

## GENERALIZED INTEGRATED BROWNIAN FIELDS FOR SIMULATION METAMODELING

Peter Salemi  
Jeremy Staum  
Barry L. Nelson

Department of Industrial Engineering and Management Sciences  
Northwestern University  
2145 Sheridan Road  
Evanston, IL, 60208-3119, U.S.A.

### ABSTRACT

We use Gaussian random fields (GRFs) that we call generalized integrated Brownian fields (GIBFs), whose covariance functions have been studied in the context of reproducing kernels, for Gaussian process modeling. We introduce GIBFs into the fields of deterministic and stochastic simulation metamodeling, and give a probabilistic representation of GIBFs that is not given in the literature on reproducing kernels. These GIBFs have differentiability that can be controlled in each coordinate, and are built from GRFs which have the Markov property. Furthermore, we introduce a new parameterization of GIBFs which allows them to be used in higher-dimensional metamodeling problems. We also show how to implement stochastic kriging with GIBFs, covering trend modeling and fitting. Lastly, we use tractable examples to demonstrate superior prediction ability as compared to the GRF corresponding to the Gaussian covariance function.

### 1 INTRODUCTION

Stochastic simulations are often used to model complex systems in industrial engineering and operations research. Although 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, which is called the simulation metamodel. Using the metamodel, quick predictions can be made at new scenarios.

A great deal of research has been directed towards experiment designs for 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 models has been remarkably successful for global metamodeling (see, for example, Santner, Williams, and Notz (2010)). Following the introduction of Gaussian process models into the design and analysis of deterministic computer experiments, Mitchell and Morris (1992) introduced Gaussian process models for representing the response surface in stochastic simulation. Since the predictions are made by fitting a Gaussian process, 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, Nelson, and Staum 2010).

In simulation metamodeling using Gaussian process models, the response surface is modeled as a sample path of a GRF. A critical choice in fitting Gaussian process models is specifying the GRF that

is used. To obtain better prediction accuracy, the GRF should have desirable properties and be flexible enough to capture the characteristics of the response surface, such as smoothness in each coordinate.

Gaussian process models were initially used in geostatistics to model the amount of gold in underground deposits (Krig 1951). For these applications, if we were interested in predicting the amount of gold in a region, knowing the amount of gold along the boundary of the region would not be sufficient information for our prediction. For example, if we also knew there was gold located just outside the region, but not necessarily at the boundary, we would expect more gold in the region. We are mainly concerned with response surfaces in operations research, which contrast with response surfaces in geostatistics. In operations research applications, 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 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: the GRF in a region, given sufficient information (level and derivatives) on the boundary, is independent of the GRF outside of the region (Pitt 1971).

The ability to control the differentiability of the GRF is a characteristic that has received considerable attention in the literature (see, for example, Santner, Williams, and Notz (2010)). A common class of GRFs that are used for metamodeling corresponds to the power exponential covariance function, for which the differentiability is controlled by a single parameter. However, these GRFs can only be non-differentiable or infinitely differentiable, depending on the value of the parameter. The generalization of the power exponential covariance function is the class of Matérn covariance functions, 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 Matérn class can have differentiability of any order. However, the differentiability cannot be controlled separately for each coordinate.

The GRFs that we use in this paper, which we call generalized integrated Brownian fields (GIBFs), have differentiability that can be controlled in each coordinate, and are built from GRFs which have the Markov property. The parameterization of GIBFs that we introduce allows for the easy use of GIBFs in higher-dimensional metamodeling problems.

The use of Gaussian process models in stochastic simulation metamodeling has led to several different metamodeling techniques. In the following section we discuss stochastic kriging, the simulation metamodeling technique we use in this paper. We present GIBFs using a probabilistic approach in Section 3, and provide a guide to using these random fields with stochastic kriging in Section 4. We conclude the paper with numerical experiments which show the improved prediction accuracy as compared to the well-known and highly-used Gaussian covariance function.

## 2 STOCHASTIC KRIGING

Gaussian process models have been used for approximating the output of deterministic computer experiments following the work of Sacks, Welch, Mitchell, and Wynn (1989), which introduced kriging into the design and analysis of deterministic computer experiments. In kriging, the response surface  $y(\mathbf{x})$  is modeled as a sample path of the GRF

$$Y_M(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + M(\mathbf{x}), \quad (1)$$

where  $\mathbf{x}$  is a design point in the design space  $\mathbb{X}$  (the space of all possible design points),  $\mathbf{f}(\mathbf{x})$  is a  $p \times 1$  vector of known functions of  $\mathbf{x}$ ,  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of unknown parameters, and  $M$  is a mean-zero GRF. In other words, sample paths of  $M$  can be thought of as being randomly sampled from a space of functions mapping  $\mathbb{R}^d \rightarrow \mathbb{R}$ , according to a Gaussian measure. The GRF  $M$  is assumed to exhibit spatial covariance, which is determined by the covariance function  $\Sigma_M(\cdot, \cdot; \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is a vector of parameters. Specifically, for two design points  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ,

$$\text{Cov}[M(\mathbf{x}_i), M(\mathbf{x}_j)] = \Sigma_M(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta}).$$

For deterministic computer experiments where 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 (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  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$  for a total of  $n_i$  replications at design point  $\mathbf{x}_i$ . Replication  $j$  at design point  $\mathbf{x}_i$  is denoted by  $\mathcal{Y}_j(\mathbf{x}_i)$ . At design point  $\mathbf{x}_i$  we collect the sample mean  $\bar{\mathcal{Y}}(\mathbf{x}_i) = (1/n_i) \sum_{j=1}^{n_i} \mathcal{Y}_j(\mathbf{x}_i)$ , and the sample standard deviation  $s^2(\mathbf{x}_i) = (1/(n_i - 1)) \sum_{j=1}^{n_i} (\mathcal{Y}_j(\mathbf{x}_i) - \bar{\mathcal{Y}}(\mathbf{x}_i))^2$ . Gaussian process methods in stochastic simulation utilize the sample means and sample standard deviations at the design points to build the Gaussian process model.

