{M}}(\cdot, \cdot; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes a generic vector of parameters, with the number of components implied from the context. Specifically, the covariance between $\mathsf{M}(\cdot)$ at two points $\mathbf{x}$ and $\mathbf{x}'$ in the design space is given by $\mathrm{Cov}[\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{x}')] = \Sigma_{\mathsf{M}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta})$.

For deterministic computer experiments in which the output of the experiment contains no noise, the response surface can be observed exactly at each of the design points at which the computer experiment is run. Kriging results in an interpolation of the data; that is, the metamodel is equal to the computer experiment output at each of the scenarios run, which fits the deterministic nature of the problem.

In the stochastic simulation case, we no longer observe the response surface without noise. Rather, we run the simulation model at k design points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ for a total of n_i replications at design point $\mathbf{x}_i$. Replication j at design point $\mathbf{x}_i$ is denoted by $\mathscr{Y}_j(\mathbf{x}_i)$. At design point $\mathbf{x}_i$ we collect the sample mean $\bar{\mathscr{Y}}(\mathbf{x}_i) = (1/n_i) \cdot \sum_{j=1}^{n_i} \mathscr{Y}_j(\mathbf{x}_i)$ and the sample variance $s^2(\mathbf{x}_i) = (1/(n_i - 1)) \sum_{j=1}^{n_i} (\mathscr{Y}_j(\mathbf{x}_i) - \bar{\mathscr{Y}}(\mathbf{x}_i))^2$. In general, GP regression in stochastic simulation utilizes the sample means and sample variances at the design points to build the GP regression model.

In stochastic kriging (Ankenman et al. 2010), the response surface is modeled as a sample path of the GRF $Y_{\mathsf{M}}(\cdot)$, given by Equation (1), with mean function $\mathbf{f}(\cdot)^{\top}\boldsymbol{\beta}$ and covariance function $\Sigma_{\mathsf{M}}(\cdot, \cdot; \boldsymbol{\theta})$. The simulation output on replication j at design point $\mathbf{x}$ is modeled as a realization of the random variable $Y_{\mathsf{M}}(\mathbf{x}) + \epsilon_j(\mathbf{x})$, where the zero mean sampling noise in the replications $\{\epsilon_j(\mathbf{x})\}_j$ at a design point $\mathbf{x}$ is independent and identically distributed across replications. The sampling noise is referred to as intrinsic uncertainty, because it is inherent in the stochastic simulation. The stochastic nature of M is called extrinsic uncertainty, because it is

imposed on the problem to aid in the development of the metamodel (Ankenman et al. 2010).

Suppose that the simulation model has been run at the k design points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ yielding the vector of observed simulation output $\bar{\mathcal{Y}} = (\bar{\mathcal{Y}}(\mathbf{x}_1), \bar{\mathcal{Y}}(\mathbf{x}_2), \ldots, \bar{\mathcal{Y}}(\mathbf{x}_k))^\top$, and we now want to predict the response surface at $\mathbf{x}_0$. Let $\mathbf{F} = [\mathbf{f}(\mathbf{x}_1) \ \mathbf{f}(\mathbf{x}_2) \ \ldots \ \mathbf{f}(\mathbf{x}_k)]^\top$ be the $k \times p$ regression matrix, let $\Sigma_M(\theta)$ be the $k \times k$ variance–covariance matrix with ijth entry $\Sigma_M(\mathbf{x}_i, \mathbf{x}_j; \theta)$, and let $\Sigma_M(\mathbf{x}_0, \cdot; \theta)$ be the $k \times 1$ vector of spatial covariances between the design points and the prediction point; that is, the ith entry of $\Sigma_M(\mathbf{x}_0, \cdot; \theta)$ is $\Sigma_M(\mathbf{x}_0, \mathbf{x}_i; \theta)$. Additionally, let Σ_ϵ be the $k \times k$ intrinsic covariance matrix with ijth entry $\mathrm{Cov}[\sum_{k=1}^{n_i} \epsilon_k(\mathbf{x}_i)/n_i, \sum_{k=1}^{n_j} \epsilon_k(\mathbf{x}_j)/n_j]$. Assuming that β, θ, and Σ_ϵ are known, the stochastic kriging prediction (Ankenman et al. 2010) at $\mathbf{x}_0$ is given by

$$\widehat{Y}_M(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \beta + \Sigma_M(\mathbf{x}_0, \cdot; \theta)^\top (\Sigma_M(\theta) + \Sigma_\epsilon)^{-1} (\bar{\mathcal{Y}} - \mathbf{F}\beta). \tag{2}$$

The mean-squared error of the prediction $\widehat{Y}_M(\mathbf{x}_0)$ is

$$\mathrm{MSE}(\widehat{Y}_M(\mathbf{x}_0)) = \Sigma_M(\mathbf{x}_0, \mathbf{x}_0; \theta) - \Sigma_M(\mathbf{x}_0, \cdot; \theta)^\top$$
$$\cdot (\Sigma_M(\theta) + \Sigma_\epsilon)^{-1} \Sigma_M(\mathbf{x}_0, \cdot; \theta). \tag{3}$$

In practice, β, θ, and Σ_ϵ are not known and must be estimated from the simulation output. Parameter estimation for GIBFs is discussed in Section 5.2. Although derivative information can be used with the covariance functions introduced in this paper, we do not discuss incorporating derivative information here and refer the reader to Chen et al. (2013) for details on implementation.

3. Generalized Integrated Brownian Fields

In stochastic kriging, the response surface is modeled as a sample path of the GRF $Y_M(\cdot)$, given by Equation (1), with mean function $\mathbf{f}(\cdot)^\top \beta$ and covariance function $\Sigma_M(\cdot, \cdot; \theta)$. The GRFs we construct in this section, called GIBFs, have desirable properties such as differentiability that can differ in each coordinate, no mean reversion, and the Markov property. We want to use GRFs with these properties in an effort to obtain better predictions. Note, in this paper, the continuity, differentiation, and integration of a stochastic process is interpreted in the mean-square sense.

A widely used GRF is the Gaussian GRF for which the covariance between the two points $\mathbf{x}$ and $\mathbf{x}'$ is given by $\Sigma_M(\mathbf{x}, \mathbf{x}'; \theta) = \sigma^2 \exp\{-\sum_i \theta_i (x_i - x_i')^2\}$, where θ_i, x_i, and x_i' are the ith coordinates of θ, $\mathbf{x}$, and $\mathbf{x}'$, respectively, and σ^2 is the variance of the GRF. The Gaussian GRF is mean-reverting and is often criticized as being too smooth because it is infinitely continuously differentiable in every coordinate. Another widely used GRF is the radial basis

function class of the Matérn GRF, for which the covariance between the two points $\mathbf{x}$ and $\mathbf{x}'$ is given by $\Sigma_M(\mathbf{x}, \mathbf{x}'; \theta) = \sigma^2 (2^{1-v}/\Gamma(v))(\sqrt{2v} \|\theta^\top(\mathbf{x} - \mathbf{x}')\|_2)^v K_v(\sqrt{2v} \|\theta^\top (\mathbf{x} - \mathbf{x}')\|_2)$, where $\Gamma(\cdot)$ is the gamma function and $K_v(\cdot)$ is the modified Bessel function of the second kind. The differentiability of the Matérn GRF is controlled by the single parameter v, so its differentiability cannot differ in each coordinate. The product class of the Matérn GRF, given by $\Sigma_M(\mathbf{x}, \mathbf{x}'; \theta) = \sigma^2 \cdot \prod_{i=1}^d (2^{1-v_i}/\Gamma(v_i))(\sqrt{2v_i}|\theta_i^\top(x_i - x_i')|)^{v_i} K_v(\sqrt{2v_i}|\theta_i(x_i - x_i')|)$, has a separate parameter, v_i, controlling the differentiability in each coordinate. Both classes of the Matérn GRF are mean-reverting, and the product class of the Matérn GRF is not often used or studied in the literature. Furthermore, for the product class of the Matérn GRF, each v_i is continuous, and there is not a one-to-one correspondence between values of v_i and the resulting differentiability of the GRF in that coordinate. For example, an uncountably infinite number of values of v_i correspond to a GRF that is once differentiable in the ith coordinate. For GIBFs, there is a one-to-one correspondence between the values of the parameters controlling the differentiability of the GRF and its differentiability. This lends itself to a simple search algorithm to find the optimal values.

GIBFs are generalized versions of integrated Brownian fields (Fill and Torcaso 2004), which are multivariate versions of integrated Brownian motions. We first construct one-dimensional GIBFs, which we call generalized integrated Brownian motions (GIBMs), and then construct multidimensional GIBFs. In the following, let $\wedge$ and $\vee$ denote the functions min and max, respectively.

3.1. One-Dimensional Generalized Integrated Brownian Motions

Consider one-dimensional Brownian motion $B(\cdot; \theta)$ on $\mathbb{R}_{\geq 0}$ with volatility θ, where $\mathbb{R}_{\geq 0} \triangleq [0, \infty)$. This process is a real-valued, continuous, and nondifferentiable zero mean Gaussian stochastic process. The covariance between $B(\cdot; \theta)$ at $x, x' \in \mathbb{R}_{\geq 0}$ is given by $\Sigma_B(x, x'; \theta) = \theta(x \wedge x')$. An m-times differentiable stochastic process can be obtained by integrating $B(\cdot; \theta)$ for m times, which gives us the well-known m-integrated Brownian motion $B_m(\cdot; \theta)$ with volatility θ. The integral representation of m-integrated Brownian motion with volatility θ at x is

$$B_m(x; \theta) = \int_0^x B_{m-1}(u; \theta) du = \int_0^x \frac{(x - u)^m}{m!} dB(u; \theta), \tag{4}$$

where the first equality expresses $B_m(x; \theta)$ recursively with $B_0(\cdot; \theta) = B(\cdot; \theta)$, and the second equality follows from integration by parts, which expresses $B_m(x; \theta)$ as an integral with respect to Brownian motion with volatility θ. From the first integral in Equation (4), it

is clear that the process $B_m(\cdot; \theta)$ and its m derivatives $B_m^{(i)}(\cdot; \theta)$, for $i = 1, 2, \ldots, m$, are zero at the boundary $x = 0$. These boundary conditions make $B_m(\cdot; \theta)$ unsuitable for metamodeling, because the response surface and its derivatives may not be zero at the boundary $x = 0$. We modify $B_m(\cdot; \theta)$ by adding a random polynomial whose coefficients are $m + 1$ independent standard normal random variables, $Z_0, Z_1, \ldots, Z_m$, scaled by some parameters, creating the novel stochastic process we call generalized m-integrated Brownian motion (m-GIBM). This process, denoted by $G_m(\cdot; \theta)$, is defined by

$$G_m(x; \theta) \triangleq \sum_{n=0}^{m} \frac{\sqrt{\theta_n} Z_n}{n!} x^n + B_m(x; \theta_{m+1}), \quad (5)$$

where θ has been relabeled as θ_{m+1} for convenience, $\theta = (\theta_0, \theta_1, \ldots, \theta_{m+1})^\top$, and $B_m(\cdot; \theta_{m+1})$ is independent of Z_n for all $n = 1, 2, \ldots, m$. For the Gaussian and Matérn covariance functions, the number of components of θ is equal to the dimension of $\mathbf{x}$ (i.e., d). For GIBMs, the number of components of θ is equal to the number of integrations plus two (i.e., $m + 2$). The first $m + 1$ parameters, $\theta_0, \theta_1, \ldots, \theta_m$, are the coefficients of the random polynomial, whereas the last parameter, θ_{m+1}, always corresponds to the volatility of the Brownian motion. Note that adding the random polynomial to $B_m(\cdot; \theta_{m+1})$ does not change its mean at $x = 0$. Figure 1 shows sample paths of GIBMs of different orders, with $\theta = (1, 1, \ldots, 1)^\top$, on the unit interval. Directly from the definitions of $B_m(\cdot; \theta)$ and $G_m(\cdot; \theta)$, it follows that the covariance between $G_m(\cdot; \theta)$ at $x, x' \in \mathbb{R}_{\geq 0}$ is given by

$$\Sigma_{G_m}(x, x'; \theta) = \sum_{n=0}^{m} \theta_n \frac{x^n (x')^n}{(n!)^2}$$
$$+ \theta_{m+1} \int_0^\infty \frac{(x - u)_+^m (x' - u)_+^m}{(m!)^2} du.$$

For any m, the integral can be easily computed and has a convenient closed-form solution, composed of terms that are products of the functions min and max. Indeed, we have

$$\Sigma_{G_m}(x, x'; \theta) = \sum_{n=0}^{m} \theta_n \frac{x^n (x')^n}{(n!)^2} + \theta_{m+1} \sum_{i=m+1}^{2m+1} (-1)^{i-m+1}$$
$$\cdot \frac{(x \wedge x')^i (x \vee x')^{2m+1-i}}{i! (2m+1-i)!}. \quad (6)$$

For $m = 0$ and $m = 1$, one can easily check that the last sum in Equation (6) reduces to the covariance functions of the well-known Brownian motion and integrated Brownian motion, respectively.

