Brownian Integrated Covariance Functions for Gaussian Process Modeling: Sigmoidal Versus Localized Basis Functions

Ning Zhang and Daniel W. Apley[1]

Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208, USA

ABSTRACT

Gaussian process modeling, or kriging, is a popular method for modeling data from deterministic computer simulations, and the most common choices of covariance function are Gaussian, power exponential, and Matérn. A characteristic of these covariance functions is that the basis functions associated with their corresponding response predictors are localized, in the sense that they decay to zero as the input location moves away from the simulated input sites. As a result, the predictors tend to revert to the prior mean, which can result in a bumpy fitted response surface. In contrast, a fractional Brownian field model results in a predictor with basis functions that are nonlocalized and more sigmoidal in shape, although it suffers from drawbacks such as inability to represent smooth response surfaces. We propose a class of Brownian integrated covariance functions obtained by incorporating an integrator (as in the white noise integral representation of a fractional Brownian field) into any stationary covariance function. Brownian integrated covariance models result in predictor basis functions that are nonlocalized and sigmoidal, but they are capable of modeling smooth response surfaces. We discuss fundamental differences between Brownian integrated and other covariance functions, and we illustrate by comparing Brownian integrated power exponential with regular power exponential kriging models in a number of examples.

KEY WORDS: Kriging, Response surface modeling, Metamodeling, Gaussian random fields, Fractional Brownian field

[1] Ning Zhang is PhD student (E-mail: ningzhang@u.northwestern.edu) and Daniel W. Apley (corresponding author) is a Professor (apley@northwestern.edu), Department of Industrial Engineering & Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208. This work is supported by xxxxxx. The authors thank xxx and xxx for xxx.

1. INTRODUCTION

Computer simulation is widely used in many branches of science and engineering and is still gaining momentum in terms of application areas. The fidelity and complexity of computer simulations have also advanced significantly, which renders many simulations still very time-consuming to run in spite of advances in computing power. To deal with this, metamodels (or emulators) are often used to approximate the simulation response surface based on a finite number of evaluations, which might be constrained by the simulation budget. Introduced for this purpose by Currin et al. (1988) and Sacks, Welch, Mitchell and Wynn (1989), the kriging (aka random field regression or Gaussian process modeling) approach is a useful and popular method in practice that has found its way into many textbooks on the topic of metamodeling the output of computer simulations (Santner, Williams and Notz 2003; Fang, Li and Sudjianto 2006).

Suppose we select a set of n distinct input sites $\mathbf{S} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and evaluate the scalar simulation response function $y(\mathbf{x})$ at these input sites, where $\mathbf{x} = [x_1, x_2, \ldots, x_d]^T$ denotes the set of d input variables. Let $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T = [y(\mathbf{x}_1), y(\mathbf{x}_2), \ldots, y(\mathbf{x}_n)]^T$ denote the corresponding n response values. Kriging models the response surface $y(\mathbf{x})$ as a realization of a stochastic random field $Y(\mathbf{x})$ over the d-dimensional input variable space. The most commonly used random fields are Gaussian random fields (GRFs), with constant mean function $E[Y(\mathbf{x})] = \mu$ and covariance function $R(\mathbf{x}, \mathbf{x}') = Cov[Y(\mathbf{x}), Y(\mathbf{x}')]$. A widely used covariance function is the power exponential (PEXP) covariance

$$R(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{i=1}^{d} exp\{-|\phi_i(x_i - x'_i)|^{\nu}\}, \quad (1)$$

where σ^2 is the prior variance of $Y(\bullet)$, and $\mathbf{\Phi} = \text{diag}\{\phi_1, \phi_2, \ldots, \phi_d\}$ and ν ($1 \leq \nu \leq 2$) are the lengthscale and exponent parameters that reflect the smoothness of the GRF. When $\nu = 2$, the covariance structure in (1) is called a Gaussian covariance. With a constant mean, the approach is referred to as *ordinary kriging*; whereas for a nonconstant mean that is a function of $\mathbf{x}$ (typically a polynomial with unknown coefficients), the approach is referred to as *universal kriging*. Many researchers believe there is little need to include a non-constant mean function in standard kriging models (Welch et al. 1992), and this is especially the case for the covariance

model that we propose in this paper because of its nonlocalized nature (see Section 4.1). Consequently, we consider only a constant mean.

The kriging predictor $\hat{y}(\mathbf{x})$ of $y(\mathbf{x})$ is defined as the best linear unbiased predictor of $Y(\mathbf{x})$ given $\mathbf{y}$, or equivalently, the posterior mean predictor in an empirical Bayesian framework (with the MLEs of the lengthscale parameters viewed as fixed), which is (Stein 1999)

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}). \quad (2)$$

Here, $\mathbf{R}$ is an $n \times n$ matrix whose ith row, jth column element is $R(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{r}(\mathbf{x})$ is an $n \times 1$ vector whose ith element is $R(\mathbf{x}, \mathbf{x}_i)$, $\mathbf{1}$ is a n-length column vector of ones, and $\hat{\mu} = [\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}]^{-1}\mathbf{1}^T\mathbf{R}^{-1}\mathbf{y}$.

Although GRFs with Gaussian, PEXP, and Matérn covariance are the most common choice of kriging model for computer simulations (Santner, Williams and Notz 2003), they have a characteristic that can be undesirable, depending on the application. Namely, the fitted response surface can have a bumpy appearance because the predictor $\hat{y}(\mathbf{x})$ reverts to the mean as the predictive location $\mathbf{x}$ deviates from the simulation sites. A number of prior works have discussed this phenomenon and/or developed methods to mitigate it (e.g., Li and Sudjianto 2005, Joseph 2006, Xiong, at al. 2007, Joseph, Hung and Sudjianto 2008, Staum 2009, Gramacy and Lee 2012, Ba and Joseph 2012, Zhang and Apley 2014). To illustrate, consider the following example from Xiong et al. (2007), also considered in Ba and Joseph (2012), in which $d = 1$ and the response surface is the function $y(x) = \sin(30(x - 0.9)^4)\cos(2(x - 0.9)) + (x - 0.9)/2$. Suppose we have observed the simulation response at the 15 input sites shown in Figure 1, which constitute 12 evenly spaced points over the region [0, 0.45] plus 3 additional points evenly spaced over the region [0.45, 1]. The input sites were more densely spaced over the [0, 0.45] region because the response surface varied more over that region. When the data are fitted by a PEXP covariance model with a constant prior mean, the maximum likelihood estimates (MLEs) of the parameters are $\hat{\mu} = -0.15$, $\hat{v} = 1.79$, $\hat{\phi} = 10.9751$, and $\hat{\sigma}^2 = 0.10$, and the corresponding kriging predictor and 95% prediction intervals (PIs) are plotted in Figure 1(a). Notice that $\hat{y}(x)$ has a bumpy appearance over the region $x > 0.5$ and tends to revert to the mean $\hat{\mu}$ as x deviates from

the input sites. Similar but even more extreme reversion to the mean occurs in this example if we use the Gaussian covariance function (not shown here) instead of PEXP.

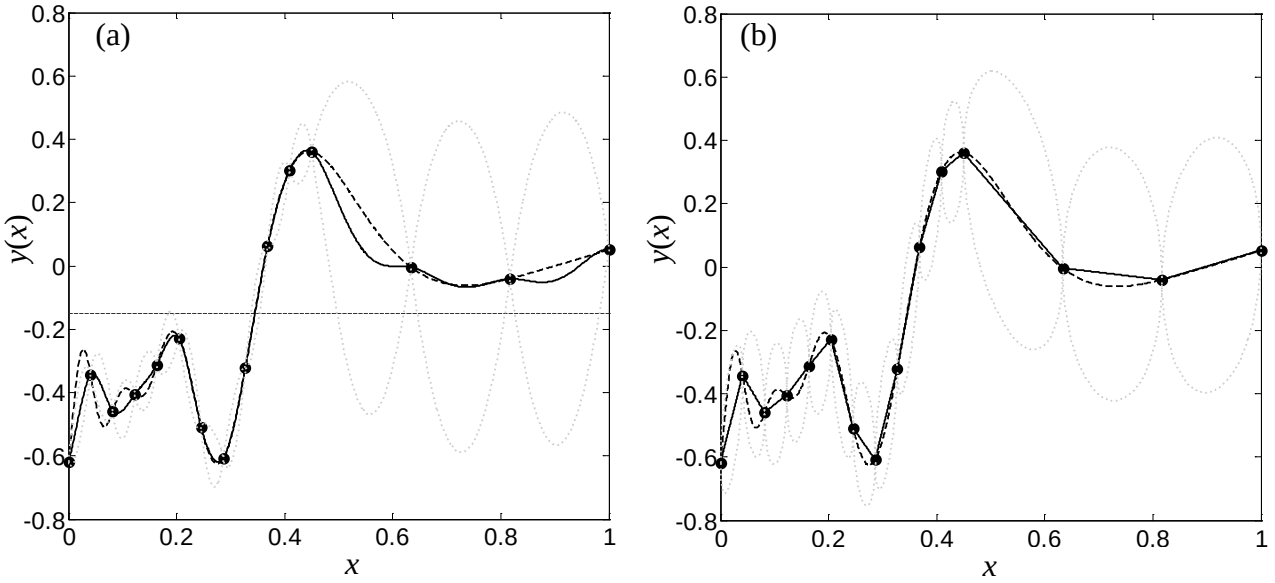

Figure 1. Plot of the function $y(x) = \sin(30(x - 0.9)^4)\cos(2(x - 0.9)) + (x - 0.9)/2$ (dark dashed line), the kriging predictor $\hat{y}(x)$ (solid line), and 95% PIs (gray dotted line) for (a) PEXP covariance function and (b) FBF covariance function. The horizontal dashed line in (a) is the fitted PEXP mean $\hat{\mu}$.