In stochastic kriging, the response surface is modeled as a sample path of the GRF  $Y_M$ , given by Equation (1), with mean  $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$  and covariance function  $\Sigma_M(\cdot, \cdot; \boldsymbol{\theta})$ . The simulation output on replication  $j$  at design point  $\mathbf{x}$  is modeled as a realization of the random variable

$$Y_{M, \varepsilon_j}(\mathbf{x}) = Y_M(\mathbf{x}) + \varepsilon_j(\mathbf{x}),$$

where the mean-zero sampling noise in the replications  $\varepsilon_1(\mathbf{x}), \varepsilon_2(\mathbf{x}), \dots$  at a design point  $\mathbf{x}$  is independent and identically distributed across replications. We also assume that the sampling noise is independent across design points, i.e., we do not use Common Random Numbers (CRN), although our method will still work when CRN are used. The sampling noise is referred to as intrinsic uncertainty, since it is inherent in the stochastic simulation. The stochastic nature of  $M$  is called extrinsic uncertainty, since it is imposed on the problem to aid in the development of the metamodel.

Suppose that the simulation model has been run at the  $k$  design points  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$  yielding the vector of observed simulation output  $\bar{\mathcal{Y}} = (\bar{\mathcal{Y}}(\mathbf{x}_1), \bar{\mathcal{Y}}(\mathbf{x}_2), \dots, \bar{\mathcal{Y}}(\mathbf{x}_k))^\top$ , and we now want to predict the response surface at  $\mathbf{x}_0$ . Let  $\widehat{\Sigma}_M$  be the  $k \times k$  variance-covariance matrix with  $ij$ th entry  $\Sigma_M(\mathbf{x}_i, \mathbf{x}_j; \widehat{\boldsymbol{\theta}})$ , where  $\widehat{\boldsymbol{\theta}}$  is the maximum likelihood estimate of  $\boldsymbol{\theta}$ , let  $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), \mathbf{f}(\mathbf{x}_2), \dots, \mathbf{f}(\mathbf{x}_k))^\top$  be the  $k \times p$  regression matrix, and let  $\widehat{\Sigma}_M(\mathbf{x}_0, \cdot)$  be the  $k \times 1$  vector of spatial covariances between the design points and the prediction point, i.e., the  $i$ th entry of  $\widehat{\Sigma}_M(\mathbf{x}_0, \cdot)$  is  $\Sigma_M(\mathbf{x}_0, \mathbf{x}_i; \widehat{\boldsymbol{\theta}})$ . Also, let  $\widehat{\Sigma}_\varepsilon = \text{diag}\{s^2(\mathbf{x}_1)/n_1, s^2(\mathbf{x}_2)/n_2, \dots, s^2(\mathbf{x}_k)/n_k\}$ . For brevity, we write  $\widehat{\Sigma} := \widehat{\Sigma}_M + \widehat{\Sigma}_\varepsilon$ . The stochastic kriging predictor is

$$\widehat{Y}_M(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \widehat{\boldsymbol{\beta}} + \widehat{\Sigma}_M(\mathbf{x}_0, \cdot)^\top \widehat{\Sigma}^{-1} (\bar{\mathcal{Y}} - \mathbf{F} \widehat{\boldsymbol{\beta}}), \quad (2)$$

where  $\widehat{\boldsymbol{\beta}} = (\mathbf{F}^\top \widehat{\Sigma}^{-1} \mathbf{F})^{-1} \mathbf{F}^\top \widehat{\Sigma}^{-1} \bar{\mathcal{Y}}$ . The mean squared error ( $\widehat{\text{MSE}}$ ) of the predictor  $\widehat{Y}_M(\mathbf{x}_0)$  is

$$\widehat{\text{MSE}}(\widehat{Y}_M(\mathbf{x}_0)) = \Sigma_M(\mathbf{x}_0, \mathbf{x}_0; \widehat{\boldsymbol{\theta}}) - \widehat{\Sigma}_M(\mathbf{x}_0, \cdot)^\top \widehat{\Sigma}^{-1} \widehat{\Sigma}_M(\mathbf{x}_0, \cdot) + \boldsymbol{\eta}^\top (\mathbf{F}^\top \widehat{\Sigma}^{-1} \mathbf{F})^{-1} \boldsymbol{\eta}, \quad (3)$$

where  $\boldsymbol{\eta} = \mathbf{f}(\mathbf{x}_0) - \mathbf{F}^\top \widehat{\Sigma}^{-1} \widehat{\Sigma}_M(\mathbf{x}_0, \cdot)$ . The last term arises because the  $p \times 1$  vector of regression coefficients needs to be estimated, which inflates the MSE of the prediction; if  $\boldsymbol{\beta}$  was known instead of estimated, the last term of  $\widehat{\text{MSE}}(\widehat{Y}_M(\mathbf{x}_0))$  would drop from the expression.

For stochastic kriging using derivative estimates, see Chen, Ankenman, and Nelson (2013), which provides details on implementation and results. Using derivative estimates, the same results as in Equations (2) and (3) follow, with the proper matrices substituted.

### 3 GENERALIZED INTEGRATED BROWNIAN FIELDS

In stochastic kriging, the response surface is modeled as a sample path of the GRF  $Y_M$ , given by Equation (1), with mean  $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$  and covariance function  $\Sigma_M(\cdot, \cdot; \boldsymbol{\theta})$ . We use GRFs that are built from  $\mathbf{m}$ -order Markov random fields, in an effort to obtain better prediction accuracy. A random field  $M$  is called an  $\mathbf{m}$ -order Markov random field, where  $\mathbf{m} = (m_1, m_2, \dots, m_d)$ , if

$$E[M(\mathbf{x}) | M^{(\alpha)}(\mathbf{z}), \forall \mathbf{z} \in \mathcal{C}^{\ell}, \forall \alpha \leq \mathbf{m}] = E[M(\mathbf{x}) | M^{(\alpha)}(\mathbf{z}), \forall \mathbf{z} \in \partial \mathcal{C}, \forall \alpha \leq \mathbf{m}],$$

for any  $\mathbf{x} \in \mathcal{C}$  and every open set  $\mathcal{C} \subseteq \mathbb{R}^d$  with smooth boundary, where  $\partial\mathcal{C} = \overline{\mathcal{C}} \cap \overline{(\mathcal{C}^c)}$  is the boundary of  $\mathcal{C}$ , and  $M^{(\alpha)}(\mathbf{x}) = \partial^{|\alpha|}M(\mathbf{x})/\partial\mathbf{x}^\alpha$ . For a more precise, measure-theoretic definition, see Pitt (1971).