3.2. Multidimensional Generalized Integrated Brownian Fields

In the multidimensional case, consider d-dimensional Brownian field $\mathbf{B}(\cdot; \theta)$ on $\mathbb{R}_{\geq 0}^d$ with volatility θ (Holden et al. 2010), where $\theta = (\theta_1, \theta_2, \ldots, \theta_d)^\top$ is a vector of parameters. Here $\mathbf{B}(\cdot; \theta)$ is the tensor product of d independent copies of one-dimensional Brownian motions with varying volatilities. This field is a real-valued, continuous, and nondifferentiable zero mean GRF. The covariance between $\mathbf{B}(\cdot; \theta)$ at $\mathbf{x}, \mathbf{x}' \in \mathbb{R}_{\geq 0}^d$ is given by $\Sigma_{\mathbf{B}}(\mathbf{x}, \mathbf{x}'; \theta) = \prod_{i=1}^{d} \theta_i (x_i \wedge x_i')$. Similar to the one-dimensional case, we can integrate Brownian field with volatility θ over each coordinate to get a differentiable process. In the multidimensional case, each coordinate can be integrated a different number of times. If we integrate m_i times in the ith coordinate for $i = 1, 2, \ldots, d$, the resulting GRF is the well-known $\mathbf{m}$-integrated Brownian field $\mathbf{B_m}(\cdot; \theta)$ with volatility θ (Fill and Torcaso 2004), where $\mathbf{m} = (m_1, m_2 \ldots, m_d)^\top$. Using integration by parts, $\mathbf{B_m}(\mathbf{x}; \theta)$ can be expressed as a multiple integral with respect to Brownian field with volatility θ,

$$\mathbf{B_m}(\mathbf{x}; \theta) \triangleq \int_0^{x_1} \cdots \int_0^{x_d} \prod_{i=1}^{d} \frac{(x_i - u_i)^{m_i}}{m_i!} d\mathbf{B}(\mathbf{u}; \theta).$$

Figure 1. Sample Paths of GIBMs on the Unit Interval

0-GIBM   1-GIBM   2-GIBM
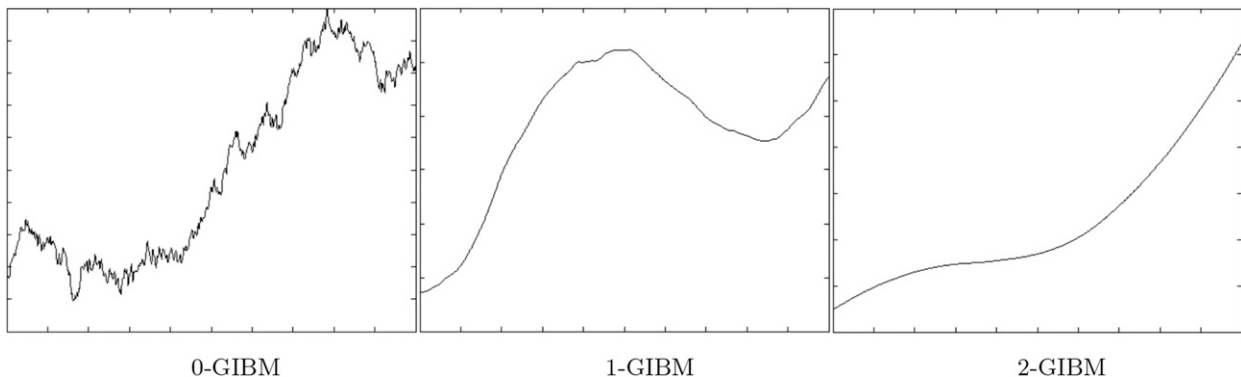
It follows immediately (Fill and Torcaso 2004) from this representation that the covariance between $\mathbf{B_m}(\cdot; \boldsymbol{\theta})$ at $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbb{R}_{\geq 0}^d$ is given by

$$\Sigma_{\mathbf{B_m}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}) = \prod_{i=1}^d \theta_i \int_0^\infty \frac{(x_i - u_i)_+^{m_i}(x_i' - u_i)_+^{m_i}}{(m_i!)^2} du_i.$$

The covariance function $\Sigma_{\mathbf{B_m}}(\cdot, \cdot; \boldsymbol{\theta})$ is the product of the covariance functions of the one-dimensional integrated Brownian motions $B_{m_1}(\cdot; \theta_1), B_{m_2}(\cdot; \theta_2), \ldots, B_{m_d}(\cdot; \theta_d)$. Similar to the one-dimensional case, $\mathbf{m}$-integrated Brownian field with volatility $\boldsymbol{\theta}$ has boundary conditions $\mathbf{B_m}(\mathbf{x}; \boldsymbol{\theta}) = 0$ and $\mathbf{B_m}^{(\alpha)}(\mathbf{x}; \boldsymbol{\theta}) = 0$ for all $\boldsymbol{\alpha} \leq \mathbf{m}$ (i.e., all $\boldsymbol{\alpha}$ such that $\alpha_i \leq m_i$ for all i, and $\mathbf{x} \in \mathbb{R}_{\geq 0}^d$ such that $x_i = 0$ for some i). We define a novel process whose covariance function is the product of covariance functions of d GIBMs, in the same way that the covariance function of $\mathbf{B_m}(\cdot; \boldsymbol{\theta})$ is the product of the covariance functions of $B_{m_1}(\cdot; \theta_1), B_{m_2}(\cdot; \theta_2), \ldots, B_{m_d}(\cdot; \theta_d)$. This novel process, which we call generalized $\mathbf{m}$-integrated Brownian field, is defined next:

Definition 1. The zero mean GRF $\mathbf{G_m}(\cdot; \boldsymbol{\theta})$ on $\mathbb{R}_{\geq 0}^d$ whose covariance at $\mathbf{x}$ and $\mathbf{x}'$ is given by

$$\Sigma_{\mathbf{G_m}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}) = \prod_{i=1}^d \left(\sum_{n=0}^{m_i} \theta_{i,n} \frac{x_i^n (x_i')^n}{(n!)^2} + \theta_{i,m_i+1} \right.$$
$$\left. \cdot \int_0^\infty \frac{(x_i - u_i)_+^{m_i}(x_i' - u_i)_+^{m_i}}{(m_i!)^2} du_i \right), \quad (7)$$

where $\boldsymbol{\theta} = (\theta_{1,0}, \ldots, \theta_{1,m_1+1}, \theta_{2,0}, \ldots, \theta_{d,m_d+1})^\top \in \mathbb{R}_{>0}^M$ is a vector of parameters and $M = \sum_{i=1}^d (m_i + 2)$ is called generalized $\mathbf{m}$-integrated Brownian field ($\mathbf{m}$-GIBF).

Similar to the one-dimensional case, the number of components of $\boldsymbol{\theta}$ corresponding to each coordinate i is equal to the number of integrations with respect to that coordinate plus two (i.e., $m_i + 2$). Thus, the total number of components of $\boldsymbol{\theta}$ is $\sum_{i=1}^d (m_i + 2)$. We restrict $\boldsymbol{\theta}$ to $\mathbb{R}_{>0}^M$ instead of $\mathbb{R}_{\geq 0}^M$ to avoid improper GRFs (i.e., GRFs with positive-semidefinite covariance matrices). Figure 2 shows sample paths of GIBFs of different orders, with $\boldsymbol{\theta} = (1, 1, \ldots, 1)^\top$, on the unit square. Equivalently, we can define $\mathbf{m}$-GIBF as

$$\mathbf{G_m}(\mathbf{x}; \boldsymbol{\theta}) \triangleq \sum_{\mathbf{n}=0}^{\mathbf{m}} C_{\mathbf{n}}(\boldsymbol{\theta}) \mathbf{x}^{\mathbf{n}} Z_{\mathbf{n}} + \sum_{i=1}^d \sum_{j_1 < \cdots < j_i} C_{j_1, \ldots, j_i}(\mathbf{x}, \boldsymbol{\theta})$$
$$\cdot \mathbf{B}_{(m_{j_1}, \ldots, m_{j_i})}^{j_1, \ldots, j_i}(x_{j_1}, \ldots, x_{j_i}; \mathbf{1}), \quad (8)$$

where the multidimensional sum is over all $\mathbf{n} = (n_1, n_2, \ldots, n_d)$ such that $\mathbf{0} \leq \mathbf{n} \leq \mathbf{m}$. Equation (8) is the multidimensional analog of Equation (5). The first term in Equation (8) is a random polynomial of degree $\mathbf{m}$, which is the linear combination of standard normal random variables with coefficients that are monomials of degree at most $\mathbf{m}$. The second term is the sum of integrated Brownian fields over every i-face of $\mathbb{R}_{\geq 0}^d$, for $i = 1, 2, \ldots, d$. In other words, we sum integrated Brownian fields over each edge, face, cell, 4-face, 5-face, etc. of $\mathbb{R}_{\geq 0}^d$. The functions $C_{\mathbf{n}}(\cdot)$ and $C_{j_1, \ldots, j_i}(\cdot, \cdot)$ are deterministic functions of $\mathbf{x}$ and $\boldsymbol{\theta}$, and although closed-form expressions can be obtained for each, they are not needed for implementation and do not add any insight into the process. Because the functions $C_{\mathbf{n}}(\cdot)$ and $C_{j_1, \ldots, j_i}(\cdot, \cdot)$ are deterministic functions of $\mathbf{x}$ and $\boldsymbol{\theta}$, the randomness in $\mathbf{G_m}(\cdot; \boldsymbol{\theta})$ is due to the standard normal random variables $Z_{\mathbf{n}}$ and the integrated Brownian

Figure 2. Sample Paths of GIBFs on the Unit Square

(0,0)-GIBF

(1,1)-GIBF

(1,2)-GIBF

(2,2)-GIBF
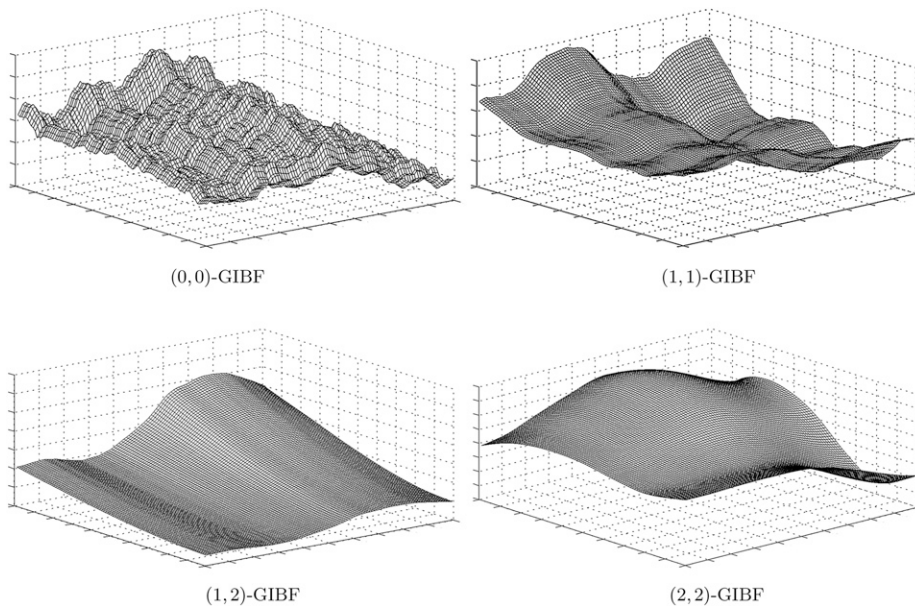
fields $\mathbf{B}_{(m_{j_1},\ldots,m_{j_i})}^{j_1,\ldots,j_i}(\cdot;\mathbf{1})$, which are all independent from each other. From the formulation of $\mathbf{G_m}(\cdot;\boldsymbol{\theta})$ given by Equation (8), it is clear that $\mathbf{m}$-GIBF does not have any boundary conditions. Furthermore, one can easily check that this representation is equivalent to Definition 1 by multiplying out the terms in the product in the covariance function given by Equation (7) and comparing it with the covariance function of the GRF given by the right-hand side of Equation (8). Because a GRF is completely determined by its mean and covariance functions, and the sum of Gaussian random variables is Gaussian, the equivalence between Definition 1 and Equation (8) follows.