Collecting more observations at additional input sites (e.g., especially over the region $x \in [0.5, 1]$ in Figure 1) would result in a less bumpy $\hat{y}(x)$. However, one could argue that the input sites are already nearly dense enough to capture the overall trend of the response surface in this region, and the prediction with our proposed covariance model (shown later, in Figure 3) looks quite reasonable. More generally, there may be barriers that prevent one from collecting a "sufficiently dense" set of input sites. A complex computer simulation may involve a large number of input variables, and input sites inevitably are sparser in higher dimensional space. Moreover, metamodels are most useful precisely when the computational expense of running the simulation is high, in which case there may not be sufficient computational budget to ensure a dense set of input sites.

Ultimately, a bumpy appearance to $\hat{y}(\mathbf{x})$ in Figure 1(a) is a consequence of the localized nature of the basis functions (as defined in Section 4.1) associated with $\hat{y}(\mathbf{x})$, each of which peaks at a simulation site $\mathbf{x}_i$ and decays to zero as $\mathbf{x}$ moves away from $\mathbf{x}_i$. Motivated by the goal of avoiding a bumpy $\hat{y}(\mathbf{x})$ that reverts to the mean, Zhang and Apley (2014) investigated a

fractional Brownian field (FBF) as the GRF covariance model. In comparison to Figure 1(a), Figure 1(b) shows that the kriging predictor for a FBF model does not revert to the mean. An FBF can be viewed as a certain integral of a white noise GRF [see (6), later], and Zhang and Apley (2014) discuss how this integration results in nonlocalized basis functions via which the FBF model avoids a bumpy predictor that reverts to the mean. They also note that a drawback of the FBF model is that it may not provide a smooth enough predictor to effectively represent many smooth response surfaces, evidence of which can be seen in Figure 1(b) and in Figure 5, discussed later.

In this article, we propose a new class of GRF covariance models that result in nonlocalized, sigmoidally shaped basis functions, but that can be smoother than the FBF basis functions and provide better representation of smooth response surfaces. The core idea is as follows. In the white noise integral representation of an FBF, we replace the white noise GRF with any GRF having a stationary covariance function (e.g., PEXP, Gaussian, Matérn, etc.). We refer to the resulting model as a Brownian integrated (BI) GRF and the associated covariance function as a BI covariance function. For all of the examples in this paper, we use a PEXP covariance function for the underlying GRF to be integrated, and we refer to this as a BIPEXP model.

The relationship between the underlying stationary GRF and its BI counterpart is analogous to the relationship between a stationary autoregressive moving average (ARMA) time series model and its ARIMA counterpart. ARIMA models often perform quite well in forecasting time series that tend to wander, whereas ARMA models are typically better for time series that fluctuate up and down while remaining centered about some long term mean. In Section 4.4 we compare the performances of PEXP and BIPEXP models for a number of real examples and mathematical test function examples from the literature. The BI covariance model performed better than its nonintegrated counterpart in all of our real examples, but the results were mixed for the mathematical test functions. We conjecture that this is because the real examples involve more nicely behaved response surfaces (from deterministic computer simulations of real physical systems, with no measurement error), whereas the mathematical test functions tend to fluctuate

and are more wiggly with more local minima and maxima. A common perception in machine learning is that nicely behaved response surfaces are usually better modeled via sigmoidal basis functions than via localized basis functions (e.g., via a neural network versus a radial basis function network), and vice-versa for response surfaces with many local minima and maxima, and the same may be true for a BIPEXP versus PEXP model.

The format of the remainder of the paper is as follows. In Section 2 we review the FBF model, and in Section 3 we define our BI covariance model and discuss a number of issues related to its use for kriging. In particular, we show that the BI covariance can be conveniently evaluated via a one-dimensional integral, regardless of the dimension *d* of the input domain over which the BI process is defined. Section 4 discusses various characteristics of a BI covariance model, such as sigmoidal versus localized basis functions, and compares the performances of PEXP versus BIPEXP models for a number of examples. Section 5 concludes the paper.

2. REVIEW OF FBFS FOR KRIGING

In ordinary kriging, the response surface is modeled as a realization of a GRF

$$Y(\mathbf{x}) = \mu + Z(\mathbf{x}), \tag{3}$$

where μ is the mean, and $Z(\mathbf{x})$ is a GRF with zero mean and covariance function $R(\bullet,\bullet)$. The predictor is given by (2), and the prediction mean square error (MSE) is

$$\sigma^2(\mathbf{x}) = R(\mathbf{x},\mathbf{x}) - \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}) + [1 - \mathbf{1}^T\mathbf{R}^{-1}\mathbf{r}(\mathbf{x})]^2/\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}, \tag{4}$$

which coincides with the posterior variance in an empirical Bayesian framework (Currin, Mitchell, Morris and Ylvisaker, 1991) in which the covariance parameters are treated as fixed, and a noninformative prior is assumed for μ.

Zhang and Apley (2014) considered an FBF as the GRF covariance model for kriging. A standard FBF of index $0 < p < 2$ is a random field B_p: $R^d \rightarrow R$ satisfying (see Lindstrom, 1993)

(i) $B_p(\mathbf{0}) = 0$ with probability one;

(ii) for all $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in R^d$, the random vector $B_p(\mathbf{x}_1)$, $B_p(\mathbf{x}_2)$, …, $B_p(\mathbf{x}_n)$ is jointly Gaussian with mean zero;

(iii) for all $\mathbf{x}_1, \mathbf{x}_2 \in R^d$, $E\left[\left(B_p(\mathbf{x}_1) - B_p(\mathbf{x}_2)\right)^2\right] = \|\mathbf{x}_1 - \mathbf{x}_2\|^p$, or equivalently,

$$Cov\left[B_p(\mathbf{x}_1), B_p(\mathbf{x}_2)\right] = \frac{1}{2}\left(\|\mathbf{x}_1\|^p + \|\mathbf{x}_2\|^p - \|\mathbf{x}_1 - \mathbf{x}_2\|^p\right); \quad (5)$$

(iv) sample paths are continuous with probability one.

From properties (i) and (ii), it follows that a FBF has increments $B_p(\mathbf{x}_0+\boldsymbol{\delta})-B_p(\mathbf{x}_0)$ that are stationary in the sense that $Cov[B_p(\mathbf{x}_0+\boldsymbol{\delta}_1)-B_p(\mathbf{x}_0),\ B_p(\mathbf{x}_0+\boldsymbol{\delta}_2)-B_p(\mathbf{x}_0)] = \frac{1}{2}\left(\|\boldsymbol{\delta}_1\|^p + \|\boldsymbol{\delta}_2\|^p - \|\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2\|^p\right)$ does not depend on $\mathbf{x}_0$. Moreover, Lindstrom (1993) has shown that a FBF can be represented as the white noise integral

$$B_p(\mathbf{x}) = \int_{\mathbb{R}^d} k_{p,d}\left(\|\mathbf{x} - \mathbf{z}\|^{(p-d)/2} - \|\mathbf{z}\|^{(p-d)/2}\right) W(\mathbf{z})\, d\mathbf{z}\,, \quad (6)$$

where $W(\bullet)$ is a Gaussian white noise field on $\mathbb{R}^d$, and $k_{p,d}$ is a constant that depends only on the index p and the dimension d (see (10) for details).

Aside from a few nuances such as choosing an arbitrary input site to serve as the origin (see Section 3.4), using an FBF model for kriging is a straightforward application of the standard kriging equations with the covariance function given by (5). The fact that $Cov[B_p(\mathbf{x}_1),\ B_p(\mathbf{x}_2)]$ does not decay to zero as $\mathbf{x}_2$ deviates from $\mathbf{x}_1$ is what results in nonlocalized, sigmoidal basis functions for the FBF $\hat{y}(\mathbf{x})$. This is ultimately due the inclusion of an integrator in an FBF model, via the white noise integral representation (6). As discussed in Zhang and Apley (2014), this is the mechanism by which the FBF predictor avoids reversion to the mean and a bumpy $\hat{y}(\mathbf{x})$. This can be observed in the FBF predictor in Figure 1(b), for which the MLE of p was 1 (with $p = d = 1$, the FBF predictor reduces to piecewise linear interpolation), and the MLE of ϕ was 0.9075. As noted in Zhang and Apley (2014), the main drawback of the FBF predictor is that it cannot adequately model many smooth response surfaces. With $p = 1$, the derivative of $\hat{y}(\mathbf{x})$ at each input site does not exist. Although the derivative exists at input sites for $p > 1$, in which case $\hat{y}(\mathbf{x})$ is smoother than for $p = 1$, the second derivative of $\hat{y}(\mathbf{x})$ does not exist. Zhang and Apley (2014) provide a number of examples that illustrate how the FBF predictor eliminates the bumpiness in $\hat{y}(\mathbf{x})$ but also lacks adequate smoothness characteristics to effectively represent some response surfaces. They conclude that it is most effective for modeling response surfaces

that have abrupt features like jumps or sharp peaks. The BI covariance model that we define in the following section includes an integrator, like the FBF model, and inherits all of the characteristics attributed to this, but it is capable of modeling smooth response surfaces.