Furthermore, we require the ability to control the differentiability in each coordinate. A popular choice of GRF corresponds to the so-called Gaussian covariance function  $\Sigma_M(\mathbf{x}, \mathbf{y}; \theta) = \exp\{-\sum_i \theta_i(x_i - y_i)^2\}$ , where  $\theta_i$ ,  $x_i$ , and  $y_i$  are the  $i$ th coordinates of  $\theta$ ,  $\mathbf{x}$ , and  $\mathbf{y}$ , respectively. This GRF has the Markov property of infinite order, i.e., infinitely many derivatives are required at the boundary for conditional independence, and is infinitely differentiable in each coordinate.

GIBFs are generalized versions of integrated Brownian fields, which are multivariate versions of integrated Brownian motions introduced in Fill and Torcaso (2004). These GIBFs have differentiability that can be controlled in each coordinate, and are built from GRFs which have the Markov property. The parameterization of GIBFs that we introduce here is new, and allows for use in high-dimensional metamodeling problems.

### Definition and Properties

ONE-DIMENSIONAL CASE: Consider one-dimensional Brownian motion  $B = (B(x; \theta) : x \in [0, 1])$  with volatility  $\theta$ . Brownian motion is a real-valued, 0-order Gaussian Markov random process with continuous, non-differentiable sample paths, mean function  $E[B(x; \theta)] = 0$ , and covariance function  $\Sigma_B(x, y; \theta) = \theta \min\{x, y\}$  for  $x, y \in [0, 1]$ . An  $m$ -times differentiable stochastic process can be obtained by integrating Brownian motion  $m$  times, which gives us  $m$ -integrated Brownian motion  $B_m(x; \theta)$  with volatility  $\theta$ , an  $m$ -order Gaussian Markov random process. The integral representation of  $m$ -integrated Brownian motion 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  $B_0(x; \theta) = B(x; \theta)$ , and the second equality follows from using integration by parts, which expresses  $B_m$  as an integral with respect to Brownian motion. From the first integral in Equation (4), it is clear that the process  $B_m$  and its  $m$  derivatives  $B_m^{(i)}$  for all  $i = 1, 2, \dots, m$  are zero at the boundary  $x = 0$ . These boundary conditions make  $B_m$  unsuitable for metamodeling, since the response surface and its derivatives may not be zero on the boundary  $x = 0$ . We modify  $B_m$  by adding a random polynomial whose coefficients are  $m + 1$  independent standard normal random variables  $Z_0, Z_1, \dots, Z_m$  scaled by some parameters. This process is denoted by  $X_m(x; \theta)$  and is called generalized  $m$ -integrated Brownian motion and is defined as

$$X_m(x; \theta) = \sum_{n=0}^m \sqrt{\theta_n} Z_n x^n + B_m(x; \theta_{m+1}), \tag{5}$$

where  $\theta$  has been relabelled as  $\theta_{m+1}$  for convenience,  $\theta = (\theta_0, \theta_1, \dots, \theta_{m+1})$ , and  $B_m$  is independent of  $Z_n$  for all  $n = 1, 2, \dots, m$ . Directly from the definitions of  $B_m(x; \theta)$  and  $X_m(x; \theta)$ , it follows that the covariance function of  $X_m(x; \theta)$  is

$$\Sigma_{X_m}(x, y; \theta) = \sum_{k=0}^m \theta_k \frac{x^k y^k}{(k!)^2} + \theta_{m+1} \int_0^1 \frac{(x-u)_+^m (y-u)_+^m}{(m!)^2} du.$$

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 with coefficients that are monomials of degree at most  $m$ . Thus, the random polynomial is an  $m$ -order Gaussian Markov random process. By specifying  $m$ , we are able to control the differentiability of the GIBF.

MULTI-DIMENSIONAL CASE: In the multi-dimensional case, let  $\mathbf{B} = (\mathbf{B}(\mathbf{x}; \theta) : \mathbf{x} \in [0, 1]^d)$  be  $d$ -dimensional Brownian field (see, for example, Holden, Øksendal, Ubøe, and Zhang (2010)) with volatility

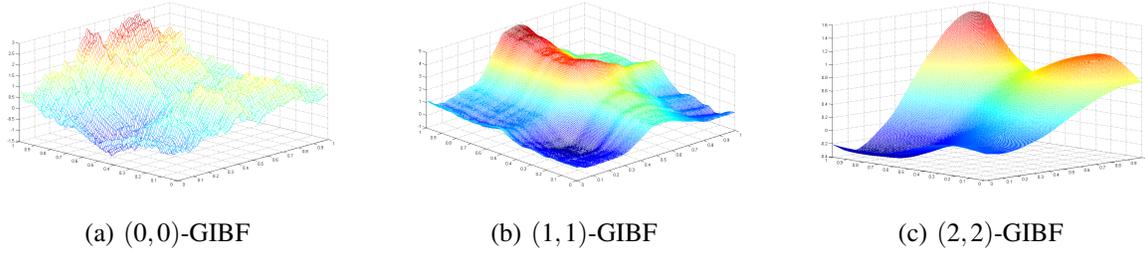


Figure 1: Sample paths of GIBFs (on the two-dimensional unit hypercube).

$\theta$ , where  $\theta = (\theta_1, \theta_2, \dots, \theta_d)$ . Brownian field is a real-valued  $\mathbf{0}$ -order Gaussian Markov random field (GMRF) with continuous, non-differentiable sample paths, mean function  $E[\mathbf{B}(\mathbf{x}; \theta)] = 0$ , and covariance function  $\Sigma_{\mathbf{B}}(\mathbf{x}, \mathbf{y}; \theta) = \prod_{i=1}^d \theta_i \min\{x_i, y_i\}$  for  $\mathbf{x}, \mathbf{y} \in [0, 1]^d$ , which is the tensor product of  $d$  copies of one-dimensional Brownian motions with varying volatilities. Similar to the one-dimensional case, we can integrate Brownian field over each coordinate to get a differentiable process. In the multi-dimensional case, each coordinate can be integrated a different number of times. If we integrate  $m_i$  times in the  $i$ th coordinate, for all  $i = 1, 2, \dots, d$ , the resulting GMRF is called  $\mathbf{m}$ -integrated Brownian field  $\mathbf{B}_{\mathbf{m}}$ , where  $\mathbf{m} = (m_1, m_2, \dots, m_d)$ , which is an  $\mathbf{m}$ -order GMRF. Using integration by parts,  $\mathbf{B}_{\mathbf{m}}$  can be expressed as a multiple integral with respect to Brownian field,