4. Properties
In GP regression, using methods such as kriging and stochastic kriging, the response surface is modeled as a sample path of a GRF. When $\mathbf{m}$-GIBF is used as the GRF, the response surface is modeled as a sample path of $\mathbf{m}$-GIBF. To obtain the best predictions possible, we would like the GRF that we use to have properties that make it suitable for metamodeling. Three desirable properties that GIBFs have are differentiability that can differ in each coordinate, no mean reversion, and the Markov property. In this section, we discuss these properties, as well as another property that has implications for metamodeling, namely, the nonstationarity of GIBFs.

4.1. Differentiability
For one-dimensional m-GIBM, the differentiability can easily be seen from the representation of m-GIBM, given by Equation (5). The random polynomial given by the first term on the right-hand side of Equation (5) is the linear combination of standard normal random variables whose coefficients are monomials of degree at most m, and is thus infinitely differentiable. The second term on the right-hand side of Equation (5) is an m-integrated Brownian motion. By definition, m-integrated Brownian motion is m times differentiable. Specifically, the derivative of m-integrated Brownian motion is $(m-1)$-integrated Brownian motion. Thus, one-dimensional m-GIBM is m times differentiable. Note, the derivative of m-GIBM, where $m \geq 1$, is $(m-1)$-GIBM. Indeed, the second-order mixed derivative of the covariance function (Chen et al. 2013) of m-GIBM exists and is continuous and is the covariance function of $(m-1)$-GIBM.

The equivalent definition of $\mathbf{m}$-GIBF, given by Equation (8), can be used to determine the differentiability of $\mathbf{m}$-GIBF. As with one-dimensional m-GIBM, the random polynomial given by the first term on the right-hand side of Equation (8) is the linear combination of standard normal random variables whose coefficients are monomials of degree at most $\mathbf{m}$, and is thus infinitely differentiable with respect to every

coordinate. The functions $C_{j_1,\ldots,j_i}(\cdot,\boldsymbol{\theta})$ are also linear combinations of monomials and are, thus, also infinitely differentiable with respect to every coordinate. The only other terms in Equation (8) that depend on $\mathbf{x}$ are the integrated Brownian fields. A $(p_1,p_2,\ldots,p_h)$-integrated Brownian field, for any $h \geq 1$, is p_i times differentiable in the ith coordinate. Specifically, the derivative of $(p_1,p_2,\ldots,p_h)$-integrated Brownian field with respect to the ith coordinate is $(p_1,p_2,\ldots,p_i-1,\ldots,p_h)$-integrated Brownian field. Thus, the ith coordinate of $\mathbf{m}$-GIBF is m_i times differentiable. Furthermore, we are able to control the differentiability of $\mathbf{m}$-GIBF in each coordinate by specifying each entry of the vector $\mathbf{m} = (m_1,m_2,\ldots,m_d)$. Note, the derivative of $\mathbf{m}$-GIBF, where $m_i \geq 1$, with respect to the ith coordinate is $(m_1,m_2,\ldots,m_i-1,\ldots,m_d)$-GIBF. Indeed, the second-order mixed derivative of the covariance function of $\mathbf{m}$-GIBF with respect to the ith coordinate exists and is continuous and is the covariance function of $(m_1,m_2,\ldots,m_i-1,\ldots,m_d)$-GIBF.

4.2. No Mean Reversion
As mentioned in the introduction, GP regression models can exhibit mean reversion if the GRF used is mean-reverting. In this paper we assume that the application of GP regression is such that mean reversion should be avoided, because some algorithms can use mean reversion to their benefit. For example, the practical performance of the correlated knowledge gradient (cKG) algorithm (Powell and Ryzhov 2012) can sometimes be improved by purposefully setting the unconditional mean of the GRF to a small value (for minimization problems). If mean reversion is present in the fitted GP regression model, then the cKG algorithm will tend to over-sample. This is because we are not confident about the current optimal solution, because several other feasible solutions also have a small conditional mean (owing to mean reversion). Thus, more exploration is encouraged, which can help prevent early stopping. In Section 6, we focus on comparing to the Gaussian and Matérn GRFs when mean reversion is not present, because we assume mean reversion should be avoided. For experiments comparing to the Gaussian and Matérn GRFs when mean reversion is present, see the e-companion.

The concept of mean reversion is well defined for stochastic processes that are parameterized on the time domain: the process is mean-reverting if it tends to drift toward its long-term mean over time. A well-known example of a mean-reverting stochastic process is the Ornstein-Uhlenbeck process. In contrast, the concept of mean reversion has not been formally defined in terms of random fields parameterized on a multidimensional spatial domain. In this case, we no longer have a concept of time. We introduce the following definition of

mean reversion for random fields parameterized on a multidimensional spatial domain:

Definition 2. Let $\mathsf{M}(\cdot)$ be a random field with mean function $m(\cdot)$ defined on a convex cone $\mathscr{C}$. The random field $\mathsf{M}(\cdot)$ is mean-reverting if

$$\mathbb{E}\left[\mathsf{M}(\lambda\mathbf{x}) - m(\lambda\mathbf{x})|\mathsf{M}(\mathbf{x}_1), \mathsf{M}(\mathbf{x}_2), \ldots, \mathsf{M}(\mathbf{x}_k)\right] \xrightarrow{p} 0 \quad (9)$$

as $\lambda \to \infty$, for any $k \geq 1$ points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathscr{C}$ and any $\mathbf{x} \in \mathscr{C}$ not equal to $\mathbf{0}$.

Essentially, we can think of the points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ as being the design points at which we are able to observe the value of the random field $\mathsf{M}(\cdot)$. Then, $\mathbb{E}[\mathsf{M}(\lambda\mathbf{x})|\mathsf{M}(\mathbf{x}_1), \mathsf{M}(\mathbf{x}_2), \ldots, \mathsf{M}(\mathbf{x}_k)]$ is the kriging predictor at the point $\lambda\mathbf{x}$, based on the observations at the design points. The difference $\mathbb{E}[\mathsf{M}(\lambda\mathbf{x})|\mathsf{M}(\mathbf{x}_1), \mathsf{M}(\mathbf{x}_2), \ldots, \mathsf{M}(\mathbf{x}_k)] - m(\lambda\mathbf{x})$ is the difference between the kriging predictor and the unconditional mean. Thus, a random field is mean-reverting if the difference between the kriging predictor and the unconditional mean converges in probability to zero as we move away from the design points. The next theorem shows that this definition of a mean-reverting random field is consistent with the behavior of the Gaussian and Matérn GRFs.

Theorem 1. *Let $\mathsf{M}(\cdot)$ be a GRF defined on $\mathbb{R}^d_{\geq 0}$ with mean function $m(\cdot)$ and covariance function $\mathrm{Cov}[\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{y})] = \tau^2 r(\mathbf{x} - \mathbf{y}; \boldsymbol{\theta})$, for some scalar τ and function $r(\cdot; \boldsymbol{\theta})$ such that $r(\mathbf{x}; \boldsymbol{\theta}) \to 0$ as $\|\mathbf{x}\| \to \infty$ and $r(\mathbf{0}; \boldsymbol{\theta}) = 1$. Then $\mathsf{M}(\cdot)$ is mean-reverting.*

The main property of the covariance function $\tau^2 r(\mathbf{x} - \mathbf{x}'; \boldsymbol{\theta})$ on which the proof of Theorem 1 relies is that $r(\mathbf{x}; \boldsymbol{\theta})$ decays to zero as $\|\mathbf{x}\| \to \infty$. As we will see in the next theorem, GIBFs do not exhibit mean reversion because their covariance functions do not have this property.

Theorem 2. The GRF $\mathbf{G_m}(\cdot; \boldsymbol{\theta})$ defined on $\mathbb{R}^d_{\geq 0}$ is not mean-reverting.

This theorem ensures us that when we use GIBFs as the GRFs for GP regression, mean reversion will not be present in the resulting metamodel. This is a key property of GIBFs and a significant reason why we prefer to use GIBFs for stochastic kriging.

4.3. Markov Property

Definitions of the Markov property for random fields have been proposed in the stochastic analysis literature, such as those found in Pitt (1971). However, these definitions are too abstract for our purposes; we only need a notion of independence across a "simple" boundary, given sufficient information about the random field on the boundary, and do not require the Markov property to hold for exotic sets. Furthermore, because we are dealing with differentiable random

fields, we would like a definition that explicitly makes use of the random fields differentiability. We introduce the following definition of a real-valued Markov random field, where $\sigma(\mathscr{A})$ denotes the sigma-algebra generated by the set $\mathscr{A}$.

Definition 3. A real-valued random field $\mathsf{M}(\cdot)$ on $\mathbb{R}^d_{\geq 0}$ is called $\mathbf{p}$-order Markov if $\sigma(\mathsf{M}(\mathbf{x}))$ is independent of $\sigma(\{\mathsf{M}(\mathbf{s}): \forall \mathbf{s} \in \times_{i=1}^d [0, t_i)\})$ given $\sigma(\{\mathsf{M}^{(\boldsymbol{\alpha})}(\mathbf{s}): \forall \mathbf{s} \in \partial(\times_{i=1}^d [0, t_i]), \forall \boldsymbol{\alpha} \leq \mathbf{p}\})$, for all $\mathbf{x} \in (\times_{i=1}^d [0, t_i])^c$, where $0 < t_i < \infty$ for all i and $\partial(\mathbb{O})$ is the boundary of $\mathbb{O}$.

It is easy to see that our definition reduces to the standard definition of one-dimensional Markov processes with no derivative information (i.e., $p = 0$). Essentially, a random field is $\mathbf{p}$-order Markov if information about the random field (level and derivatives) on the boundary of a d-dimensional rectangle is sufficient for predicting the values that the random field takes at points outside of the d-dimensional rectangle (i.e., our predictions do not change if we know the level and derivatives of the random field inside). Similarly, information about the random field (level and derivatives) on the boundary of a d-dimensional rectangle is sufficient for predicting the values that the random field takes at points inside of the d-dimensional rectangle. We have the following result:

Theorem 3. *The GRF $\mathbf{G_m}(\cdot; \boldsymbol{\theta})$ defined on $\mathbb{R}^d_{\geq 0}$ is $\mathbf{m}$-order Markov.*

This theorem shows that GIBFs indeed have the Markov property, as defined in Definition 3. This is another key property of GIBFs, as well as another significant reason why we prefer to use GIBFs for stochastic kriging.

4.4. Nonstationarity

Multidimensional $\mathbf{m}$-GIBF at $\mathbf{x} \in \mathbb{R}^d_{\geq 0}$ is a Gaussian random variable with mean zero and variance given by

$$\sigma^2_{\mathbf{G_m}}(\mathbf{x}; \boldsymbol{\theta}) = \prod_{i=1}^d \left(\sum_{k=0}^{m_i} \theta_k \frac{x^{2k}}{(k!)^2} + \theta_{m_i+1} \int_0^\infty \frac{(x-u)_+^{2m_i}}{(m_i!)^2} du \right).$$

Because the distribution of $\mathbf{m}$-GIBF depends on $\mathbf{x}$, it follows that $\mathbf{m}$-GIBF is not stationary. Indeed, $\sigma^2_{\mathbf{G_m}}(\mathbf{x}; \boldsymbol{\theta})$ is an increasing function of $\mathbf{x}$, so the variance increases as we move away from the origin. This is in contrast to the stationary Gaussian and Matérn GRFs. However, the conditional distribution of any GRF, stationary or not, is always nonstationary; as we move away from the design points, the conditional variance increases. The key difference between the conditional variances of GIBFs and the conditional variances of stationary GRFs is that the conditional variances of GIBFs can be asymmetric, given symmetrically placed design points.

In the case of 1-GIBM with $\boldsymbol{\theta} = \mathbf{1}$ on the unit interval, the conditional variance at x, given the value of the process at two design points, for two different sets of

design points is plotted in Figure 3. From this plot we can see that the conditional variance is smaller near $x = 0$ and larger near $x = 1$, even though the design points have been placed symmetrically with respect to the center of the design space. Note that the conditional variance of GIBM is positive at the origin, even though the prior and conditional variances of integrated Brownian motion at the origin are zero. This is because the random polynomial added to integrated Brownian motion to form GIBM contributes to the prior (and conditional) variance at the origin. Thus, we do not need to be concerned with having an overconfident prediction at the origin, as would be the case if we were using integrated Brownian motion.