3. THE BI-COVARIANCE MODEL FOR KRIGING

3.1 Definition of the BI Gaussian Random Field

We define a BI GRF as (6), but with the white noise field $W(\bullet)$ replaced by any zero-mean GRF $C(\bullet)$ having stationary, isotropic covariance function:

$$B_p^C(\mathbf{x}) = \int_{\mathbb{R}^d} k_{p,d}\left(\|\mathbf{x}-\mathbf{z}\|^{(p-d)/2} - \|\mathbf{z}\|^{(p-d)/2}\right) C(\mathbf{z})\, d\mathbf{z}\,. \tag{7}$$

Let $R_C(\|\mathbf{x}-\mathbf{x}'\|) = Cov[C(\mathbf{x}), C(\mathbf{x}')]$ denote the covariance function of $C(\bullet)$. The reason we assume isotropic covariance for $C(\bullet)$ is to enable convenient calculation of the covariance function of $B_p^C(\mathbf{x})$ as described in Section 3.3. However, because anisotropy is usually necessary in kriging to capture lengthscale characteristics that differ in each input coordinate direction, we allow for this by introducing lengthscale parameters as described in Section 3.4.

With the notational convention in (7), a standard FBF is denoted by $B_p^W(\mathbf{x})$. As mentioned in the introduction, we view the integration of the stationary field $C(\bullet)$ in (7) as analogous to integrating a stationary ARMA time series model to yield an ARIMA model. Any stationary, isotropic covariance model for $C(\bullet)$ can be used, although we focus attention on a PEXP covariance model because of its popularity for modeling response surfaces in computer simulations of engineering systems.

3.2 Some Statistical Characteristics of a BI Gaussian Random Field

It is straightforward to show that $B_p^C(\mathbf{0}) = 0$, and $E\left[B_p^C(\mathbf{x})\right] = 0$ for all $\mathbf{x}$. Below, we derive an expression for

$$V_B(\mathbf{x},\mathbf{x}') = Var\left[B_p^C(\mathbf{x}) - B_p^C(\mathbf{x}')\right] \tag{8}$$

and show that $B_p^C(\mathbf{x})$ is intrinsic stationary in the sense that $V_B(\mathbf{x},\mathbf{x}')$ depends only on $\|\mathbf{x}-\mathbf{x}'\|$ (see Cressie, 1991, for a discussion of intrinsic stationarity).

From (8) and (7),

$$V_B(\mathbf{x},\mathbf{x}') = Var\left[B_p^C(\mathbf{x}) - B_p^C(\mathbf{x}')\right] = E\left[\left(B_p^C(\mathbf{x}) - B_p^C(\mathbf{x}')\right)^2\right]$$

$$= E\left[\int_{\mathbb{R}^d} k_{p,d}\left(\|\mathbf{x}-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{z}\|^{\frac{p-d}{2}}\right) C(\mathbf{z})d\mathbf{z} \int_{\mathbb{R}^d} k_{p,d}\left(\|\mathbf{x}-\mathbf{z}'\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{z}'\|^{\frac{p-d}{2}}\right) C(\mathbf{z}')d\mathbf{z}'\right]$$

$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k_{p,d}^2\left(\|\mathbf{x}-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{z}\|^{\frac{p-d}{2}}\right)\left(\|\mathbf{x}-\mathbf{z}'\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{z}'\|^{\frac{p-d}{2}}\right) E[C(\mathbf{z})C(\mathbf{z}')]d\mathbf{z}d\mathbf{z}'$$

$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k_{p,d}^2\left(\|\mathbf{x}-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{z}\|^{\frac{p-d}{2}}\right)\left(\|\mathbf{x}-\mathbf{z}'\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{z}'\|^{\frac{p-d}{2}}\right) R_C(\|\mathbf{z}-\mathbf{z}'\|)d\mathbf{z}d\mathbf{z}'.$$

Defining $\mathbf{z}' = \mathbf{z} + \mathbf{y}$ and using a change of variables in the preceding gives

$$V_B(\mathbf{x},\mathbf{x}') = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k_{p,d}^2\left(\|\mathbf{x}-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{z}\|^{\frac{p-d}{2}}\right)\left(\|\mathbf{x}-\mathbf{z}-\mathbf{y}\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{z}-\mathbf{y}\|^{\frac{p-d}{2}}\right)$$

$$\times R_C(\|\mathbf{y}\|)d\mathbf{z}d\mathbf{y}$$

$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k_{p,d}^2 \cdot \frac{1}{2}\Big[\left(\|\mathbf{x}-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{y}-\mathbf{z}\|^{\frac{p-d}{2}}\right)^2$$

$$-\left(\|\mathbf{x}-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}-\mathbf{y}-\mathbf{z}\|^{\frac{p-d}{2}}\right)^2 + \left(\|\mathbf{x}'-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}-\mathbf{y}-\mathbf{z}\|^{\frac{p-d}{2}}\right)^2$$

$$-\left(\|\mathbf{x}'-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{y}-\mathbf{z}\|^{\frac{p-d}{2}}\right)^2\Big] \cdot R_C(\|\mathbf{y}\|)d\mathbf{z}d\mathbf{y}$$

$$= \int_{\mathbb{R}^d} k_{p,d}^2 \cdot \frac{1}{2} \cdot R_C(\|\mathbf{y}\|)\Big\{\int_{\mathbb{R}^d}\Big[\left(\|\mathbf{x}-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{y}-\mathbf{z}\|^{\frac{p-d}{2}}\right)^2$$

$$-\left(\|\mathbf{x}-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}-\mathbf{y}-\mathbf{z}\|^{\frac{p-d}{2}}\right)^2 + \left(\|\mathbf{x}'-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}-\mathbf{y}-\mathbf{z}\|^{\frac{p-d}{2}}\right)^2$$

$$-\left(\|\mathbf{x}'-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{x}'-\mathbf{y}-\mathbf{z}\|^{\frac{p-d}{2}}\right)^2\Big]d\mathbf{z}\Big\}d\mathbf{y}. \tag{9}$$

Lindstrom (1993) derives the result

$$\int_{\mathbb{R}^d}\left(\|\mathbf{x}-\mathbf{z}\|^{\frac{p-d}{2}} - \|\mathbf{y}-\mathbf{z}\|^{\frac{p-d}{2}}\right)^2 d\mathbf{z} = \|\mathbf{x}-\mathbf{y}\|^p k_{p,d}^{-2}. \tag{10}$$

Combining (9) and (10) gives

$$V_B(\mathbf{x},\mathbf{x}') = \int_{\mathbb{R}^d} \frac{1}{2} \cdot k_{p,d}^2 \cdot R_C(\|\mathbf{y}\|) \cdot k_{p,d}^{-2} \cdot 2(\|\mathbf{x}-\mathbf{x}'+\mathbf{y}\|^p - \|\mathbf{y}\|^p)d\mathbf{y}$$

$$= \int_{\mathbb{R}^d}\{\|\mathbf{x}-\mathbf{x}'+\mathbf{y}\|^p - \|\mathbf{y}\|^p\}R_C(\|\mathbf{y}\|)d\mathbf{y}, \tag{11}$$

where we have used the result $\int_{\mathbb{R}^d}\|\mathbf{x}-\mathbf{x}'+\mathbf{y}\|^p R_C(\|\mathbf{y}\|)d\mathbf{y} = \int_{\mathbb{R}^d}\|\mathbf{x}-\mathbf{x}'-\mathbf{y}\|^p R_C(\|-\mathbf{y}\|)d\mathbf{y}$ $= \int_{\mathbb{R}^d}\|\mathbf{x}'-\mathbf{x}+\mathbf{y}\|^p R_C(\|\mathbf{y}\|)d\mathbf{y}$, with the first equality following via the change of variables $\mathbf{y} \to -\mathbf{y}$.