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

It follows immediately from this representation that the covariance function of  $\mathbf{B}_{\mathbf{m}}$  is

$$\Sigma_{\mathbf{B}_{\mathbf{m}}}(\mathbf{x}, \mathbf{y}; \theta) = \prod_{i=1}^d \theta_i \int_0^1 \frac{(x_i - u_i)_+^{m_i} (y_i - u_i)_+^{m_i}}{(m_i!)^2} du_i.$$

Note that  $\Sigma_{\mathbf{B}_{\mathbf{m}}}$  is the tensor product of the covariance functions of the one-dimensional processes  $B_{m_1}, B_{m_2}, \dots, B_{m_d}$ . As in the one-dimensional case,  $\mathbf{m}$ -integrated Brownian field has boundary conditions  $\mathbf{B}_{\mathbf{m}}(\mathbf{x}; \theta) = 0$  and  $\mathbf{B}_{\mathbf{m}}^{\mathbf{n}}(\mathbf{x}; \theta) = 0$  for all  $\mathbf{n}$  such that  $|\mathbf{n}| \leq |\mathbf{m}|$  and  $\mathbf{x} \in \{\mathbf{y} \in [0, 1]^d : \exists i, y_i = 0\}$ . We define a new process  $\mathbf{X}_{\mathbf{m}}$  whose covariance function is the tensor product of the covariance functions of the one dimensional processes  $X_{m_1}, X_{m_2}, \dots, X_{m_d}$ , given by Equation (5), in the same way that the covariance function of  $\mathbf{B}_{\mathbf{m}}$  is the tensor product of the covariance functions of  $B_{m_1}, B_{m_2}, \dots, B_{m_d}$ . The process  $\mathbf{X}_{\mathbf{m}}(\mathbf{x}; \theta)$  has mean function  $E[\mathbf{X}_{\mathbf{m}}(\mathbf{x}; \theta)] = 0$  and covariance function

$$\Sigma_{\mathbf{X}_{\mathbf{m}}}(\mathbf{x}, \mathbf{y}; \theta) = \prod_{i=1}^d \left( \sum_{k=0}^{m_i} \theta_{i,k} \frac{x_i^k y_i^k}{(k!)^2} + \theta_{i, m_i+1} \int_0^1 \frac{(x_i - u_i)_+^{m_i} (y_i - u_i)_+^{m_i}}{(m_i!)^2} du_i \right), \quad (6)$$

where  $\theta = (\theta_{1,0}, \dots, \theta_{1, m_1+1}, \theta_{2,0}, \dots, \theta_{d, m_d+1})$ .

The GRF  $\mathbf{X}_{\mathbf{m}}$  is called generalized  $\mathbf{m}$ -integrated Brownian field ( $\mathbf{m}$ -GIBF) on  $[0, 1]^d$ . Figure 1 shows sample paths of (0,0)-GIBF, (1,1)-GIBF, and (2,2)-GIBF. We have defined  $\mathbf{X}_{\mathbf{m}}$  in terms of its mean and covariance functions which is all that is required to fully define a GRF. To get a better understanding of  $\mathbf{m}$ -GIBF, an equivalent formulation of  $\mathbf{X}_{\mathbf{m}}$  is given by

$$\mathbf{X}_{\mathbf{m}}(\mathbf{x}) := \sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{m}} C_{\mathbf{n}}(\theta) \mathbf{x}^{\mathbf{n}} Z_{\mathbf{n}} + \sum_{\substack{1 \leq i \leq d \\ 1 \leq j_1 < j_2 < \dots < j_i \leq d}} \sum_{\substack{0 \leq k_j \leq m_j \\ j \neq j_i, \forall l}} C_{\mathbf{j}_i, \mathbf{k}_{d-i}}(\mathbf{x}, \theta) \mathbf{B}_{(m_{j_1}, \dots, m_{j_i})}^{\mathbf{j}_i, \mathbf{k}_{d-i}}(x_{j_1}, \dots, x_{j_i}; \mathbf{1}), \quad (7)$$

where  $\mathbf{j}_i = \{j_1, j_2, \dots, j_i\}$ ,  $\mathbf{k}_{d-i} = \{k_j : j \neq j_l, \forall l\}$ , and the multidimensional sum is over all  $\mathbf{n} = (n_1, n_2, \dots, n_d)$  such that  $\mathbf{0} \leq \mathbf{n} \leq \mathbf{m}$ . The functions  $C_{\mathbf{n}}(\theta)$  and  $C_{\mathbf{j}_i, \mathbf{k}_{d-i}}(\mathbf{x}, \theta)$  are deterministic functions of  $\mathbf{x}$  and  $\theta$ , 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, so they are omitted. Equation (7) is the multi-dimensional analog of Equation (5). The first term in Equation (7) 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  $\mathbf{m}$ . Thus, the random polynomial is an  $\mathbf{m}$ -order GMRF. The second term is the sum of integrated Brownian fields over every  $i$ -face (for all  $i = 1, 2, \dots, d$ ) of the unit hypercube that contains the origin. In other words, we sum integrated Brownian fields over each edge, face, cell, 4-face, 5-face, etc. of the unit hypercube that contains the origin. Since the functions  $C_{\mathbf{n}}(\theta)$  and  $C_{\mathbf{j}_i, \mathbf{k}_{d-i}}(\mathbf{x}, \theta)$  are deterministic functions of  $\mathbf{x}$  and  $\theta$ , the randomness in  $\mathbf{X}_{\mathbf{m}}$  is due to the standard normal random variables  $Z_{\mathbf{n}}$  and the integrated Brownian fields  $\mathbf{B}_{(m_{j_1}, \dots, m_{j_i})}^{\mathbf{j}_i, \mathbf{k}_{d-i}}$ , which are all independent from each other. From the formulation of  $\mathbf{X}_{\mathbf{m}}$  given by Equation (7), it is clear that  $\mathbf{m}$ -GIBF does not have any boundary conditions. Furthermore, we are able to control the differentiability in each coordinate by specifying each entry of the vector  $\mathbf{m} = (m_1, m_2, \dots, m_d)$ .