Although the conditional variances of GIBFs can be asymmetric, given symmetrically placed design points, the conditional variances still possess the property we desire for metamodeling: namely, the conditional variance increases as we move away from the design points. Consider kriging with Brownian motion: although Brownian motion is nonstationary (as we move away from the origin, the variance of Brownian motion increases), when we condition on the simulation output at two design points, the resulting process between the design points is a Brownian bridge. The variance of the Brownian bridge will be largest in the center of the design points and decrease as we get closer to either design point. In the e-companion, we show that the conditional variances of GIBFs provide highly-desirable inference and can be used in methods utilizing the conditional distribution (e.g., expected improvement methods in global optimization [Jones et al. 1998]) because they possess this property.

5. Stochastic Kriging with Generalized Integrated Brownian Fields

For stochastic kriging with $\mathbf{m}$-GIBF, the response surface $y(\cdot)$ at $\mathbf{x}$ is modeled as a realization of the random variable $Y_{\mathbf{G}_{\mathbf{m}}}(\mathbf{x}; \theta) = \mathbf{f}(\mathbf{x})^\top \beta + \widetilde{\mathbf{G}}_{\mathbf{m}}(\mathbf{x}; \theta)$, where $\mathbf{f}(\cdot)$ and β are as before, and $\widetilde{\mathbf{G}}_{\mathbf{m}}(\cdot; \theta)$ is a modified version of $\mathbf{m}$-GIBF, discussed in Section 5.1, which accounts for the basis functions in $\mathbf{f}(\cdot)$. To implement stochastic kriging with GIBFs, we need to choose the vector of basis functions $\mathbf{f}(\cdot)$ to be used for trend modeling and values for the parameters $\mathbf{m}$, β, and θ. This section discusses these aspects of fitting GIBFs, including trend modeling in Section 5.1, followed by parameter estimation in Section 5.2, assuming that the vector of basis functions has been fixed. The approximation capability and the resulting differentiability of metamodels built using GIBFs are discussed in Sections 5.3 and 5.4, respectively.

5.1. Trend Modeling

To maintain the differentiability of the metamodel (discussed in Section 5.4 below), we assume that each basis function in the $p \times 1$ vector of basis functions $\mathbf{f}(\cdot)$ is m_i times continuously differentiable in the ith coordinate. Any function can be a basis function as long as it satisfies this differentiability condition. If any prior information about the response surface is known, the trend model should be chosen to reflect it. However, if no prior information about the response surface is known, we recommend using no trend model.

For certain basis functions, the covariance function needs to be modified. For stochastic kriging with $\mathbf{m}$-GIBF, when a basis function is a monomial $\mathbf{x}^\alpha$, where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)^\top$ and $\alpha_i \le m_i$ for $i = 1, 2, \ldots, d$, we need to subtract $\prod_{i=1}^d \theta_{i,\alpha_i} x_i^{\alpha_i} y_i^{\alpha_i}/(\alpha_i!)^2$ from the covariance function given by Equation (7). The need for this modification of the covariance function is the following. For stochastic kriging with the GRF $Y_{\mathsf{M}}(\cdot)$, given by Equation (1), the difference $y(\cdot) - \mathbf{f}(\cdot)\beta$ is modeled as a sample path of the zero mean GRF $\mathsf{M}(\cdot)$. When $\mathbf{x}^\alpha$ is included in $\mathbf{f}(\cdot)$, the variability of the simulation output $\widetilde{\mathcal{Y}}$ associated with the subspace spanned by $\mathbf{x}^\alpha$ is eliminated by taking the difference $\widetilde{\mathcal{Y}} - \mathbf{F}\widehat{\beta}$, where $\widehat{\beta}$ is the

Figure 3. Plots of the Conditional Variance of $G_1(x; \mathbf{1})$ Given the Value of the Process at the Design Points
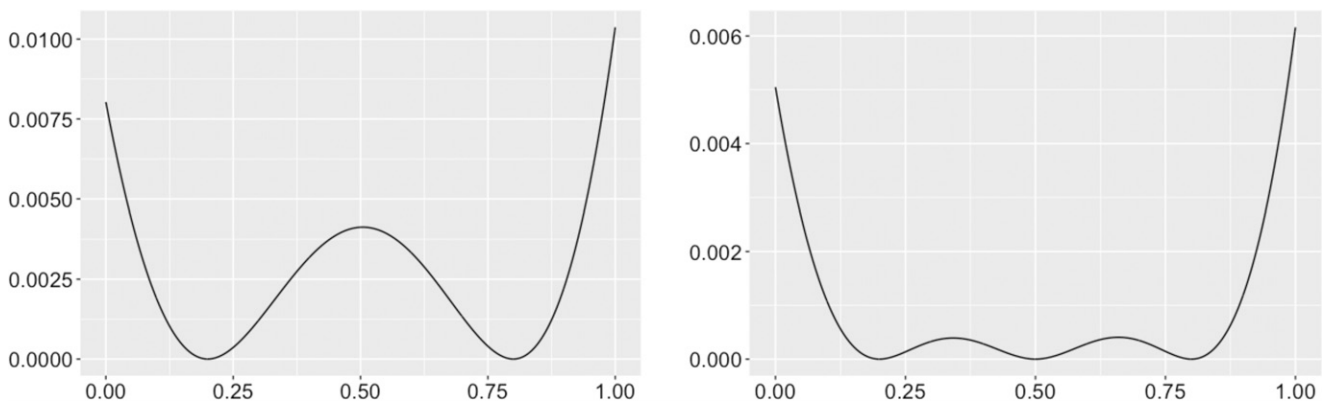
maximum likelihood estimator (MLE) of β (see Section 5.2). To avoid redundancy when we use an **m**-GIBF as the zero mean GRF $\mathsf{M}(\cdot)$, we remove the term $C_\alpha(\theta)\mathbf{x}^\alpha Z_\alpha$ in the random polynomial in Equation (8). This term is the GRF whose covariance at $\mathbf{x}, \mathbf{y} \in [0,1]^d$ is given by $\prod_{i=1}^d \theta_{i,\alpha_i} x_i^{\alpha_i} y_i^{\alpha_i}/(\alpha_i!)^2$.

Another explanation for the modification of the covariance function can be given in terms of boundary conditions. The formulation of **m**-GIBF given by Equation (8) is the sum of a term involving an **m**-integrated Brownian field and other terms that compensate for its boundary conditions. When the basis function $\mathbf{x}^\alpha$ is included in $\mathbf{f}(\cdot)$, we do not need the term $C_\alpha(\theta)\mathbf{x}^\alpha Z_\alpha$ in the random polynomial $\sum_{\mathbf{n}=0}^{\mathbf{m}} C_\mathbf{n}(\theta)\mathbf{x}^\mathbf{n} Z_\mathbf{n}$, because the corresponding boundary condition is compensated for by the term involving $\mathbf{x}^\alpha$ in the trend function. For example, consider the GRF $\mathsf{Y}_{\mathsf{G_m}}(\cdot;\theta)$ with $\mathbf{f}(\cdot) = (1)^\top$, whose value at $\mathbf{x}$ is given by $\mathsf{Y}_{\mathsf{G_m}}(\mathbf{x};\theta) = \beta_0 + \widetilde{\mathsf{G}}_\mathbf{m}(\mathbf{x};\theta)$, where $\widetilde{\mathsf{G}}_\mathbf{m}(\mathbf{x};\theta) = \mathsf{G}_\mathbf{m}(\mathbf{x};\theta) - C_0(\theta)Z_0$. Although $\widetilde{\mathsf{G}}_\mathbf{m}(\cdot;\theta)$ has the boundary condition $\widetilde{\mathsf{G}}_\mathbf{m}(0;\theta) = 0$, $\mathsf{Y}_{\mathsf{G_m}}(\cdot)$ has no boundary conditions because the constant trend compensates for the boundary condition of $\widetilde{\mathsf{G}}_\mathbf{m}(\cdot;\theta)$ at the origin, that is, $\mathsf{Y}_{\mathsf{G_m}}(0;\theta) = \beta_0$. In general, we define the GRF $\widetilde{\mathsf{G}}_\mathbf{m}(\cdot;\theta)$ to be the zero mean GRF whose covariance function is the covariance function of the **m**-GIBF with the proper terms subtracted. We denote the covariance function of $\widetilde{\mathsf{G}}_\mathbf{m}(\cdot;\theta)$ by $\Sigma_{\widetilde{\mathsf{G}}_\mathbf{m}(\cdot,\cdot;\theta)}$.

5.2. Parameter Estimation

We first discuss finding the MLEs $\widehat{\beta}$ and $\widehat{\theta}$ of β and θ, respectively, with the order **m** of the GIBF fixed, in Section 5.2.1. Then, we discuss finding the optimal order $\widehat{\mathbf{m}}$ of the GIBF in Section 5.2.2. In the following, we assume that the vector of basis functions has been chosen and fixed. Furthermore, following Ankenman et al. (2010) and Chen et al. (2012), we assume independent sampling (i.e., no CRN) and use the plug-in estimator $\widehat{\Sigma}_\epsilon \triangleq \mathrm{diag}\{s^2(\mathbf{x}_1)/n_1, s^2(\mathbf{x}_2)/n_2, \ldots, s^2(\mathbf{x}_k)/n_k\}$ for Σ_ϵ.

After we obtain $\widehat{\beta}$, $\widehat{\theta}$, and $\widehat{\mathbf{m}}$ using the methods in this section, we substitute them and $\widehat{\Sigma}_\epsilon$ into Equation (2) to get the estimated stochastic kriging prediction $\widehat{\mathsf{Y}}_{\mathsf{G}_{\widehat{\mathbf{m}}}}(\mathbf{x}_0)$ at $\mathbf{x}_0$ (Ankenman et al. 2010). Furthermore, the estimated MSE of $\widehat{\mathsf{Y}}_{\mathsf{G}_{\widehat{\mathbf{m}}}}(\mathbf{x}_0)$ is given by substituting $\widehat{\beta}$, $\widehat{\theta}$, $\widehat{\mathbf{m}}$, and $\widehat{\Sigma}_\epsilon$ into Equation (3) and adding the term $\eta^\top(\mathbf{F}^\top(\Sigma_{\widetilde{\mathsf{G}}_{\widehat{\mathbf{m}}}}(\widehat{\theta}) + \widehat{\Sigma}_\epsilon)^{-1}\mathbf{F})^{-1}\eta$, where $\eta \triangleq \mathbf{f}(\mathbf{x}_0) - \mathbf{F}^\top(\Sigma_{\widetilde{\mathsf{G}}_{\widehat{\mathbf{m}}}}(\widehat{\theta}) + \widehat{\Sigma}_\epsilon)^{-1}\Sigma_{\widetilde{\mathsf{G}}_{\widehat{\mathbf{m}}}}(\mathbf{x}_0,\cdot;\widehat{\theta})$. This additional term accounts for the uncertainty introduced by estimating the vector β of regression coefficients.

5.2.1. Finding the MLEs of β and θ, with m Fixed.
Assume that the order **m** of the GIBF has been fixed. Finding the MLEs $\widehat{\beta}$ and $\widehat{\theta}$ involves solving an optimization problem with continuous decision variables.

Given a fixed value for θ, the MLE of β is $\widehat{\beta}(\theta) \triangleq (\mathbf{F}^\top \Sigma(\theta)^{-1}\mathbf{F})^{-1}\mathbf{F}^\top \Sigma(\theta)^{-1}\bar{\bar{\mathcal{Y}}}$, where $\Sigma(\theta) = \Sigma_{\widetilde{\mathsf{G}}_\mathbf{m}}(\theta) + \widehat{\Sigma}_\epsilon$, and $\widehat{\beta}$ and Σ have been written as functions of θ to explicitly show dependence. If we profile over the MLE of β and ignore constants, then the profile log-likelihood function (Shao 2010) is given by

$$\mathcal{L}(\theta|\bar{\bar{\mathcal{Y}}}) = -\frac{1}{2}\log(|\Sigma(\theta)|)$$
$$-\frac{1}{2}\left(\bar{\bar{\mathcal{Y}}} - \mathbf{F}\widehat{\beta}(\theta)\right)^\top \Sigma(\theta)^{-1}\left(\bar{\bar{\mathcal{Y}}} - \mathbf{F}\widehat{\beta}(\theta)\right),$$

where $\bar{\bar{\mathcal{Y}}}$ is the vector of simulation output. Note, because we use the plug-in estimator $\widehat{\Sigma}_\epsilon$ for Σ_ϵ, the likelihood function $\mathcal{L}(\theta|\bar{\bar{\mathcal{Y}}})$ is not the full data likelihood (Binois et al. 2018). Assuming that unknown variances and hyperparameters of the GRF are indeed known can be a problem if doing so leads to underestimation of the uncertainty in the metamodel. However, in this paper, we are only concerned with obtaining a metamodel that yields point estimates, so we use the plug-in estimator $\widehat{\Sigma}_\epsilon$ for Σ_ϵ. The MLE $\widehat{\theta}$ is given by the solution to $\arg\min_\theta\{-\mathcal{L}(\theta|\bar{\bar{\mathcal{Y}}})|\theta \in \mathbb{R}_{>0}^M\}$, where $M = \sum_{i=1}^d (m_i + 2)$ and $\mathbb{R}_{>0}^M$ is the set of feasible values for θ.