From (11), and because $R_C(\|\mathbf{y}\|)$ is isotropic, it follows that the variance $V_B(\mathbf{x},\mathbf{x}')$ of the increment is a function of only $\|\mathbf{x}-\mathbf{x}'\|$. To see this, suppose we have two other vectors $\mathbf{z}$ and $\mathbf{z}'$

such that $\|\mathbf{z}-\mathbf{z}'\| = \|\mathbf{x}-\mathbf{x}'\|$. Then we can write $\mathbf{x}-\mathbf{x}' = \mathbf{Q}(\mathbf{z}-\mathbf{z}')$ for some orthogonal matrix $\mathbf{Q}$. If we change the integration variable in (11) for $V_B(\mathbf{z},\mathbf{z}')$ from $\mathbf{y} \to \mathbf{Q}\mathbf{y}$, it follows that $V_B(\mathbf{z},\mathbf{z}') = V_B(\mathbf{x},\mathbf{x}')$, which implies that $V_B(\mathbf{x},\mathbf{x}')$ depends only on $\|\mathbf{x}-\mathbf{x}'\|$. Thus, $B_p^C(\mathbf{x})$ is an intrinsic stationary GRF. Furthermore, let $R_B(\mathbf{x},\mathbf{x}') = Cov\left[B_p^C(\mathbf{x}), B_p^C(\mathbf{x}')\right]$ denote the covariance function of $B_p^C(\mathbf{x})$. Because $B_p^C(\mathbf{0}) = 0$, it follows that $V_B(\mathbf{x},\mathbf{x}')$ can be specified in terms of $R_B(\mathbf{x},\mathbf{x}')$, and vice-versa, via the relationship

$$\begin{aligned} 2R_B(\mathbf{x},\mathbf{x}') &= V_B(\mathbf{x},\mathbf{0}) + V_B(\mathbf{x}',\mathbf{0}) - V_B(\mathbf{x},\mathbf{x}') \\ &= V_B(\|\mathbf{x}\|) + V_B(\|\mathbf{x}'\|) - V_B(\|\mathbf{x}-\mathbf{x}'\|), \end{aligned} \tag{12}$$

where, with a slight abuse of notation, we have written $V_B(\mathbf{x},\mathbf{x}') = V_B(\|\mathbf{x}-\mathbf{x}'\|)$. We note that the function $V_B(\cdot)$ depends on $R_C(\cdot)$ and p, although we have suppressed this in the notation.

As stated in Lindstrom (1993), (6) is not valid for the case that $p = d = 1$, and neither is (7). However, (11) and (12) do still yield a valid covariance function for $p = d = 1$. Because the kriging approach described subsequently uses the covariance function, but not (7) directly, the approach remains valid for $p = d = 1$.

3.3 Computing the Covariance Function of a BI Gaussian Random Field

To be used for kriging, it is necessary to repeatedly compute $R_B(\mathbf{x},\mathbf{x}')$ [via $V_B(\mathbf{x},\mathbf{x}')$ and (12)] for various input values $\mathbf{x}$ and $\mathbf{x}'$. For $d = 1$, (11) can be used to evaluate $V_B(\mathbf{x},\mathbf{x}')$ via the one-dimensional integral. For $d > 1$, however, the d-dimensional integration in (11) is clearly computationally prohibitive. In this section we show that (11), and thus (12), can be evaluated via a one-dimensional integral for any d.

Because (11) is a function of only $\|\mathbf{x}-\mathbf{x}'\|$, without loss of generality we can assume that $\mathbf{x}-\mathbf{x}'$ is aligned with the first coordinate axis, as illustrated in Figure 2. Define $H = \|\mathbf{x}-\mathbf{x}'\|$. For fixed H, because $R_C(\|\mathbf{y}\|)$ is isotropic, the integrand in (11) depends only on r and θ, where we have defined $r = \|\mathbf{y}\|$, and $\theta = \cos^{-1}[\mathbf{y}^T(\mathbf{x}'-\mathbf{x})/(r \cdot H)]$ to be the angle between $\mathbf{y}$ and $\mathbf{x}'-\mathbf{x}$. Hence, the d-dimensional integral in (11) can be reduced to a two-dimensional integral using polar coordinates as follows and as illustrated in Figure 2.

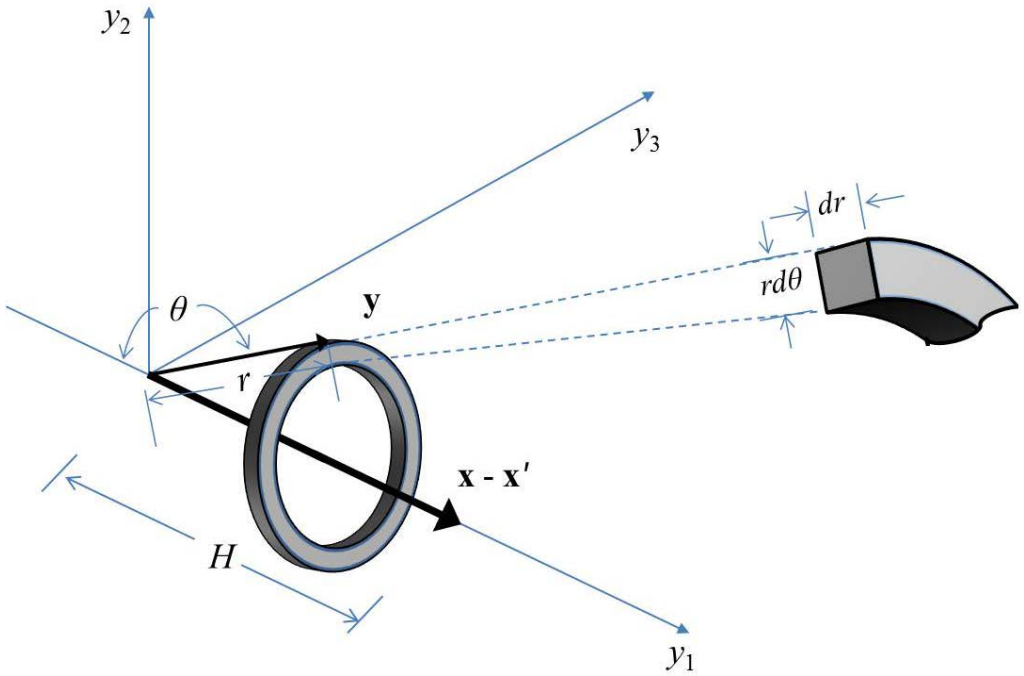

Figure 2. Illustration of how the variance of the BI increment in (11) can be calculated using a two-dimensional integral.

First note that the integrand of (11) assumes the constant value $\{\|\mathbf{x}-\mathbf{x}'+\mathbf{y}\|^p - \|\mathbf{y}\|^p\}R_C(\|\mathbf{y}\|) = \{(H^2+r^2-2Hr\cos\theta)^{p/2} - r^p\}R_C(r)$ over the ring in Figure 2 that corresponds to the fixed values of r and θ. Moreover, the integral of a unit integrand over the ring is the circumference of the ring (for the d = 3 case), which is $V(r,\theta) = 2\pi r\sin\theta$, and the differential cross-sectional area element for the ring is $rd\theta dr$. Hence, for the d = 3 case illustrated in Figure 2, the polar coordinate version of the integral of (11) becomes

$$V_B(H) = \int_{\mathbb{R}^3}\{\|\mathbf{x}-\mathbf{x}'+\mathbf{y}\|^p - \|\mathbf{y}\|^p\}R_C(\|\mathbf{y}\|)d\mathbf{y}$$
$$= \int_0^\infty\int_0^\pi\{(H^2+r^2-2Hr\cdot\cos\theta)^{p/2} - r^p\}R_C(r)\,V(r,\theta)rd\theta dr$$

More generally, in higher dimensions, the circumference of the ring over which the integrand in (11) is constant (for fixed r and θ) is replaced by the surface volume of a d–2 dimensional sphere of radius $r\sin\theta$ in d-dimensional space, which is $V(r,\theta) = 2\pi^{(d-1)/2}(r\sin\theta)^{(d-2)}/\Gamma\left(\frac{d-1}{2}\right)$, where Γ denotes the gamma function. Thus, for any $d \geq 2$, the polar coordinate version of (11) becomes

$$V_B(H) = \int_0^\infty\int_0^\pi\{(H^2+r^2-2Hr\cdot\cos\theta)^{p/2} - r^p\}R_C(r)\,V(r,\theta)rd\theta dr$$
$$= \int_0^\infty\int_0^\pi\{(H^2+r^2-2Hr\cdot\cos\theta)^{p/2} - r^p\}R_C(r)\frac{2\pi^{(d-1)/2}r^{d-1}(\sin\theta)^{d-2}}{\Gamma\left(\frac{d-1}{2}\right)}d\theta dr, \qquad (13)$$

which further simplifies to (Wolfram, 2014)

$$V_B(H) = \frac{2\pi^{\left(\frac{d}{2}\right)}}{\Gamma\left(\frac{d}{2}\right)} \int_0^\infty R_C(r) \cdot r^{d-1} \cdot \left[-r^p + (H+r)^p \cdot F\left(\frac{d-1}{2}, -\frac{p}{2}; d-1; \frac{4Hr}{(H+r)^2}\right)\right] dr \qquad (14)$$

where $F(\bullet,\bullet;\bullet;\bullet)$ denotes the Gaussian or ordinary hypergeometric function (often denoted by ${}_2F_1(\bullet,\bullet;\bullet;\bullet)$ in the literature), which in general is defined via an infinite sum.

Evaluating the infinite sum in the hypergeometric function may negate the benefits of reducing the two-dimensional integral in (13) to the one-dimensional integral in (14). However, when *d* is odd, the hypergeometric function simplifies to a function that can be easily computed without an infinite summation, as described in the online supplement. Consequently, if *d* is even, we recommend the trick of artificially increasing the dimension of the input space by one, adding an extra redundant dimension whose coordinate value is always 0 in the (*d*+1)-dimensional augmented input vector **x**. In this case, all of the input sites and any input value at which one wishes to predict the response will lie in a subspace that has one fewer dimension than the (*d*+1)-dimensional space over which the BI Gaussian random process is defined. We have encountered no numerical problems (or problems of any type) caused by augmenting the dimension in this manner, and it allows the BI covariance function to be evaluated via the one-dimensional integral of (14) without having an integrand that must be evaluated as an infinite sum.