*Remark:* The covariance function of  $\mathbf{X}_{\mathbf{m}}$  is the reproducing kernel of the tensor product Hilbert space  $\mathcal{H} = \otimes_{i=1}^d H^{m_i}(0, 1)$ , where  $H^{m_i}(0, 1) = \{\phi : \phi, \phi^{(1)}, \dots, \phi^{(m_i-1)} \text{ absolutely continuous, } \phi^{(m_i)} \in \mathbb{L}^2[0, 1]\}$ , endowed with the norm  $\|\phi\|^2 = \sum_{k=0}^{m_i-1} \theta_k \phi^{(k)}(0)^2 + \theta_{i, m_i+1} \int_0^1 \phi^{(m_i)}(t)^2 dt$ . Therefore, the same results can be obtained by considering multi-dimensional tensor product smoothing splines on the space  $\mathcal{H}$  (see Berlinet and Thomas-Agnan (2004) for more information).

#### 4 STOCHASTIC KRIGING WITH GENERALIZED INTEGRATED BROWNIAN FIELDS

For stochastic kriging with  $\mathbf{m}$ -GIBF, the response surface is modeled as a sample path of the GMRF

$$Y_{\mathbf{x}_{\mathbf{m}}}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \beta + \widehat{\mathbf{X}}_{\mathbf{m}}(\mathbf{x}), \quad (8)$$

where  $\mathbf{f}(\mathbf{x})$  and  $\beta$  are as before, and  $\widehat{\mathbf{X}}_{\mathbf{m}}$  is a modified version of  $\mathbf{m}$ -GIBF, which is discussed in Section 4.1. To implement stochastic kriging with GIBFs, we need to choose the vector of basis functions  $\mathbf{f}$  to be used for trend modeling, and values for the parameters  $\mathbf{m}$ ,  $\beta$ , and  $\theta$ . This section discusses these aspects of fitting GIBFs: we discuss trend modeling in Section 4.1, followed by the choice of order of the GIBF in Section 4.2. Maximum likelihood estimation of the parameters is presented in Section 4.3, assuming that the vector of basis functions and order have been fixed. The properties of metamodels built using  $\mathbf{m}$ -GIBF are given in Section 4.4.

##### 4.1 Trend Modeling

To maintain the differentiability of the metamodel, we assume that each basis function in the  $p \times 1$  vector of basis functions  $\mathbf{f}$  is  $m_i$  times continuously differentiable in the  $i$ th coordinate. Any function can be a basis function as long as it satisfies this differentiability condition.

For certain basis functions, the covariance function needs to be modified: for stochastic kriging with  $\mathbf{m}$ -GIBF, when a basis function is the monomial  $\mathbf{x}^\alpha$ , where  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)^\top$  and  $\alpha_i \leq m_i$  for all  $i = 1, 2, \dots, d$ , 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 (6). The need for this modification of the covariance function is the following: for stochastic kriging with the GRF  $Y_{\mathbf{M}}$ , the residuals  $\mathcal{Y} - \mathbf{F}\beta$  are modeled as a sample path of the mean-zero GRF  $\mathbf{M}$ , plus Monte Carlo noise. When  $\mathbf{x}^\alpha$  is a basis function, the deviation of the residuals from the mean of  $\mathbf{M}$  (zero) is eliminated in the subspace spanned by  $\mathbf{x}^\alpha$ . To avoid redundancy when we use  $\mathbf{m}$ -GIBF, we need to subtract from  $\mathbf{m}$ -GIBF the variability associated with the subspace spanned by  $\mathbf{x}^\alpha$ , and this variability results from the GMRF whose covariance function is  $\prod_{i=1}^d \theta_{i, \alpha_i} x_i^{\alpha_i} y_i^{\alpha_i} / (\alpha_i!)^2$ , i.e., the GMRF  $C_\alpha(\theta) \mathbf{x}^\alpha Z_\alpha$  in the random polynomial in Equation (7).

Another explanation for the modification of the covariance function can be given in terms of boundary conditions. The formulation of  $\mathbf{m}$ -GIBF given by Equation (7) is the sum of  $\mathbf{B}_{\mathbf{m}}$  and other terms that

compensate for the boundary conditions of  $\mathbf{B}_m$ . When the basis function  $\mathbf{x}^\alpha$  is included in  $\mathbf{f}$ , we do not need the term  $C_\alpha(\theta)\mathbf{x}^\alpha Z_\alpha$  in the random polynomial  $\sum_{n=0}^m C_n(\theta)\mathbf{x}^n Z_n$ , since the corresponding boundary condition is being compensated for in the trend function. In other words, since the response surface is modeled as a sample path of  $Y_{\mathbf{X}_m}$ , we want  $Y_{\mathbf{X}_m}$  to have no boundary conditions, and when  $\mathbf{x}^\alpha$  is a basis function for the trend, we no longer need the corresponding boundary condition compensated for in  $\mathbf{B}_m$ . For example, consider the process  $Y_{\mathbf{X}_m}(\mathbf{x})$  with  $\mathbf{f} = (1)$ , i.e.,  $Y_{\mathbf{X}_m}(\mathbf{x}) = \beta + \widehat{\mathbf{X}}_m(\mathbf{x})$ , where  $\widehat{\mathbf{X}}_m(\mathbf{x}) = \mathbf{X}_m(\mathbf{x}) - C_0(\theta)Z_0$ . Although  $\widehat{\mathbf{X}}_m$  has the boundary condition  $\widehat{\mathbf{X}}_m(\mathbf{0}) = 0$ ,  $Y_{\mathbf{X}_m}$  has no boundary conditions since the constant trend compensates for the boundary condition of  $\widehat{\mathbf{X}}_m$  at the origin, i.e.,  $Y_{\mathbf{X}_m}(\mathbf{0}) = \beta$ .

The random field  $\widehat{\mathbf{X}}_m$  is the (mean-zero) GMRF whose covariance function is the modified covariance function of  $\mathbf{m}$ -GIBF. For example, if we use stochastic kriging with (1,1)-GIBF and the vector of basis functions  $\mathbf{f}(\mathbf{x}) = (1, x_1)^\top$  used for trend modeling, the response surface would be modeled as the sample path of the GRF  $Y_{(1,1)}(\mathbf{x}) = (1, x_1)^\top \beta + \widehat{\mathbf{X}}_{(1,1)}(\mathbf{x})$ , where

$$\Sigma_{\widehat{\mathbf{X}}_{(1,1)}}(\mathbf{x}, \mathbf{y}; \theta) = \prod_{i=1}^2 \left( \theta_{i,0} + \theta_{i,1}x_i y_i + \theta_{i,2} \int_0^1 (x_i - u_i)_+(y_i - u_i)_+ du_i \right) - \theta_{1,0}\theta_{2,0} - \theta_{2,0}\theta_{1,1}x_1 y_1$$

is the covariance function of the GMRF  $\widehat{\mathbf{X}}_{(1,1)} = \mathbf{X}_{(1,1)} - C_{0,0}(\theta)Z_{0,0} - C_{1,0}(\theta)x_1 Z_{1,0}$ . Thus, using the vector of basis functions  $\mathbf{f}(\mathbf{x}) = (1, x_1)^\top$  results in the loss of the random polynomial terms corresponding to the monomials 1 and  $x_1$ .