Instead of searching over the unbounded space $\mathbb{R}_{>0}^M$ for the MLE of θ, we add a dummy parameter τ, which allows θ to be restricted to the M-dimensional unit hypercube. In other words, only the magnitudes of the parameters in θ relative to each other are important because the actual magnitude is absorbed in τ. The reparameterized covariance function for **m**-GIBF is

$$\Sigma_{\mathsf{G_m}}(\mathbf{x},\mathbf{y};\theta,\tau) = \tau\prod_{i=1}^d\left(\sum_{k=0}^{m_i}\theta_{i,k}\frac{x_i^k y_i^k}{(k!)^2}\right.$$
$$\left. + \theta_{i,m_i+1}\int_0^\infty \frac{(x_i - u_i)_+^{m_i}(y_i - u_i)_+^{m_i}}{(m_i!)^2}du_i\right),$$

where now θ lies in the M-dimensional unit hypercube and $\tau \geq 0$. The MLE $\widehat{\theta}$ can now be found by solving $\arg\min_\theta\{-\mathcal{L}(\theta,\tau^*(\theta)|\bar{\bar{\mathcal{Y}}})|\theta \in [0,1]^M\}$, where

$$\mathcal{L}(\theta,\tau|\bar{\bar{\mathcal{Y}}}) = -\frac{1}{2}\log(|\Sigma(\theta,\tau)|)$$
$$-\frac{1}{2}\left(\bar{\bar{\mathcal{Y}}} - \mathbf{F}\widehat{\beta}(\theta,\tau)\right)^\top \Sigma(\theta,\tau)^{-1}\left(\bar{\bar{\mathcal{Y}}} - \mathbf{F}\widehat{\beta}(\theta,\tau)\right)$$

is the reparameterized profile log-likelihood function, $\widehat{\beta}$ and Σ have been written as functions of θ and τ to explicitly show dependence, and $\tau^*(\theta)$ is the value of τ that minimizes $\mathcal{L}(\theta,\tau|\bar{\bar{\mathcal{Y}}})$ with θ fixed. Finding $\tau^*(\theta)$ can be done efficiently using a line search method and supplying the solver with the gradient $\partial\mathcal{L}(\theta,\tau|\bar{\bar{\mathcal{Y}}})/\partial\tau$. We can now solve the constrained optimization problem by evaluating $-\mathcal{L}(\theta,\tau^*(\theta)|\bar{\bar{\mathcal{Y}}})$ at a low-discrepancy point-set in the M-dimensional unit

hypercube and use the point that minimizes this quantity as the starting solution for a nonlinear optimization algorithm.

5.2.2. Finding the Optimal m.
The set of feasible values for $\mathbf{m}$ is the set $\mathbb{Z}_{\geq 0}^d$, because $\mathbf{m}$ must be a vector of integers. To find the optimal order $\widehat{\mathbf{m}}$, we first fix $\mathbf{m}$ to some value $\mathbf{m}' \in \mathbb{Z}_{\geq 0}^d$ and calculate the Monte Carlo cross-validation error for $\mathbf{m}'$-GIBF as follows:

0. (Initialization) Let Ξ denote the set of indices of the design points not on the boundary of the convex hull of the design points (to avoid extrapolation). Let p_{CV} denote the number of design points in Ξ to hold out on each iteration, where $1 \leq p_{CV} < |\Xi|$. Furthermore, let n_{CV} denote the number of iterations. Set $i = 1$.

1. Randomly select p_{CV} indices from Ξ and let Ξ_{CV} denote the set of these indices. Similarly, let $\Xi_{\backslash CV}$ denote the set of indices $\Xi \backslash \Xi_{CV}$.

2. Build a metamodel $\widehat{Y}_{\mathbf{G}_{\mathbf{m}'}}$ using the design points $\{\mathbf{x}_j | j \in \Xi_{\backslash CV}\}$. The MLE of θ for $\widetilde{G}_{\mathbf{m}'}(\cdot; \theta)$ and the MLE of β are obtained using the method in Section 5.2.1 with the order fixed to $\mathbf{m}'$.

3. Calculate

$$\text{MSE}_{\mathbf{m}'}^i = \frac{1}{p_{CV}} \sum_{j \in \Xi_{CV}} \left(\widehat{Y}_{\mathbf{G}_{\mathbf{m}'}}(\mathbf{x}_j) - \bar{\mathcal{Y}}(\mathbf{x}_j) \right)^2.$$

4. If $i = n_{CV}$, go to 5. Else, set $i \leftarrow i + 1$ and go to 1.
5. Calculate

$$\text{CV}_{\mathbf{m}'} = \frac{1}{n_{CV}} \sum_{i=1}^{n_{CV}} \text{MSE}_{\mathbf{m}'}^i.$$

The cross-validation error $\text{CV}_{\mathbf{m}'}$ is used to evaluate $\mathbf{m}'$-GIBF. After we have calculated $\text{CV}_{\mathbf{m}'}$, we choose a different value of $\mathbf{m} \neq \mathbf{m}'$ in $\mathbb{Z}_{\geq 0}^d$ and repeat the process until we are satisfied with our solution in terms of the cross-validation error (i.e., we do not exhaust the search space $\mathbb{Z}_{\geq 0}^d$ of $\mathbf{m}$). The optimal order $\widehat{\mathbf{m}}$ is the order that gave the smallest cross-validation error. Our justification for using cross-validation instead of maximum likelihood estimation to choose the order is given in the e-companion.

Instead of searching over the unbounded space $\mathbb{Z}_{\geq 0}^d$ for the optimal order $\widehat{\mathbf{m}}$, we limit our search to the bounded set $\{1, 2\}^d$, which has 2^d elements. We only search the bounded set $\{1, 2\}^d$ for $\widehat{\mathbf{m}}$, because we have found in our practical experience with metamodeling of engineering simulations that it is sufficient to only consider GIBFs that are at least once-differentiable in each coordinate and at most twice-differentiable in each coordinate (i.e., $\mathbf{m}$-GIBF with $1 \leq m_i \leq 2$ for all i). These GIBFs are flexible enough for most response surfaces. When the order of the GIBF is increased in a coordinate, the computational cost of finding the MLE of θ increases because the number of parameters in θ increases. When

m_i is at most two in each coordinate, the number of parameters is manageable.

Assuming, without loss of generality, that the coordinates are ordered in terms of least important to most important (using any variable importance method), we recommend first evaluating $\mathbf{m} = \mathbf{1}^\top$ and then, starting at $i = 1$, setting the ith coordinate of $\mathbf{m}$ to 2. If the cross-validation error is improved, continue with the $(i + 1)$th coordinate, keeping the ith coordinate set to 2. Otherwise, continue with the $(i + 1)$st coordinate, setting the ith coordinate back to 1. Repeat until all coordinates are exhausted. Finally, evaluate $\mathbf{m} = \mathbf{2}^\top$ if it has not already been evaluated and choose the value of $\mathbf{m}$ that minimizes the cross-validation error. When $m_i = 0$ for some $i = 1, 2, \ldots, d$, the metamodel will be nondifferentiable in that coordinate. If prior knowledge about the response surface suggests that this might be appropriate, then we recommend broadening the search space to allow for this possibility.

5.3. Approximation Capability
Although we investigate the approximation capability on several real-world examples and test functions in Section 6, we discuss the approximation capability in a theoretical setting here. We first discuss the case of approximating polynomials and then provide the space of functions for which linear combinations of $\Sigma_{\mathbf{G}_{\mathbf{m}}}(\cdot, \cdot; \theta)$ are dense. For ease of discussion, we focus on the case of deterministic simulations (i.e., kriging with GIBFs).

Roughly and asymptotically speaking, when the underlying response surface y is a polynomial, we can reproduce y exactly if a polynomial trend model and a GIBF of sufficiently high order (determined by the order of y) is used. Indeed, when a trend model is used in kriging with GIBFs, the deviations from the trend are modeled as a realization of a GIBF. When a polynomial trend model of sufficiently high order is used, the polynomial trend itself will be able to reproduce y because it is a polynomial of appropriate order. Even without a polynomial trend model (or any trend model), we can reproduce polynomials exactly by using a GIBF of sufficiently high order. This follows from the fact that the integral part of $\Sigma_{G_m}(s, t; \theta)$ coincides with a polynomial in s of degree less than or equal to $2m + 1$ for $s \leq t$, and with a polynomial in s of degree less than or equal to m for $s \geq t$ (Berlinet and Thomas-Agnan 2004). Thus, the resulting metamodel will be the linear combination of polynomial terms.

To provide the space of functions for which linear combinations of $\Sigma_{\mathbf{G}_{\mathbf{m}}}(\cdot, \cdot; \theta)$ are dense, we first construct the RKHS corresponding to the reproducing kernel $\Sigma_{\mathbf{G}_{\mathbf{m}}}(\cdot, \cdot; \theta)$ (which we now denote by $K_{\mathbf{m}+1}(\cdot, \cdot; \theta)$ to be consistent with the functional analysis literature). For more details or proofs regarding this construction, see Berlinet and Thomas-Agnan (2004). Let $\mathcal{D}'(\mathbb{R})$ denote the space of Schwartz distributions (generalized

functions) (i.e., the space of continuous linear functionals of the space of infinitely differentiable functions with compact support). Let D denote the derivative operator, and let $\mathbb{L}^2(0, 1)$ denote the space of square integrable functions on $(0, 1)$ with respect to Lebesgue measure defined up to almost everywhere equality. Furthermore, for the purposes of this section, consider GIBFs restricted to $[0, 1]^d$. The Sobolev space of functions

$$H^{m+1}(0, 1) \triangleq \left\{ \phi \in \mathcal{D}'(\mathbb{R}) \, | \, D^\alpha \phi \in \mathbb{L}^2(0, 1), \, \forall \, \alpha \le m + 1 \right\}$$

is an RKHS endowed with the inner product

$$\langle \phi_1, \phi_2 \rangle_{H^{m+1}(0,1)} = \sum_{n=0}^m \frac{1}{\theta_n} \phi_1^{(n)}(0) \phi_2^{(n)}(0) \\ + \frac{1}{\theta_{m+1}} \int_0^1 \phi_1^{(m+1)}(t) \phi_2^{(m+1)}(t) d\lambda(t),$$

where $\phi_1, \phi_2 \in H^{m+1}(0, 1)$ and λ denotes Lebesgue measure on the set $\mathbb{R}$. The reproducing kernel of $H^{m+1}(0, 1)$ is given by

$$K_{m+1}(s, t; \theta) = \sum_{n=0}^m \theta_n \frac{s^n t^n}{(n!)^2} + \theta_{m+1} \int_0^1 \frac{(s - u)_+^m (t - u)_+^m}{(m!)^2} du.$$

Now, let $\mathcal{H} = \otimes_{i=1}^d H^{m_i+1}(0, 1)$ denote the functional completion of the tensor product $\tilde{\otimes}_{i=1}^d H^{m_i+1}(0, 1)$ of the vector spaces $H^{m_i+1}(0, 1)$ for $i = 1, 2, \ldots, d$. $\mathcal{H}$ is an RKHS endowed with the reproducing kernel

$$K_{\mathbf{m}+1}(\mathbf{s}, \mathbf{t}; \theta) = \prod_{i=1}^d \left(\sum_{n=0}^{m_i} \theta_{i,n} \frac{s_i^n t_i^n}{(n!)^2} \right. \\ \left. + \theta_{i,m_i+1} \int_0^1 \frac{(s_i - u_i)_+^{m_i} (t_i - u_i)_+^{m_i}}{(m_i!)^2} du_i \right).$$

Theorem 4. *The span of the functions* $\{K_{\mathbf{m}+1}(\cdot, \mathbf{x}; \theta) \, | \, \mathbf{x} \in [0, 1]^d\}$ *is dense in* $\mathcal{H}$.