Although this even-*d* trick does not generally give the same results as if (14) were evaluated exactly for even *d*, the resulting covariance model may have no less conceptual appeal. One way to view the trick is that it is like having a simulation model in which *d*+1 inputs can be varied, but the experimenter decides to only vary the first *d* inputs when collecting the data, keeping the $(d+1)^{st}$ input constant. This happens frequently in practice, as complex simulation experiments rarely vary every input that can be varied. The primary presumption behind the even-*d* trick is that, in the hypothetical scenario in which the experimenter also were to vary the $(d+1)^{st}$ input, the response would behave as a BI covariance model in (*d*+1)-dimensional input space.

The expression (14) is valid for any isotropic covariance function $R_C(r)$, such as the isotropic Gaussian covariance, isotropic Matérn covariance, or the isotropic version

$$R(\mathbf{x}, \mathbf{x}') = \sigma^2 exp\{-\|\mathbf{x} - \mathbf{x}'\|^{\nu}\} \tag{15}$$

of the PEXP covariance. Notice that the PEXP covariance of (1) for $\nu < 2$ is not isotropic (under Euclidean distance) even when $\mathbf{\Phi}$ is the identity matrix. In the remainder of this paper, we use the isotropic PEXP covariance of (15) in the BIPEXP covariance model (lengthscale parameters are included via (16) later) Unless otherwise noted, the PEXP covariance model by itself will refer to the version in (1). In the literature, this seems to be more popular than the version in (15) with $\mathbf{x}–\mathbf{x}'$ replaced by $\mathbf{\Phi}[\mathbf{x}–\mathbf{x}']$.

3.4 Using the BI Covariance Model for Kriging

Except for a few nuances discussed in this section, using the BI covariance model for kriging is fairly straightforward, and (2) and (4) can be applied directly with the covariance calculated via (12) and (14) (for $d \geq 2$) or (12) and (11) (for $d = 1$). In terms of handling the nuances, our treatment in this section mirrors that of Zhang and Apley (2014).

If we model the response surface as a BI GRF plus an unknown constant mean μ, then the constraint $B_p^C(\mathbf{0}) = 0$ implies that $Y(\mathbf{0}) = \mu$. In light of this, we arbitrarily choose one of the input sites (which we label $\mathbf{x}_1$) as the origin, view the observed response y_1 at that input site as perfect prior knowledge of μ, and model the response surface as

$$Y(\mathbf{x}) = y_1 + B_p^C\big(\mathbf{\Phi}(\mathbf{x} - \mathbf{x}_1)\big), \tag{16}$$

where $\mathbf{\Phi} = \text{diag}\{\phi_1, \phi_2, \ldots, \phi_d\}$ are the lengthscale parameters with each $\phi_i > 0$. The same as for the PEXP lengthscale parameters, a smaller ϕ_i means the response surface varies more slowly in the ith spatial coordinate.

For the remainder of the paper, let $\mathbf{y}$ denote the vector of $n–1$ remaining observations, $\mathbf{R}$ denote the $(n–1)\times(n–1)$ covariance matrix of $\mathbf{y}$, etc. For an FBF model, Zhang and Apley (2014) have shown that the choice of input site to serve as the origin is arbitrary in the sense that neither the likelihood function of $\mathbf{y}$ (and hence the MLE of the parameters), nor the posterior predictive distribution of $Y(\mathbf{x})$ at any $\mathbf{x}$ (given $\mathbf{y}$, and treating the MLEs as fixed parameters), depends on

which input site is chosen as the origin. The arguments used in Zhang and Apley (2014) apply directly to the present case of a BI GRF model.

It is well-known (see Cressie, 1991) that the kriging predictor can be expressed in terms of the variogram, which we have denoted by $V_B(H)$ for the BI random field, and one could use the variogram form to avoid having to choose an arbitrary input site as the origin. However, the covariance function is still required to calculate the MLEs of the parameters. Because of this, and because the covariance form of the kriging predictor is more familiar in engineering response surface modeling, we adopt the preceding convention in which an arbitrary input site is chosen as the origin. In this case, the BI kriging predictor in (2) becomes

$$\hat{y}(\mathbf{x}) = y_1 + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - y_1 \cdot \mathbf{1}), \tag{17}$$

where **1** now denotes the (n–1)-length column vector of ones, and **r**(**x**) denotes the (n–1)-length column vector whose elements are $\{Cov[Y(\mathbf{x}), Y(\mathbf{x}_i)]: i = 2, 3, \ldots, n\}$. The elements of **r** and **R** are defined via $Cov[Y(\mathbf{x}), Y(\mathbf{x}')] = R_B(\mathbf{\Phi}(\mathbf{x}-\mathbf{x}_1), \mathbf{\Phi}(\mathbf{x}'-\mathbf{x}_1))$ using (12) and (14) to calculate $R_B(\bullet, \bullet)$. The prediction MSE is $\sigma^2(\mathbf{x}) = R(\mathbf{x},\mathbf{x}) - \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x})$, which is of the same form as (4) but with the right-most term omitted. The right-most term in (4) was due to the uncertainty associated with estimating μ, which is irrelevant for a BI covariance model.

To use any random field model for kriging, one must either choose or estimate each parameter (unless one marginalizes them in a fully Bayesian approach, which we do not use in this work). For the PEXP model, the parameters are μ, σ^2, ν and lengthscale parameters $\{\phi_1, \phi_2, \ldots, \phi_d\}$. For the BIPEXP models, the parameters are μ (estimated implicitly as y_1), p, σ^2, ν and lengthscale parameters $\{\phi_1, \phi_2, \ldots, \phi_d\}$. Hence, relative to the PEXP model, the BIPEXP model involves only a single additional parameter p.

The likelihood function for the BIPEXP model is

$$L(\mathbf{\Phi}, p; \mathbf{y}) = \frac{1}{(2\pi)^{(n-1)/2}|\mathbf{R}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - y_1 \cdot \mathbf{1})^T\mathbf{R}^{-1}(\mathbf{y} - y_1 \cdot \mathbf{1})\right\} \tag{18}$$

where $[\mathbf{R}]_{i,j} = R_B(\mathbf{\Phi}(\mathbf{x}_{i+1}-\mathbf{x}_1))$, and $\mathbf{\Phi}(\mathbf{x}_{j+1}-\mathbf{x}_1)$ is calculated as described above. Given specific values of $\hat{\nu}, \hat{p}$ and $\{\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_d\}$, the MLE of σ^2 can be easily derived from (18) as

$$\hat{\sigma}^2\left(\hat{v},\hat{p},\{\hat{\phi}_1,\hat{\phi}_2,\ldots,\hat{\phi}_d\}\right)=\frac{1}{n-1}(\mathbf{y}-y_1\cdot\mathbf{1})^T\mathbf{R}'^{-1}(\mathbf{y}-y_1\cdot\mathbf{1}) \quad (19)$$

where $[\mathbf{R}']_{i,j}$ is the same as $[\mathbf{R}]_{i,j}$ but with $\sigma^2 = 1$. Thus, only v, p and $\{\phi_1, \phi_2, \ldots, \phi_d\}$ are estimated directly in the MLE procedure.

In the PEXP covariance model, the convention is to restrict *v* to lie in the closed interval [1, 2], and we keep with this convention when we use a PEXP $R_C(\cdot)$ in the BI covariance model. In a FBF model *p* is restricted to lie in the open interval (0, 2). However, we recommend restricting *p* to the interval [1, 2), which we do for all examples in this paper, unless otherwise noted. We use a lower bound of 1.0 on *p* for reasons discussed in the online supplement. Moreover, because $R_B(\bullet, \bullet)$ is continuous in *p*, numerical problems may surface if *p* is too close to 2. In light of this, we placed an upper bound of 1.99 on *p* for all examples in this paper in which *p* was estimated. How far below 2 the upper bound should be set depends on the numerical precision of the software and hardware. In all of our examples, we observed no numerical problems or significant difference in the predictions when the upper bound on *p* was raised to 1.9999, versus 1.99. However, there may be little practical benefit of allowing a *p* higher than 1.99, and we prefer the numerically safer upper bound of 1.99. Ranjan, et al. (2011) found that a similar upper bound on *v* for the PEXP model gave much better numerical stability, even though the limiting case *v* =2 is still a valid covariance.

4. DISCUSSION AND EXAMPLES

4.1 Sigmoidal Versus Localized Basis Functions

Returning to the example in Figure 1, we fit the same data using a BIPEXP covariance model with all parameters estimated by MLE ($\hat{v} = 2, \hat{p} = 1, \hat{\phi} = 33.14, \hat{\sigma}^2 = 0.02$). Figure 3 shows the resulting BIPEXP predictor $\hat{y}(x)$ with 95% PIs. The predictor $\hat{y}(x)$ for the BIPEXP model in Figure 3 is less bumpy than for the PEXP model in Figure 1(a), and it is smoother than for the FBF model in Figure 1(b). The root mean square error (RMSE) for 1,000 test sites evenly spaced over the interval [0,1] is 0.0317 for the BIPEXP model, versus 0.0484 for the PEXP model. The

95% PIs have similar shape as the PIs for the PEXP model in Figure 1(a), but are slightly narrower with different centers $\hat{y}(x)$.