## 4.2 Choosing Order

The parameters for stochastic kriging with GIBFs are  $(\mathbf{m}, \beta, \theta)$ , whose optimal values are given by the maximum likelihood estimates  $(\widehat{\mathbf{m}}, \widehat{\beta}, \widehat{\theta})$ . Thus, the optimal order would be given by its maximum likelihood estimate  $\widehat{\mathbf{m}}$ . However, finding the maximum likelihood estimates  $(\widehat{\mathbf{m}}, \widehat{\beta}, \widehat{\theta})$  involves solving an integer programming problem, since  $\mathbf{m}$  must be a vector of integers. To simplify the optimization, we use a heuristic, stepwise procedure that considers each coordinate one at a time, and increases the order of the GIBF in a coordinate if the log-likelihood increases. Let  $D_j$ , for all  $j = 1, 2, \dots, d$ , be the highest order derivative estimate obtained from the simulation output for the  $j$ th coordinate. We assume that every derivative estimate obtained from running the simulation model is to be used in building the metamodel. The procedure is given by the following:

1. Initialize  $i = 1$ ,  $j = 0$ , and  $m_k = D_k$  for all  $k = 1, 2, \dots, d$ .
2. Fit  $(m_1, m_2, \dots, m_d)$ -GIBF using the procedure given in Section 4.3. Denote the log-likelihood by  $\mathcal{L}_j$ . Set  $m_i = m_i + 1$ , and go to 3.
3. Fit  $(m_1, m_2, \dots, m_d)$ -GIBF using the procedure given in Section 4.3. Denote the log-likelihood by  $\mathcal{L}_{j+1}$ . If  $\mathcal{L}_j < \mathcal{L}_{j+1}$ , set  $j = j + 1$ ,  $m_i = m_i + 1$  and go to 3. Otherwise, set  $m_i = D_i + j$ ,  $i = i + 1$ ,  $j = 0$ , and go to 2.

## 4.3 Parameter Estimation

Assuming that the trend vector and order of the GIBF have been fixed, the parameters  $\beta$  and  $\theta$  are found using maximum likelihood estimation. For stochastic kriging with derivative estimates, parameter fitting is exactly the same as below, with the proper matrices substituted. Given fixed values for  $\theta$ , the maximum likelihood estimate of  $\beta$  is  $\widehat{\beta}(\theta) := (\mathbf{F}^\top \Sigma(\theta)^{-1} \mathbf{F})^{-1} \mathbf{F}^\top \Sigma(\theta)^{-1} \mathcal{Y}$ , where  $\Sigma(\theta) = \Sigma_{\widehat{\mathbf{X}}_m}(\theta) + \widehat{\Sigma}_\varepsilon$ ,  $\Sigma_{\widehat{\mathbf{X}}_m}(\theta)$  is the  $k \times k$  variance-covariance matrix with  $ij$ th entry  $\Sigma_{\widehat{\mathbf{X}}_m}(\mathbf{x}_i, \mathbf{x}_j; \theta)$ , and  $\widehat{\beta}$  and  $\Sigma$  have been written as functions of  $\theta$  to explicitly show dependence. If we profile over the MLE of  $\beta$ , the profile log-likelihood function (see, for example, Shao (2010)) is given by

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

where  $k$  is the number of design points, and  $\mathcal{Y}$  is the vector of simulation output. The maximum likelihood estimate of  $\theta$  is the solution to the constrained minimization problem  $\min \{-\mathcal{L}(\theta) | \theta \in \mathbb{R}_+^d\}$ , where  $\mathbb{R}_+^d$  is the set of feasible values of  $\theta$  for **m**-GIBF. To solve this constrained optimization problem, we first search for the best starting solution by evaluating  $-\mathcal{L}(\theta)$  at a low-discrepancy sequence of points, and take the point in the sequence that minimizes this quantity as the starting solution. Instead of searching over  $\mathbb{R}_+^d$ , which is unbounded, we add a dummy parameter  $\tau$  which allows  $\theta$  to be restricted to the unit hypercube  $[0, 1]^d$ . In other words, only the magnitude of the parameters in  $\theta$  relative to each other are important since the actual magnitude is absorbed in  $\tau$ . The re-parameterized covariance function for **m**-GIBF is

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

where now  $\theta \in [0, 1]^d$ , and  $\tau \geq 0$ . The maximum likelihood estimate of  $\theta$  is now found by solving the constrained optimization problem  $\min \{-\mathcal{L}(\theta, \tau^*(\theta)) | \theta \in [0, 1]^d\}$ , where

$$\mathcal{L}(\theta, \tau) := -\frac{k}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma(\theta, \tau)|) - \frac{1}{2} (\mathcal{Y} - \mathbf{F}\hat{\beta}(\theta, \tau))^\top \Sigma(\theta, \tau)^{-1} (\mathcal{Y} - \mathbf{F}\hat{\beta}(\theta, \tau)) \quad (9)$$

is the re-parameterized profile log-likelihood function,  $\hat{\beta}$  and  $\Sigma$  have been written as functions of  $\theta$  and  $\tau$  to explicitly show dependence, and  $\tau^*(\theta)$  is the value of  $\tau$  that minimizes  $\mathcal{L}(\theta, \tau)$  with  $\theta$  fixed. Finding  $\tau^*(\theta)$  can be done very efficiently using a line search method, and supplying the solver with the gradient  $\partial \mathcal{L}(\theta, \tau) / \partial \tau$ , which can be easily computed using matrix calculus.

The solution to the constrained minimization problem  $\min \{-\mathcal{L}(\theta, \tau^*(\theta)) | \theta \in [0, 1]^d\}$  is obtained using the following procedure based on the R package **mlepp** (Dancik 2012).