This theorem is a direct corollary of the Moore-Aronszajn theorem (Berlinet and Thomas-Agnan 2004). Thus, asymptotically speaking (i.e., as the number of design points goes to infinity), we are able to approximate any function in $\mathcal{H}$ to within a given level of error using kriging with GIBFs. Here, error is measured with respect to the norm induced by $\langle \cdot, \cdot \rangle_{H^{m+1}(0,1)}$.

5.4. Metamodel Differentiability

To analyze the differentiabilty of metamodels constructed using stochastic kriging with GIBFs, we rewrite the estimated stochastic kriging predictor $\widehat{Y}_{G_{\widehat{\mathbf{m}}}}(\cdot)$ as the affine combination of the k basis functions $\Sigma_{\widetilde{G}_{\widehat{\mathbf{m}}}}(\cdot, \mathbf{x}_i; \widehat{\theta})$, for $i = 1, 2, \ldots, k$. Indeed, for $\mathbf{x} \in \mathcal{X}$, we have

$$\widehat{Y}_{G_{\widehat{\mathbf{m}}}}(\mathbf{x}) = \sum_{i=1}^p f_i(\mathbf{x}) \widehat{\beta}_i + \sum_{i=1}^k c_i \Sigma_{\widetilde{G}_{\widehat{\mathbf{m}}}}(\mathbf{x}, \mathbf{x}_i; \widehat{\theta}),$$

where $\mathbf{c} = \Sigma(\widehat{\theta})^{-1}(\bar{\mathcal{Y}} - \mathbf{F}\widehat{\beta})$. Using this formulation of the stochastic kriging predictor, we can see that $f_i(\cdot)$ and $\Sigma_{\widetilde{G}_{\widehat{\mathbf{m}}}}(\cdot, \mathbf{x}_i; \widehat{\theta})$ are the only terms that depend on $\mathbf{x}$ in this expression, so the differentiability of the metamodel is determined by $\mathbf{f}(\cdot)$ and $\Sigma_{\widetilde{G}_{\widehat{\mathbf{m}}}}(\cdot, \cdot; \widehat{\theta})$. If each function, $f_i(\cdot)$, in the trend vector is a polynomial then it is infinitely differentiable with respect to the ith coordinate.

The ith term in the product of $\Sigma_{\mathbf{G}_{\widehat{\mathbf{m}}}}(\cdot, \cdot; \widehat{\theta})$ is $2\widehat{m}_i$ times differentiable with respect to the ith coordinate with the $(2\widehat{m}_i + 1)$ derivative with respect to the ith coordinate having a discontinuity (Berlinet and Thomas-Agnan 2004). Thus, the metamodel is $2\widehat{m}_i$ times differentiable in the ith coordinate. Of course, the result that we can control the differentiability separately for each coordinate is expected because $\mathbf{G}_{\widehat{\mathbf{m}}}(\cdot; \widehat{\theta})$ is $\widehat{m}_i$ times differentiable with respect to the ith coordinate.

6. Numerical Experiments

The purpose of the experiments is to assess the prediction capability of stochastic kriging with GIBFs. We are mainly concerned with how different types of response surfaces and different levels of Monte Carlo noise, including no noise in the simulation output (i.e., we are able to observe the actual response surface at the design points), affect our predictions.

We compare stochastic kriging with GIBFs versus stochastic kriging with the Gaussian and Matérn GRFs. The Gaussian and Matérn GRFs can result in metamodels with mean reversion. Because we wish to demonstrate the superiority of GIBFs to the Gaussian and Matérn GRFs even when the latter do not exhibit mean reversion, we use experiment designs with sufficiently many design points; the case in which the latter exhibit mean reversion is left to the e-companion. Because the Markov property is a theoretical, rather than practical, property we desire of a GRF, we do not investigate the effect this property has on prediction capability directly; isolating the impact of the Markov property on the prediction capability is also left to the e-companion.

In addition to comparing to the Gaussian and Matérn GRFs, we also compare with standard integrated Brownian fields (IBFs) to assess the impact of using GIBFs over IBFs, and thus determining whether accounting for the boundary conditions of IBFs is indeed necessary.

In our experiments, we use no trend model for GIBFs and a constant trend model (i.e., $\mathbf{f}(\mathbf{x}) = (1)^\top$) for the Gaussian and Matérn GRFs and set the number p_{CV} of design points to hold out on each CV iteration to 3 and the number n_{CV} of CV iterations to find the optimal order to 50. In each experiment, the design points are the first k points from a scrambled Sobol point-set rescaled to fit within the design space. The prediction points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{1023}$ are the first 1,023 points

from the Halton point-set rescaled to fit within the design space and the measure of prediction ability is the root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{1023} \sum_{i=1}^{1023} \left(\hat{y}_m(\mathbf{p}_i) - y_a(\mathbf{p}_i) \right)^2},$$

where $\hat{y}_m(\mathbf{p}_i)$ is the value of the simulation metamodel $\hat{y}_m$ at $\mathbf{p}_i$, and $y_a(\mathbf{p}_i)$ is the actual value of the response surface at $\mathbf{p}_i$. We perform 50 macroreplications (i.e., we repeat each experiment 50 times). Plots of the response surfaces and some metamodels, as well as further recommendations and discussion for using GIBFs, are given in the e-companion. Metamodels corresponding to the Gaussian GRF were fit using the R package *mlegp*, and the metamodels corresponding to the Matérn GRF were fit using maximum likelihood estimation.

### 6.1. Credit Risk Simulation
In this example, the response surface is the expected loss of a credit portfolio, given values of latent variables that trigger the default of the obligors (Glasserman et al. 2008). Consider a credit portfolio with n obligors, and let Y_k be the default indicator (=1 for default, =0 otherwise) and l_k be the deterministic loss resulting from default of the kth obligor. The dependence among the default indicators Y_k is modeled by a multifactor Gaussian copula model with a finite number of types, which is a function of the d-dimensional

standard Gaussian random vector $\mathbf{Z}$. The total loss from defaults is $L_n = \sum_{k=1}^n l_k Y_k$, which is a discrete random variable. The response surface is $y_{cr}(\mathbf{x}) = \mathbb{E}[L_n|\mathbf{Z} = \mathbf{x}]$, and a closed-form expression for $y_{cr}(\mathbf{x})$ is available and is used to obtain noiseless observations of the response surface, as well as to determine the accuracy of the predictions. To obtain noisy observations (simulation output) of the response surface, we use the importance sampling method of Glasserman et al. (2008) to estimate the expected loss $\mathbb{E}[L_n|\mathbf{Z} = \mathbf{x}]$ of the credit portfolio.

In our experiments, the number of replications run at scenario $\mathbf{x}$ is chosen so that the sample standard deviation across replications is $\sigma y_{cr}(\mathbf{x})$, where we control σ to achieve different levels of noise in the simulation output. For our particular example, consider the case with two factors and four types of obligors: $\mathbf{a}_1^\top = (0.85, 0)$, $\mathbf{a}_2^\top = (0.25, 0)$, $\mathbf{a}_3^\top = (0, 0.85)$, and $\mathbf{a}_4^\top = (0, 0.25)$. The $\mathbf{a}_i$, $i = 1, 2, \ldots, 4$, are defined in Glasserman et al. (2008) and provided here so that the experiments can be reproduced. Each type has three obligors (i.e., $n = 12$), with $l_k = 1$ and $p_k = 0.01$ for every obligor. The design space for this example is the square $[-5, 10]^2$. The actual value of the response surface is computed using the closed-form expression given in Glasserman et al. (2008).

6.1.1. Experiment Results. The experiment results for the credit risk simulation are given in Figure 4 for

Figure 4. (Color online) Box-Plots of RMSE for the Credit Risk Simulation
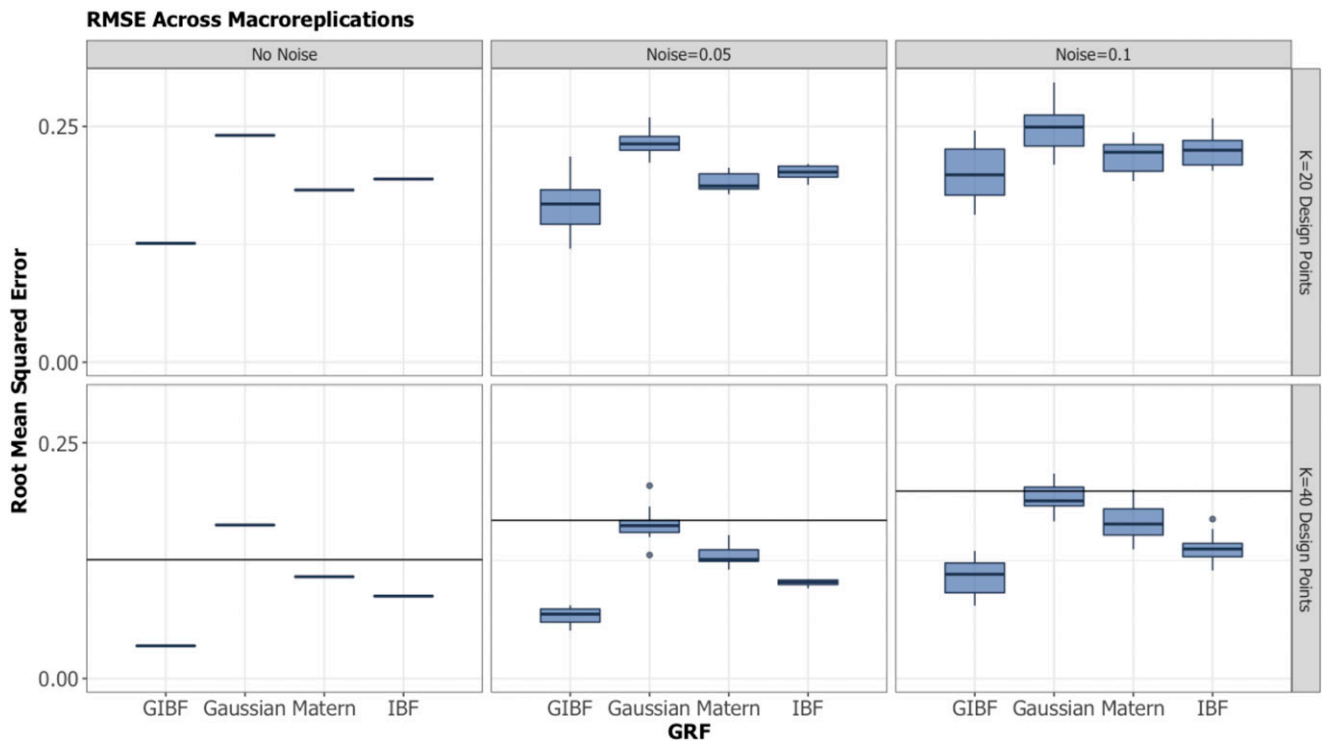
varying numbers of design points and Monte Carlo noise. An interesting characteristic of the credit risk response surface y_{cr} occurs in regions of the design space where there is a change in the number of types of obligors that are likely to default. In these regions of the design space, there is an abrupt change in the response surface, which causes a rapid change in the first partial derivatives. We can see from the experiment results that using GRFs whose differentiability can be controlled (i.e., GIBF and the Matérn GRF) resulted in better predictions than the infinitely differentiable Gaussian GRF. When we use a GRF whose differentiability can be controlled, the order of differentiability can be adjusted to account for these abrupt changes (or lack of smoothness) in the response surface. In this example, the optimal values for the parameters of both GIBF and the Matérn GRF resulted in GRFs with lower orders of differentiability. Between GIBF and the Matérn GRF, the metamodels constructed using GIBF resulted in better predictions. When the variance of the noise in the simulation output was increased, the RMSE corresponding to all GRFs increased as expected. However, the benefit (in terms of the ratio between RMSEs) of using GIBF over the Gaussian and Matérn GRFs was greatest when no noise was present in the simulation output and decreased as the noise level increased.

In Figure 4, as well as following figures displaying experiment results, a solid horizontal line is plotted in each plot for the row corresponding to $k = 40$ design points. This solid horizontal line marks the median of the GIBF RMSEs from the experiment with the same level of noise and $k = 20$ design points. In general, we can see that roughly half as many design points are needed by GIBF to obtain similar RMSEs for the Gaussian and Matérn GRFs; this becomes even clearer in following experiments. Furthermore, for fixed k, the median RMSEs of GIBF using noisy simulation output ($\sigma = 0.1$) is similar to the RMSEs of the Gaussian and Matérn GRFs using simulation output with no noise.