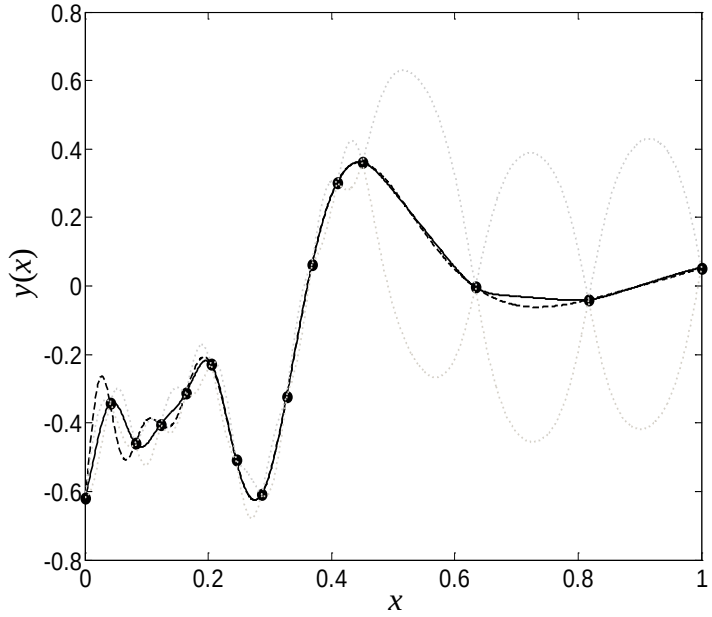

Figure 3. For the example of Figure 1, $y(x)$ (dark dashed line), the predictor $\hat{y}(x)$ (solid line), and 95% PIs (gray dotted line) for the BIPEXP model (compare to Figure 1).

To investigate a fundamental difference between a BI covariance model and models like PEXP, which is related to the preceding, consider the basis function representation of the general kriging predictor in (2):

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \sum_{i=1}^{n} c_i g_i(\mathbf{x}), \tag{20}$$

where $\{c_i: i = 1, 2, \ldots, n\}$ are coefficients that depend on the data, and $g_i(\mathbf{x}) = Cov[Y(\mathbf{x}), Y(\mathbf{x}_i)]$ ($i = 1, 2, \ldots, n$) are viewed as basis functions for the response surface predictor. The right panels of Figures 4 and 5 show example basis functions for various PEXP, FBF, and BIPEXP models for the $d = 1$ case. The left panels, discussed in Section 4.2, show five random realizations for each model. The basis functions in the right panels are for two input sites ($x_i = 1.0$ and $x_i = 5$) with the origin taken to be $x = 0$. The variance parameter was $\sigma^2 = 1$ for all models. Because the convention described in Section 3.4 is to translate the BIPEXP and FBF models so that the origin $Y(0) = 0$ coincides with an observed response value, their covariance functions should really be viewed as the conditional covariance function $g_i(\mathbf{x}) = Cov[Y(\mathbf{x}), Y(\mathbf{x}_i) \mid Y(0) = 0]$. Consequently,

to provide a more common basis for comparison, the PEXP basis functions plotted in Figure 4 are also the conditional covariance function $g_i(\mathbf{x}) = Cov[Y(\mathbf{x}), Y(\mathbf{x}_i) \mid Y(0) = 0] = R(\mathbf{x}, \mathbf{x}_i) - R(\mathbf{x}, \mathbf{0})R(\mathbf{x}_i, \mathbf{0})/R(\mathbf{0}, \mathbf{0})$, where $R(\mathbf{x}, \mathbf{x}')$ is the usual PEXP covariance function given by (1).

The form of the basis functions strongly influences the nature of the fitted response surface $\hat{y}(\mathbf{x})$. For a covariance function like PEXP, in which $Cov[Y(\mathbf{x}), Y(\mathbf{x}_i)]$ decays to zero as $\|\mathbf{x} - \mathbf{x}_i\|$ increases, the kriging predictor in (20) reverts to the mean, by definition. The resulting basis functions are localized around $\mathbf{x}_i$, as seen in the right column of Figure 4, and the degree of localization is governed by the parameters ν and ϕ. Depending on the degree of localization, this can result in a bumpy $\hat{y}(\mathbf{x})$, as in Figure 1(a). In contrast, for the FBF model, $Cov[Y(\mathbf{x}), Y(\mathbf{x}_i)]$ in (5) does not decay to zero as $\|\mathbf{x} - \mathbf{x}_i\|$ increases. As a result, the FBF basis functions are not localized about $\mathbf{x}_i$ The FBF basis functions for $p = 1$ (see the online supplement) are sigmoidal, albeit piecewise linear. This is consistent with the piecewise linear nature of $\hat{y}(\mathbf{x})$ in Figure 1(b). As can be seen in the right column of Figure 4, for $p > 1$, the FBF basis functions are more smoothly sigmoidal and result in a smoother $\hat{y}(\mathbf{x})$. However, achieving a high level of smoothness requires $p \approx 2$, which introduces serious problems discussed in Section 4.3.

Similarly, $Cov[Y(\mathbf{x}), Y(\mathbf{x}_i)]$ for a BI covariance model does not decay to zero as $\|\mathbf{x} - \mathbf{x}_i\|$ increases, and, as a result, the BI covariance basis functions are not localized about $\mathbf{x}_i$. Basis functions for the BIPEXP model are shown in Figure 5 for various p, ν, and ϕ. For $p = 1$, the basis functions are completely sigmoidal and flatten out as $\mathbf{x}$ grows. Although possessing similar global shape and trends, the basis functions for $p = 1$ in the top two rows of Figure 5 have different smoothness near the knots (due to different ν and ϕ), and the corresponding realizations also have different smoothness. For $p < 1$ (third row of Figure 5), the BIPEXP basis functions have a slightly more localized appearance with features that resemble peaks, while for $p > 1$ (bottom row of Figure 5), the basis functions continue to increase as $\mathbf{x}$ grows. See the online supplement for further discussion and examples.

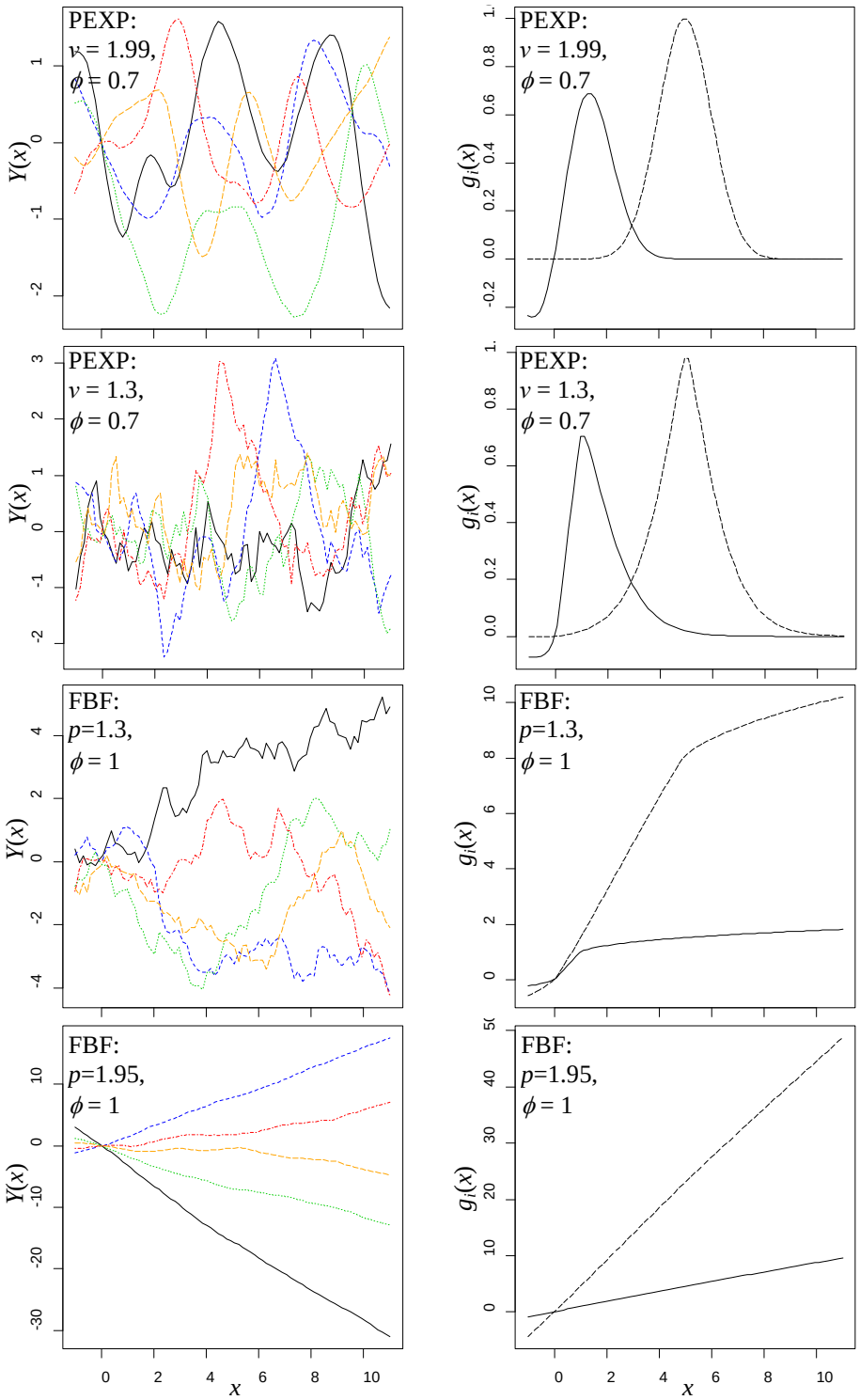

Figure 4. Examples of five random realizations of $Y(x)$ (left column) and two basis functions $g_i(x)$ (right column) for PEXP (top two rows) and FBF (bottom two rows) models with $d = 1$ and various $\{p, \nu, \phi\}$. The two basis functions are for $x_i = 1$ (solid line) and $x_i = 5$ (dashed line).