1. Generate  $m$  points  $\{\mathbf{p}_i\}_{i=1}^m$  from a low-discrepancy sequence.
2. Evaluate  $\mathcal{L}(\theta, \tau^*(\theta))$  at each point in the low discrepancy sequence  $\{\mathbf{p}_i\}_{i=1}^m$ . Let  $\mathbf{p}^*$  be the point in  $\{\mathbf{p}_i\}_{i=1}^m$  that minimizes  $\mathcal{L}(\theta, \tau^*(\theta))$ .
3. Use the Nelder-Mead method (Nelder and Mead 1965) on  $\mathcal{L}(\theta, \tau^*(\theta))$ , with the starting solution  $\mathbf{p}^*$ . Let the optimal solution returned from the Nelder-Mead method be denoted by  $\theta'$ .
4. Use the Limited-Memory BFGS method (Liu and Nocedal 1989) on  $\mathcal{L}(\theta, \tau^*(\theta))$ , with the starting solution  $\theta'$ . The optimal solution is the MLE  $\hat{\theta}$ .

Roughly speaking, we search the unit hypercube for the best starting point  $\mathbf{p}^*$ , for which we then use the Nelder-Mead and Limited-Memory BFGS methods to find the MLE  $\hat{\theta}$ .

#### 4.4 Properties

The properties of metamodels built using stochastic kriging with **m**-GIBF can be analyzed by viewing the stochastic kriging predictor as a linear combination of the  $k$  basis functions  $\Sigma_{\hat{\mathbf{X}}_m}(\cdot, \mathbf{x}_i; \hat{\theta})$ , for all  $i = 1, 2, \dots, k$ . Indeed, we can rearrange Equation (2) for  $\mathbf{x} \in \mathbb{X}$  into the following form:

$$\hat{\mathbf{Y}}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \hat{\beta} + \sum_{i=1}^k c_i \Sigma_{\hat{\mathbf{X}}_m}(\mathbf{x}, \mathbf{x}_i; \hat{\theta}),$$

where  $\mathbf{c} = \hat{\Sigma}^{-1}(\mathcal{Y} - \mathbf{F}\hat{\beta})$ . The only terms that depend on  $\mathbf{x}$  in this expression are  $\mathbf{f}$  and  $\Sigma_{\hat{\mathbf{X}}_m}$ , so the differentiability of the metamodel is determined by  $\mathbf{f}$  and  $\Sigma_{\hat{\mathbf{X}}_m}$ . Since we are assuming that each function in the trend vector is  $m_i$  times differentiable in the  $i$ th coordinate, and the fact that the covariance function  $\Sigma_{\hat{\mathbf{X}}_m}(\mathbf{x}, \mathbf{x}_i; \hat{\theta})$  is  $m_i$  times differentiable in the  $i$ th coordinate, the metamodel is  $m_i$  times differentiable in the  $i$ th coordinate.

The invariance properties of the metamodel are determined by the properties of GIBFs. The metamodel is not invariant to rotations of the design space, since GIBFs are non-stationary. However, since we are

assuming that the design space is the  $d$ -dimensional unit hypercube, which can be obtained from scaling and shifting the original design space, the metamodel is translation invariant. Furthermore, from directly inspecting the covariance function in Equation (6), we can see that the metamodel is invariant to permutations of the coordinates.

## 5 EXPERIMENTS

The purpose of the experiments is to assess the performance, in terms of prediction accuracy, of stochastic kriging with GIBFs. We are mainly concerned with the affect on performance of different types of response surfaces and whether or not we incorporate derivative information. We use two examples, and compare our results to stochastic kriging with the GRF corresponding to the Gaussian covariance function. The GRF corresponding to the Gaussian covariance function can result in metamodels that have mean reversion. Since we are interested in comparing to the Gaussian covariance function when mean reversion is not a factor, we use an experiment design that results in no mean reversion.

Each example is two-dimensional. The design points are the first 50 points of a scrambled Sobol sequence, and the prediction points  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{1,000}$  are 1,000 points uniformly sampled from the unit hypercube  $[0, 1]^2$ , re-scaled to fit inside the hypercube  $[0.05, 0.95]^2$ . We re-scale the prediction points to avoid boundary effects, since stochastic kriging (and interpolation and smoothing methods in general) is known not to predict well near the boundary of the design space. We repeat the experiment 50 times to get 50 metamodels for each response surface. We evaluate the predictions using the Root Empirical Mean Squared Error

$$\text{REMSE} = \sqrt{\frac{1}{50,000} \sum_{j=1}^{50} \sum_{i=1}^{1,000} (\hat{y}_j(\mathbf{p}_i) - y(\mathbf{p}_i))^2},$$

where  $\hat{y}_j$  is the metamodel built during the  $j$ th experiment. A constant trend is used in all of the experiments, and the estimates of the gradients are based on infinitesimal perturbation analysis (IPA). The two response surfaces are described below.

### 5.1 Multi-factor Portfolio Credit Risk

This example is taken from Glasserman, Kang, and Shahabuddin (2008). We are interested in estimating the expected credit loss of a portfolio, given values of latent variables that trigger the default of the obligors. Consider a portfolio with  $m$  obligors, and let  $Y_k$  be the default indicator ( $= 1$  for default,  $= 0$  otherwise) for the  $k$ th obligor,  $p_k$  be the marginal probability that the  $k$ th obligor defaults, and  $l_k$  be the loss resulting from default of the  $k$ th obligor. The dependence among the default indicators  $Y_k$  is modeled by a multifactor Gaussian copula model with a finite number of types, i.e.  $Y_k = \mathbf{1}\{X_k > \Phi^{-1}(1 - p_k)\}$ , where  $\Phi$  is the cumulative normal distribution, and  $X_1, X_2, \dots$  are correlated standard normal random variables. To model the correlation of the standard normal random variables, we assume that there are  $d$  factors and  $t$  types of obligors. If obligor  $k$  is of type  $j$ , then the latent variable is given by  $X_k = \mathbf{a}_j^\top \mathbf{Z} + b_j \varepsilon_k$ , where  $\mathbf{a}_j \in \mathbb{R}^d$ ,  $\mathbf{Z}$  is a  $d$ -dimensional standard normal random vector,  $b_j = \sqrt{1 - \mathbf{a}_j^\top \mathbf{a}_j}$ , and the  $\varepsilon_k$  are independent standard normal random variables. The total loss from defaults is  $L_m = \sum_{i=1}^m l_i Y_i$ .