The optimal order for GIBF, obtained using the method in Section 5.2.2, gave the best predictions in every case; this optimal order was $(1, 1)$. Although there are more parameters for GIBFs (specifically, in the vector θ) than the Gaussian and Matérn GRFs, finding the MLE of θ for GIBFs can be done very fast because, from our practical experience, the prediction ability of GIBFs is not sensitive to θ. The prediction ability of the Gaussian and Matérn GRFs is sensitive to their θ parameters (as defined in the beginning of Section 3). For example, if the θ parameters for the Gaussian are too large, then the metamodel will exhibit mean reversion. Conversely, if the θ parameters

for the Gaussian are too small, then the correlations among the design points and the prediction point will be too strong and more likely to result in an ill-conditioned covariance matrix.

By comparing the experiment results for GIBFs and IBFs in Figure 4, we can see that a benefit is obtained from accounting for the boundary conditions of IBFs. The optimal order of IBF (for this experiment and following experiments) was also obtained using the method in Section 5.2.2.

6.2. Expected Profit of a Two-Product Assortment

In this example, the response surface is the expected profit of a two-product assortment, as a function of their prices, where the stock levels are chosen optimally for each price pair (Aydin and Porteus 2008). We provide the necessary simulation model inputs used for our experiments and refer the reader to Aydin and Porteus (2008) for specific details about calculating the optimal stock levels and simulating the system. We assume the demand model of Aydin and Porteus (2008) and consider a two-product inventory and pricing problem with stochastic logit demand, where $\alpha_1 = 10$, $\alpha_2 = 25$, $c_1 = 12$, $c_2 = 24$ (defined in Aydin and Porteus 2008 and provided here so that the experiments can be reproduced), and the random error terms are uniformly distributed between 100 and 400. The price of the two products, denoted by the vector $\mathbf{x}$, varies over the rectangle $[7, 17] \times [21, 51]$; that is, the price of the first product varies over $[7, 17]$ and the price of the second product varies over $[21, 51]$. Similar to the credit risk simulation, the number of replications run at scenario $\mathbf{x}$ is chosen so that the sample standard deviation across replications is $\sigma y_{ep}(\mathbf{x})$, where the closed-form solution y_{ep} is obtained using the method in Aydin and Porteus (2008).

6.2.1. Experiment Results. The experiment results for the expected profit simulation are given in Figure 5 and are similar to the results of the credit risk simulation. Namely, using GIBF resulted in better predictions than the Gaussian and Matérn GRFs. We can also see that when twice as many design points ($k = 40$) were used with the Gaussian and Matérn GRFs, the prediction ability was still not as accurate as when using GIBF with half as many design points ($k = 20$). Furthermore, for both $k = 20$ and $k = 40$, GIBF built a more accurate metamodel using noisy observations ($\sigma = 0.1$) than the Gaussian and Matérn GRFs using observations with no noise. In each experiment, the optimal order of GIBF was either $(1, 1)$ or $(2, 2)$, and in every case the optimal order found using the method in Section 5.2.2 also gave the best predictions.

Figure 5. (Color online) Box-Plots of RMSE for the Expected Profit Simulation
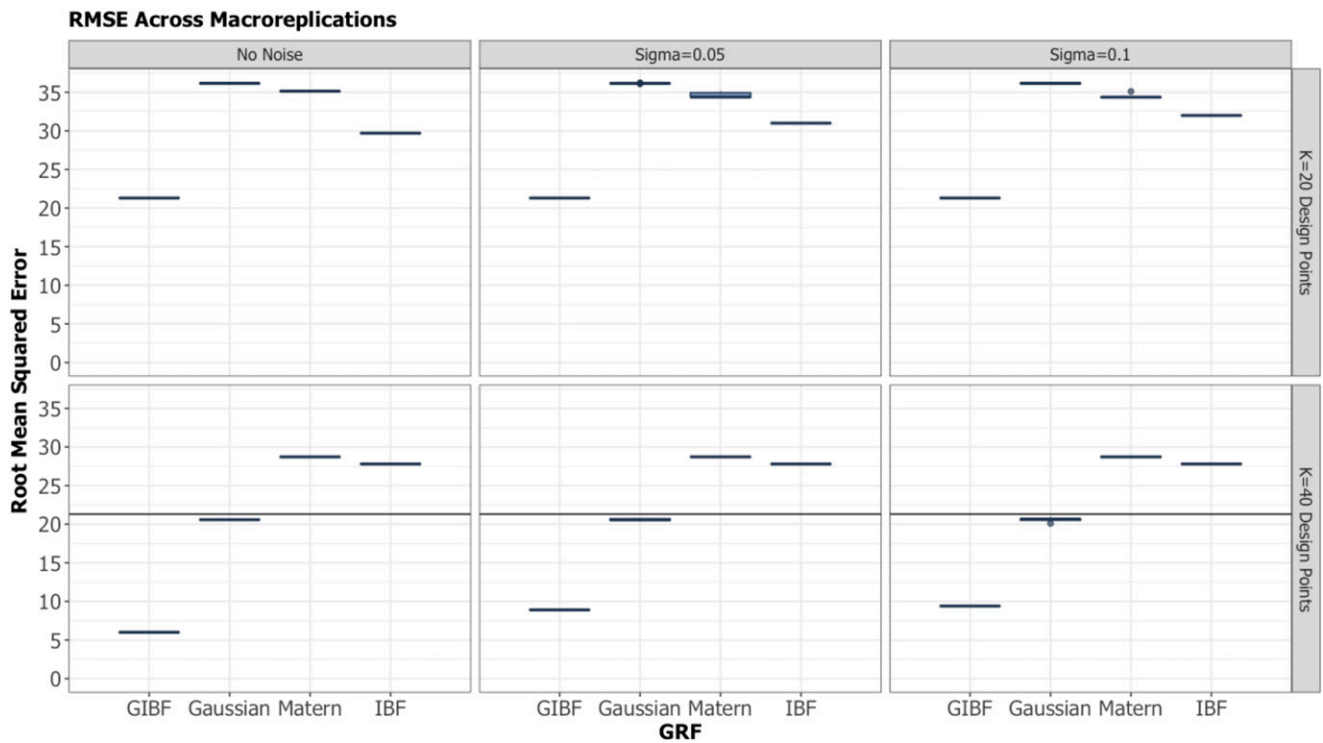
6.3. Test Functions

In this section we use three test functions with varying properties to determine how GIBFs can handle different types of response surfaces. The three test functions we use are:

$$y_{Alpine-1}(\mathbf{x}) = |x_1 \sin(x_1) + 0.1x_1| + |x_2 \sin(x_2) + 0.1x_2|$$
$$\text{on } [0,4] \times [-1,1]$$

$$y_{Camel-6}(\mathbf{x}) = (4 - 2.1x_1^2 + \frac{x_1^4}{3})x_1^2 + x_1 x_2 + (-4 + 4x_2^2)x_2^2$$
$$\text{on } [-2,2] \times [-1,1]$$

$$y_{Camel-3}(\mathbf{x}) = 2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} + x_1 x_2 + x_2^2 \quad \text{on } [-2,2]^2.$$

The Alpine-1 test function, on the region we have chosen, is nondifferentiable in the first coordinate and differentiable in the second coordinate. We use this test function to assess the benefit gained from utilizing the differentiability property of GIBFs, namely, the differentiability of GIBFs can vary in each coordinate. The Camel-6 and Camel-3 test functions represent other complicated response surfaces. To obtain noisy observations, we add a mean zero Gaussian random variable with standard deviation $\sigma y(\mathbf{x})$ to the test function value.

6.3.1. Experiment Results. The experiment results for each of the test functions are given in Figures 6–8. As

can be seen in these figures, stochastic kriging with GIBFs resulted in better predictions than with the Gaussian and Matérn GRFs. Similar to the credit risk and expected profit simulation results, the benefit (in terms of the ratio between RMSEs) was greatest when no noise was present in the simulation ouput and decreased as the noise level increased.

As mentioned earlier, the Alpine-1 test function on the region we have chosen is nondifferentiable in the first coordinate and differentiable in the second coordinate. The optimal order of GIBF chosen using the method in Section 5.2.2 was $(1,2)$ for every experiment with the Alpine-1 test function and gave the best predictions in every case. For the other test functions, the optimal order obtained using the method in Section 5.2.2 did not always give the best predictions. In two cases (Camel-6 with $k = 20$ and no noise, and Camel-3 with $k = 40$ and no noise), the order obtained using the method in Section 5.2.2 for some macroreplications led to the second-best predictions out of all GIBFs. From our practical experience, these situations occur when the GIBF whose order is chosen by the method in Section 5.2.2 and the GIBF whose order leads to the best predictions result in metamodels with similar accuracy and quality. In other words, if both metamodels are sufficiently accurate, then the method in Section 5.2.2 may choose the order of GIBF leading to the slightly less accurate metamodel. However, in our practical experience, if

Figure 6. (Color online) Box-Plots of RMSE for the Alpine–1 Test Function
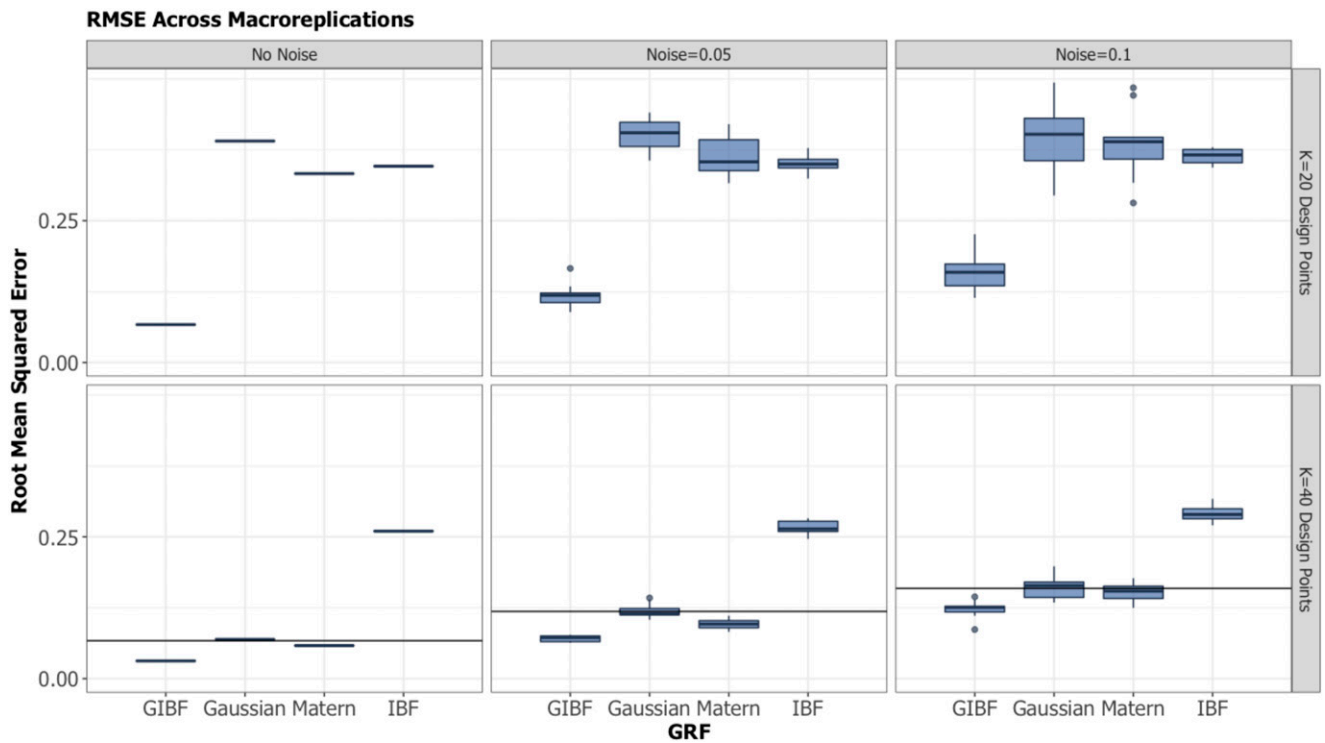
there is an order of GIBF that results in a clearly better metamodel, the method in Section 5.2.2 will choose that order.

For the Camel-3 test function, with $k = 40$ and no noise, using the Gaussian GRF led to numerical

instabilities in the inversion of the covariance matrix. Thus, a stochastic kriging metamodel could not be constructed (and thus is omitted from the plot). However, GIBF and the Matérn GRF did not experience numerical instabilities in any case.