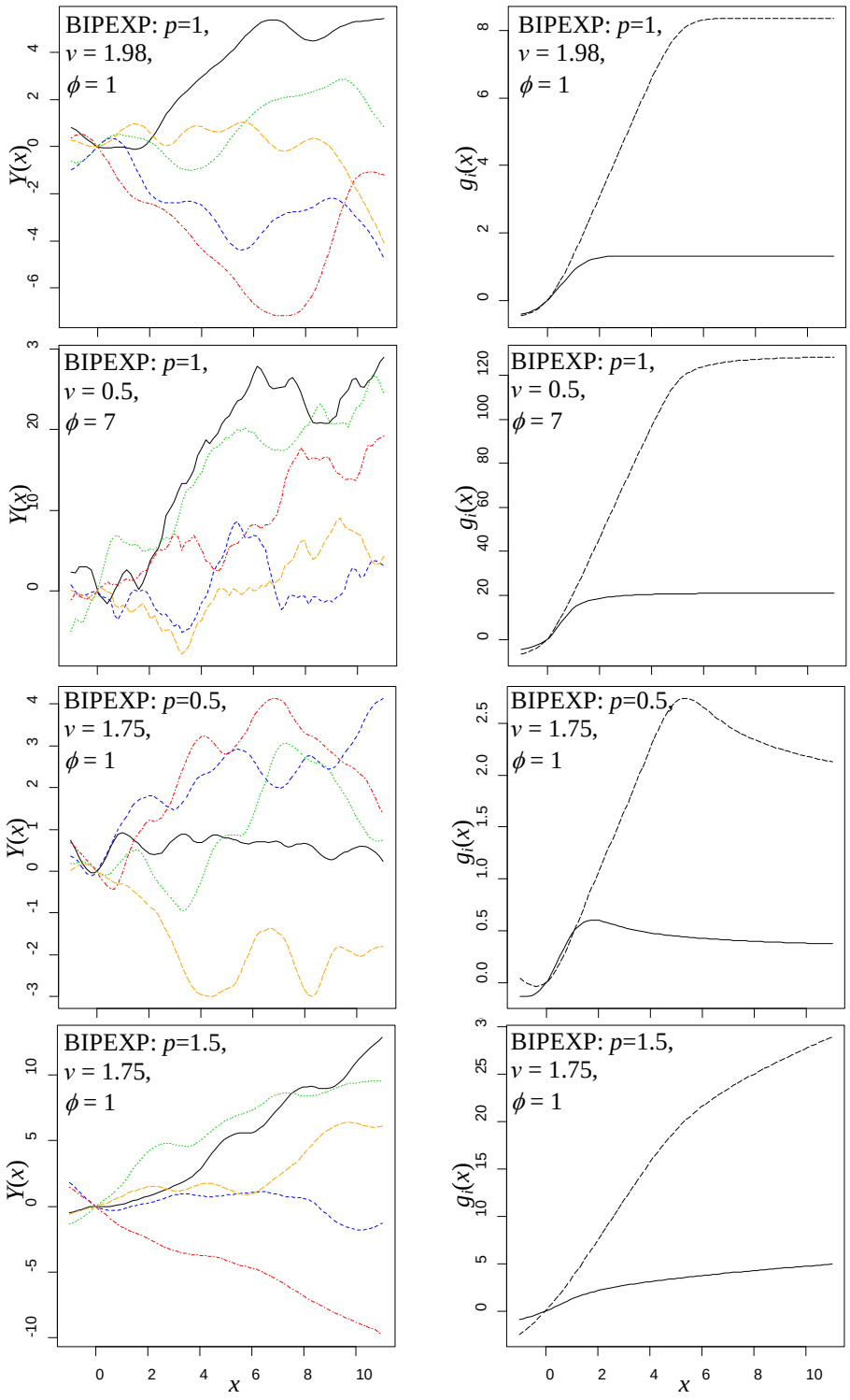

Figure 5. Examples of five random realizations of $Y(x)$ (left column) and two basis functions $g_i(x)$ (right column) for BIPEXP models with $d = 1$ and various $\{p, \nu, \phi\}$. The two basis functions are for $x_i = 1$ (solid line) and $x_i = 5$ (dashed line).

The issue of localized versus sigmoidal basis functions has an interesting parallel in machine learning. Artificial neural networks are typically comprised of sigmoidal activation functions, but a radial basis function network is a competing architecture wherein the sigmoidal activation functions are replaced with radial ones (typically, Gaussian functions). A common perception in machine learning is that certain "nicely-behaved" response surfaces (e.g., ones that are quite smooth with few high-frequency components, monotonic, nearly constant over some input regions, etc.) are better modeled by sigmoidal functions than by radial basis functions, and vice-versa for response surfaces with many local minima and maxima and higher-frequency components (see, e.g., Flake, 2012; Lee, et al., 1999; Barron, 1993). The same may be true for GRF modeling, a conjecture that seems to be supported by the examples in Section 4.4.

4.2 Scale-Nonstationarity Versus BI-Nonstationarity

There is another fundamental difference between BI covariance models and their stationary counterparts that is evident from their covariance functions and that can be seen in the random realizations in the left panels of Figures 4—5. The stationary PEXP GRFs (top two rows of Figure 4) fluctuate up and down in a random manner but stay centered about the mean over the long term. If a PEXP $Y(\mathbf{x})$ drifts above its mean at a particular $\mathbf{x}$, it will eventually drift back down below the mean as $\mathbf{x}$ increases. In contrast, a BI covariance $Y(\mathbf{x})$ wanders in a manner that is not centered about some long term mean, and it may grow unbounded as $\mathbf{x}$ increases. Mathematically, this is easiest to see for the FBF model, for which the variance of the increment is $Var[Y(\mathbf{x}) - Y(\mathbf{x}_i)] = \|\mathbf{x} - \mathbf{x}_i\|^p$, which grows unboundedly as $\mathbf{x}$ moves further from a simulation site $\mathbf{x}_i$. If the x-axes in the left panels of Figures 4—5 were extended, the plotted realizations of $Y(\mathbf{x})$ for the FBF and BIPEXP models would grow to very large values. This also implies that the prediction error variance (and, consequently, the width of a PI) for $Y(\mathbf{x})$ continues to grow unbounded as $\mathbf{x}$ moves further from all of the simulation sites. In contrast, for the stationary PEXP model, the prediction error variance asymptotically approaches the constant prior variance σ^2. However, unless one is extrapolating well beyond the experimental input

range (which is generally ill advised), this distinction may be more of theoretical interest than of practical consequence. For most of the examples in which we have compared PIs for PEXP and BIPEXP models (using their respective MLEs, with both models fit to the same data), the PIs tend to have similar shape and width.

We will refer to the preceding type of nonstationarity – for which the increments $Y(\mathbf{x}+\mathbf{z}) - Y(\mathbf{x})$ are stationary in $\mathbf{x}$ for a fixed spatial increment $\mathbf{z}$, but the variance of $Y(\mathbf{x}+\mathbf{z}) - Y(\mathbf{x})$ grows unbounded as $\|\mathbf{z}\|$ grows – as BI-nonstationarity. Zhang and Apley (2014) discuss other distinguishing characteristics of FBFs related to BI-nonstationarity. For example, for two input locations $\mathbf{x}$ and $\mathbf{x}'$ separated by an angle θ and with $\|\mathbf{x}\| = \|\mathbf{x}'\|$, the correlation function $Corr[Y(\mathbf{x}), Y(\mathbf{x}')] = 1 - 2^{(p-2)/2}[1 - cos(\theta)]^{p/2}$ depends only on θ. Thus, the correlation between $Y(\mathbf{x})$ and $Y(\mathbf{x}')$ is exactly the same as the correlation between $Y(\mathbf{u})$ and $Y(\mathbf{u}')$ for unit vectors $\mathbf{u}$ and $\mathbf{u}'$ aligned with $\mathbf{x}$ and $\mathbf{x}'$, no matter how large is $\|\mathbf{x}\|$. Moreover, for a fixed simulation input site $\mathbf{x}_i$, although $Cov[Y(\mathbf{x}), Y(\mathbf{x}_i)]$ does not decay to zero as $\mathbf{x}$ moves further from $\mathbf{x}_i$, the correlation $Corr[Y(\mathbf{x}), Y(\mathbf{x}_i)] = Cov[Y(\mathbf{x}), Y(\mathbf{x}_i)]/\{Var[Y(\mathbf{x})]Var[Y(\mathbf{x}_i)]\}^{1/2}$ does indeed decay to zero. These characteristics may or may not be desirable, depending on the nature of the response surface being modeled (see Zhang and Apley, 2014, for further discussion).