For our experiments, 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  $\mathbf{a}_4^\top = (0, 0.25)$ . Each type has the same number of obligors, with  $E[l_k] = 1$  and  $p_k = 0.01$ . The response surface, which is twice-differentiable everywhere for  $\mathbf{x} \in [-10, 30]^2$ , is

$$y_{cr}(\mathbf{x}) = E[L_m | \mathbf{Z} = \mathbf{x}] = \frac{1}{4} \sum_{j=1}^t \Phi \left( \frac{\mathbf{a}_j^\top \mathbf{x} + \Phi^{-1}(0.01)}{\sqrt{1 - \mathbf{a}_j^\top \mathbf{a}_j}} \right).$$



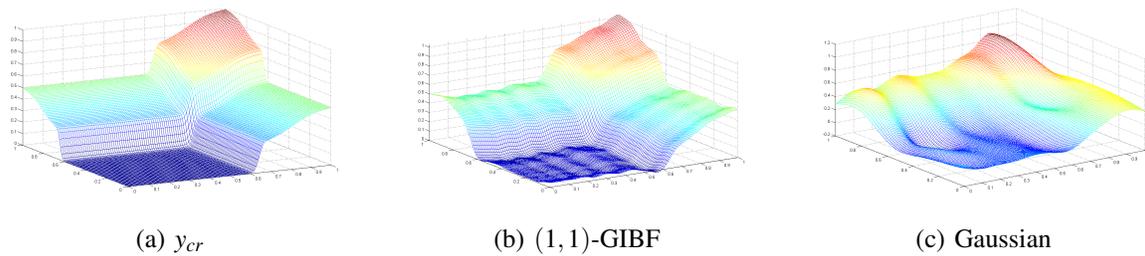


Figure 2: The credit risk response surface, and fitted metamodels using noiseless data with (1,1)-GIBF and the Gaussian covariance function.

characterizes and takes into account the intrinsic error inherent in a stochastic simulation, we showed how to implement GIBFs, and used two examples to assess the performance of stochastic kriging with GIBFs. These examples exhibited the benefit gained from using stochastic kriging with GIBFs instead of the GRF associated with the Gaussian covariance function. The examples also showed the improvement in performance when gradient estimates were included in the prediction. Additional properties of GIBFs require further investigation, as well as additional numerical experimentation to assess for which response surfaces GIBFs work best.

## ACKNOWLEDGEMENTS

This article is based upon work supported by the National Science Foundation under Grant No. CMMI-0900354. The authors would also like to thank the two anonymous referees for helpful comments and corrections.

## REFERENCES

- Ankenman, B., B. L. Nelson, and J. Staum. 2010. “Stochastic Kriging for Simulation Metamodeling”. *Operations Research* 58 (2): 371–382.
- Aydin, G., and E. L. Porteus. 2008. “Joint Inventory and Pricing Decisions for an Assortment”. *Operations Research* 56 (5): 1247–1255.
- Berlinet, A., and C. Thomas-Agnan. 2004. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, New York, NY.
- Chen, X., B. Ankenman, and B. L. Nelson. 2013. “Enhancing Stochastic Kriging Metamodels with Gradient Estimators”. *Operations Research*. in press.
- Dancik, G. M. 2012. “Maximum Likelihood Estimation of Gaussian Processes”. R Package Documentation.
- Fill, J. A., and F. Torcaso. 2004. “Asymptotic Analysis via Mellin Transforms for Small Deviations in  $L_2$ -norm of Integrated Brownian Sheets”. *Probability Theory and Related Fields* 130 (2): 259–288.
- Glasserman, P., W. Kang, and P. Shahabuddin. 2008. “Fast Simulation of Multifactor Portfolio Credit Risk”. *Operations Research* 56 (5): 1200–1217.
- Holden, H., B. Øksendal, J. Ubøe, and T. Zhang. 2010. *Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach*. Second ed. Springer Science and Business Media (Universitext), New York, NY.
- Krige, D. G. 1951. “A Statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand”. Master’s thesis, University of Witwatersrand.
- Liu, D. C., and J. Nocedal. 1989. “On the Limited Memory BFGS Method for Large Scale Optimization”. *Mathematical Programming* 45:503–528.
- Mitchell, T. J., and M. D. Morris. 1992. “The Spatial Correlation Function Approach to Response Surface Estimation”. In *Proceedings of the 1992 Winter Simulation Conference*, edited by J. J. Swain,

- D. Goldsman, R. C. Crain, and J. R. Wilson, 565–571. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Nelder, J. A., and R. Mead. 1965. “A Simplex Method for Function Minimization”. *Computer Journal* 7:308–313.
- Pitt, L. D. 1971. “A Markov Property for Gaussian Processes with a Multidimensional Parameter”. *Archive for Rational Mechanics and Analysis* 43 (5): 367–391.
- Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn. 1989. “Design and Analysis of Computer Experiments”. *Statistical Science* 4 (4): 409–423.
- Santner, T. J., B. J. Williams, and W. I. Notz. 2010. *The Design and Analysis of Computer Experiments*. Springer-Verlag New York, LLC.
- Shao, J. 2010. *Mathematical Statistics*. Springer Texts in Statistics, New York, NY.

### AUTHOR BIOGRAPHIES

**PETER SALEMI** is a Ph.D. candidate in the Department of Industrial Engineering and Management Sciences at Northwestern University. His research interests are in simulation metamodeling and simulation optimization. His email address is [petersalemi2014@u.northwestern.edu](mailto:petersalemi2014@u.northwestern.edu).

**JEREMY STAUM** is an Associate Professor of Industrial Engineering and Management Sciences at Northwestern University. He coordinated the Risk Analysis track of the 2007 and 2011 Winter Simulation Conferences and serves as department editor for financial engineering at *IIE Transactions* and as an associate editor at *Management Science*. His email and web addresses are [j-staum@northwestern.edu](mailto:j-staum@northwestern.edu) and [users.iems.northwestern.edu/~staum](http://users.iems.northwestern.edu/~staum).

**BARRY L. NELSON** is the Walter P. Murphy Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University, and a Fellow of INFORMS. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. His email and web addresses are [nelsonb@northwestern.edu](mailto:nelsonb@northwestern.edu) and [www.iems.northwestern.edu/~nelsonb/](http://www.iems.northwestern.edu/~nelsonb/).