Figure 7. (Color online) Box-Plots of RMSE for the Camel–6 Test Function
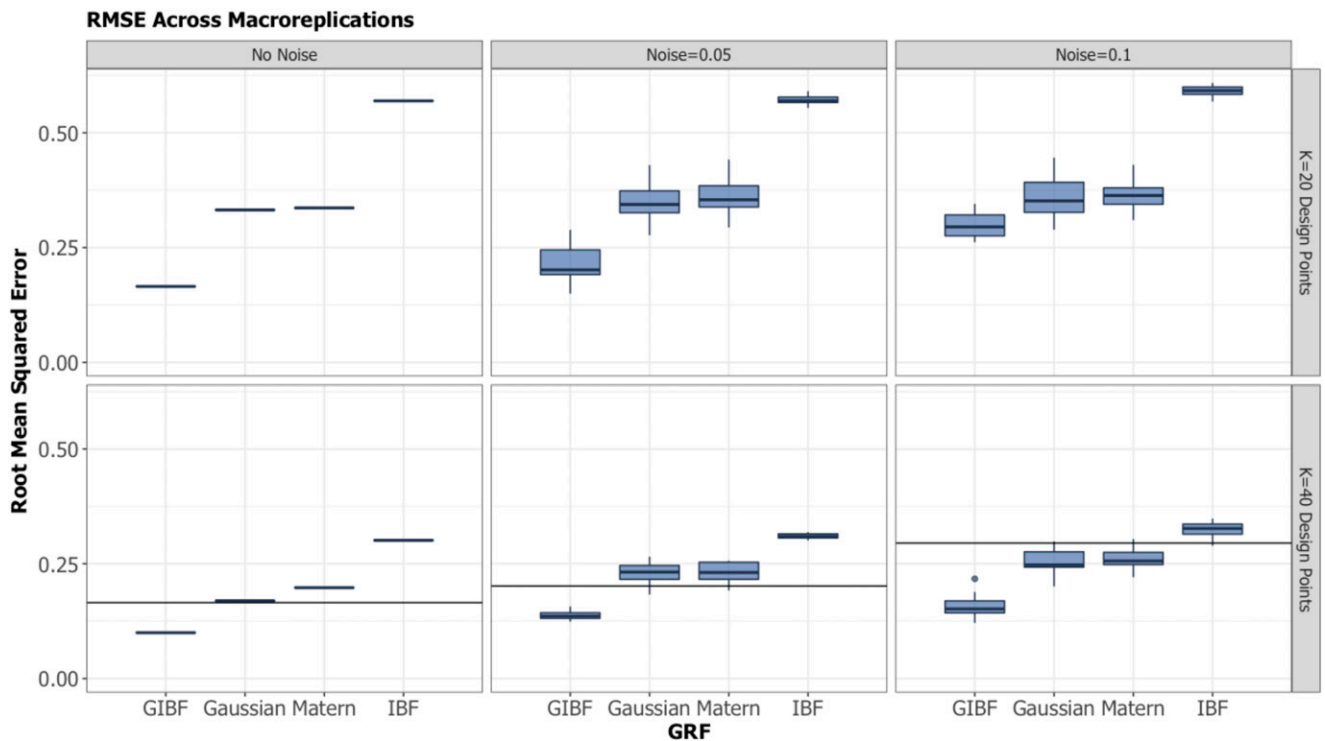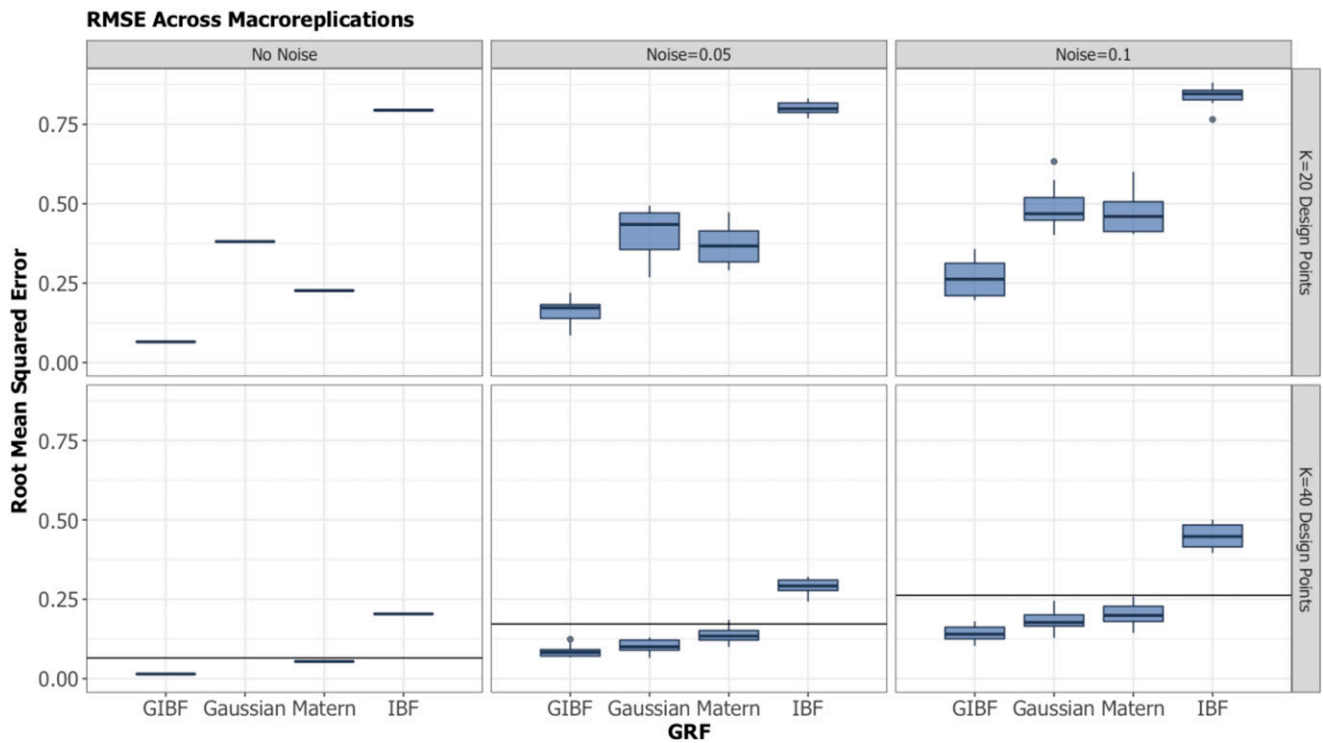
Figure 8. (Color online) Box-Plots of RMSE for the Camel–3 Test Function

RMSE Across Macroreplications

7. Conclusion

In this paper, we introduced a novel class of GRFs called generalized integrated Brownian fields (GIBFs), focusing on their use with GP regression for deterministic and stochastic simulation metamodeling. We constructed GIBFs in a probabilistic setting and discussed several of their properties, including differentiability that can differ in each coordinate, no mean reversion, and the Markov property. We showed how to build GP regression metamodels using stochastic kriging with GIBFs, discussed their approximation capability and metamodel differentiability, and used several examples to assess their prediction capability. These examples exhibited both the flexibility and the substantial improvement in predictions when using stochastic kriging with GIBFs instead of the Gaussian and Matérn GRFs. The e-companion for this paper provides additional material, including (1) proofs of all theorems, (2) a justification for using cross-validation to choose the optimal order, (3) plots of the response surfaces and some metamodels from the numerical experiments, (4) further experiments investigating the impact of no mean reversion, and the nonstationarity and Markov property of GIBFs, (5) a discussion of when GIBFs should be used, and (6) how the number of design points, the number of replications, and the dimensionality of the problem influence performance.

References

Ankenman B, Nelson BL, Staum J (2010) Stochastic kriging for simulation metamodeling. *Oper. Res.* 58(2):371–382.

Aydin G, Porteus EL (2008) Joint inventory and pricing decisions for an assortment. *Oper. Res.* 56(5):1247–1255.

Berlinet A, Thomas-Agnan C (2004) *Reproducing Kernel Hilbert Spaces in Probability and Statistics* (Kluwer Academic Publishers, New York).

Binois M, Gramacy RB, Ludkovski M (2018) Practical heteroskedastic Gaussian process modeling for large simulation experiments. *J. Comput. Graphical Statist.* 27(4):808–821.

Chen X, Ankenman B, Nelson BL (2012) The effects of common random numbers of stochastic kriging metamodels. *ACM Trans. Modeling Comput. Simulation* 22(2):1–20.

Chen X, Ankenman B, Nelson BL (2013) Enhancing stochastic kriging metamodels with gradient estimators. *Oper. Res.* 61(2):512–528.

Fill JA, Torcaso F (2004) Asymptotic analysis via Mellin transforms for small deviations in *l*2-norm of integrated Brownian sheets. *Probab. Theory Related Fields* 130(2):259–288.

Glasserman P, Kang W, Shahabuddin P (2008) Fast simulation of multifactor portfolio credit risk. *Oper. Res.* 56(5):1200–1217.

Holden H, Øksendal B, Ubøe J, Zhang T (2010) *Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach,*

2nd ed., Universitext (Springer Science and Business Media, New York).

Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J. Global Optim.* 13(4):455–492.

Joseph VR (2006) Limit kriging. *Technometrics* 48(4):458–466.

Joseph VR, Hung Y, Sudjianto A (2008) Blind kriging. *J. Mech. Design* 130(3):031102-1–031102-8.

Kleijnen J, van Beers W (2005) Robustness of kriging when interpolating in random simulation with heterogeneous variance: Some experiments. *Eur. J. Oper. Res.* 165(3):826–834.

Krige DG (1951) A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master's thesis, University of Witwatersrand, Johannesburg, South Africa.

Li R, Sudjianto A (2005) Analysis of computer experiments using penalized likelihood in Gaussian kriging models. *Technometrics* 47(2):111–120.

Liu M, Staum J (2010) Stochastic kriging for efficient nested simulation of expected shortfall. *J. Risk* 12(3):3–27.

Mitchell TJ, Morris MD (1992) The spatial correlation function approach to response surface estimation. *Proc. 1992 Winter Simulation Conf.* (IEEE, Piscataway, NJ), 565–571.

Paciorek CJ, Schervish MJ (2004) Nonstationary covariance functions for Gaussian process regression. *Proc. Conf. Neural Inform. Processing Systems* (MIT Press, Cambridge, MA).

Pitt LD (1971) A Markov property for Gaussian processes with a multidimensional parameter. *Archive Ration. Mech. Anal.* 43(5): 367–391.

Powell WB, Ryzhov IO (2012) *Optimal Learning* (John Wiley and Sons, Hoboken, NJ).

Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. *Statist. Sci.* 4(4):409–423.

Salemi P, Staum J, Nelson BL (2013) Generalized integrated Brownian fields for simulation metamodeling. *Proc. 2013 Winter Simulation Conf.* (IEEE, Piscataway, NJ).

Santner TJ, Williams BJ, Notz WI (2010) *The Design and Analysis of Computer Experiments* (Springer, New York).

Shao J (2010) *Mathematical Statistics* (Springer Texts in Statistics, New York).

Sun L, Hong LJ, Hu Z (2014) Balancing exploitation and exploration in discrete optimization via simulation through a Gaussian process-based search. *Oper. Res.* 62(6):1416–1438.

van Beers W, Kleijnen J (2003) Kriging for interpolation in random simulation. *J. Oper. Res. Soc.* 54(3):255–262.

Xie W, Nelson BL, Staum J (2010) The influence of correlation functions on stochastic kriging metamodels. *Proc. 2010 Winter Simulation Conf.* (IEEE, Piscataway, NJ).

Yin J, Ng SH, Ng KM (2011) Kriging metamodel with modified nugget-effect: The heteroscedastic variance case. *Comput. Indust. Engrg.* 61(3):760–777.

Zhang N, Apley DW (2014) Fractional Brownian fields for response surface metamodeling. *J. Quality Tech.* 46(4):285–301.

Zhang N, Apley DW (2016) Brownian integrated covariance functions for Gaussian process modeling: Sigmoidal versus localized basis functions. *J. Amer. Statist. Assoc.* 111(515):1182–1195.

Peter Salemi is an operations research analyst/data scientist in the Operations Research Department at The MITRE Corporation. His research interests include simulation metamodeling, optimization via simulation, and applications of Gaussian random fields.

Jeremy Staum is a former associate professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. His research interests include simulation metamodeling, optimization via simulation, and risk management.

Barry L. Nelson is the Walter P. Murphy Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. His research focus is on the design and analysis of computer simulation experiments on models of discrete-event, stochastic systems. Nelson is a Fellow of INFORMS and IIE.