In the kriging literature, BI-nonstationarity is frequently discussed in the context of intrinsic kriging (Cressie, 1991). A second type of nonstationarity, which we refer to as scale-nonstationarity, is also frequently encountered but is fundamentally different than BI-nonstationarity. Scale-nonstationarity is associated with processes that have scale/smoothness characteristics that vary spatially. The response surface in Figures 1 and 3 appears to exhibit scale-nonstationarity, as there are more local peaks and valleys over the region $x < 0.5$ than over the region $x > 0.5$. Most approaches to handle scale-nonstationarity modify standard stationary covariance functions by allowing the lengthscale parameters to vary spatially or by transforming $\mathbf{x}$ so that $Y(\mathbf{x})$ becomes stationary in the transformed inputs (e.g., Sampson and Guttorp, 1992; Schmidt and O'Hagan, 2003; Paciorek and Schervish, 2006; Xiong, et al. 2007; Fuentes, Chen and Davis, 2008; Gramacy and Lee, 2008; and Huang, Wang, Breidt and Davis, 2011).

Although the BIPEXP model did a better job than the stationary PEXP model for this example (compare Figures 1 and 3), a BI covariance model is not scale-nonstationary, nor is it inherently capable of handling scale-nonstationary surfaces. This follows because the increments $Y(\mathbf{x}+\mathbf{z}) - Y(\mathbf{x})$ are stationary in $\mathbf{x}$ for a BI covariance model, which implies that the local scale/smoothness characteristics are constant over the entire input domain. A possible explanation for why the BIPEXP model did a better job than the stationary PEXP model for this example is that the former avoids reversion to the mean, which may be exacerbated in scale-nonstationary situations like in Figure 1 (for which the estimated scale parameter is relatively large because of the waviness of the surface in the left half, which results in more reversion to the mean in the more sparsely simulated right half).

If a response surface is truly scale-nonstationary, then this may be best handled via use of a scale-nonstationary model. For situations like in Figure 1, a scale nonstationary model also could have mitigated reversion to the mean, although it would accomplish this via a fundamentally different mechanism than the BI covariance model. The BI covariance model avoids reversion-to-the-mean because its sigmoidal basis functions are not localized and do not decay to zero. In contrast, although a scale-nonstationary model still has localized basis functions that decay to zero, in smoother regions they would ideally decay to zero more slowly (via a smaller local lengthscale parameter ϕ) and result in a less bumpy $\hat{y}(\mathbf{x})$. This assumes that the spatially varying scale parameterization can capture the scale-nonstationarity and that the model fitting is well behaved, so that the estimated lengthscale parameters are indeed smaller in the smoother regions.

4.3 Limitations of an FBF model with $p \approx 2$

An FBF model has nonlocalized, sigmoidal basis functions, but it is typically too locally rough to represent many response surfaces. This was concluded in Zhang and Apley (2014), and it is somewhat evident from the $p = 1.3$ case in Figure 4 (also see Figure S2 in the online supplement) and from Figure 1(b), in which $\hat{y}(\mathbf{x})$ is piecewise linear. However, because an FBF is much smoother locally for $p \approx 2$, one might wonder whether using an FBF model with the

restriction $p \approx 2$ will achieve sigmoidal basis functions while handling smooth response surfaces. The p = 1.95 case in Figure 4 illustrates why this is not a viable approach and why the BI covariance model is preferable. Specifically, in the limit as $p \to 2$, an FBF $Y(\mathbf{x})$ becomes a linear function of $\mathbf{x}$. To see this mathematically, notice that for p = 2 the FBF covariance of (5) reduces to $Cov[B_2(\mathbf{x}), B_2(\mathbf{x}')] = \mathbf{x}^T\mathbf{x}'$. This implies that the basis functions $Cov[B_2(\mathbf{x}), B_2(\mathbf{x}_i)] = \mathbf{x}^T\mathbf{x}_i$ are linear functions of $\mathbf{x}$, which can be surmised from the right panel of the p = 1.95 case in Figure 4. This also implies that all eigenfunctions (associated with nonzero eigenvalues) of the FBF covariance kernel are linear functions of $\mathbf{x}$. Consequently, using a Karhunen-Loeve representation, it follows that an FBF $Y(\mathbf{x})$ for p = 2 must be itself a linear function of $\mathbf{x}$. This, too, can be surmised from the left panel of the p = 1.95 case in Figure 4, in which the realizations of $Y(\mathbf{x})$ are nearly linear (they would have been exactly linear if we had used p = 2). Consequently, an FBF model with $p \approx 2$ is incapable of modeling anything other than a response surface that is very nearly linear.

There is a related numerical problem with using $p \approx 2$ in an FBF model. Although FBF processes are defined for p = 2 when d = 1, they are not defined for p = 2 when d > 1. When d > 1, using p = 2 in the FBF covariance function will result in a covariance matrix $\mathbf{R}$ that is not positive definite. In fact, Zhang and Apley (2014) have shown that the rank of $\mathbf{R}$ for an FBF model with p = 2 is at most d (regardless of n) and, as a result, $\mathbf{R}$ is extremely rank-deficient for $n >> d$. By continuity of the determinant of a matrix, $\mathbf{R}$ will be nearly singular if $p \approx 2$. See Ranjan et al. (2011) for related discussion of the numerical behavior for the PEXP model for $v \approx 2$.

4.4 Examples and Comparison of Predictive Performance

In this section, we compare the predictive performances of the PEXP and BIPEXP models for a number of examples that have been previously analyzed in the computer metamodeling literature. Our first four examples use models of real physical systems, and the remaining examples use mathematical test functions. In each of the examples, the PEXP and BIPEXP

models were fit to a set of data from one or more designed experiments, and the fitted models were then used to predict the response values at a separate set of test sites. The test prediction root mean square error (RMSE) results are summarized in Table 1 for the real examples and in Table 2 for the mathematical test functions. We also report the ratio RMSE/SD(y) and the percent improvement in test RMSE for the BIPEXP model, defined as ($RMSE_{PEXP}$ $-RMSE_{BIPEXP}$)/$RMSE_{PEXP}$×100. Here, SD(y) is the sample standard deviation of the response at the test sites. Notice that the test $r^2 = 1 - [RMSE/SD(y)]^2$. We prefer the ratio RMSE/SD(y) over r^2 as a measure of fit, because r^2 can be misleading when modeling deterministic computer experiments. For example, an r^2 of 0.99 is considered a nearly perfect fit in many regression modeling contexts, but not necessarily for modeling deterministic computer experiments. As a point of reference, r^2 is 0.9659 for the fit in Figure 1(a) and 0.9854 for the fit in Figure 3.

	PEXP model		BIPEXP model		% Improve-
Example, Design	RMSE	RMSE/SD(y)	RMSE	RMSE/SD(y)	ment
Heat Exchanger	5.133	52.23%	2.1573	21.95%	58.0
G-Protein	0.0162	7.71%	0.0145	6.91%	10.5
Nilson-Kuusk, n = 100	0.0187	9.03%	0.0183	8.84%	2.1
Nilson-Kuusk, n= 150	0.0169	8.43%	0.0168	8.31%	0.6
Borehole, n = 27, OA design	4.645	10.25%	1.6849	3.72%	63.7
Borehole, n = 27, LHD	3.248	7.05%	2.502	5.43%	23.0

Table 1. Comparison of test RMSE values for the PEXP and BIPEXP models for examples involving real physical systems. A positive % Improvement means the BIPEXP model had lower test RMSE than the PEXP model.

	PEXP model		BIPEXP model		% Improve-
Example	RMSE	RMSE/SD(y)	RMSE	RMSE/SD(y)	ment
Xiong et al. (2007)	0.0484	18.47%	0.0317	12.08%	34.5
Function 1	0.5404	19.49%	0.5634	20.32%	-4.1
Function 2	4.4716	52.18%	3.6637	42.75%	18.1
Function 3	0.5026	50.16%	0.5878	58.66%	-14.5
Function 4	0.0416	34.32%	0.047	38.80%	-11.5
Function 5	0.2892	42.13%	0.2924	42.59%	-1.1
Function 6	0.0447	13.79%	0.0269	8.30%	39.8
Function 7	4.1517	7.60%	6.0323	11.05%	-31.2
Function 8	4.72E+03	31.36%	4.42E+03	29.32%	6.5
Function 9	6.82E+04	48.54%	6.60E+04	46.99%	3.2

Table 2. Comparison of test RMSE values for the PEXP and BIPEXP models for examples using mathematical test functions. A positive % Improvement means the BIPEXP model had lower test RMSE than the PEXP model.

Details on the input and response variables, the experimental design, the test sites, the estimated parameters, etc., for all examples are in the online supplement. Briefly, the four real examples are the *heat exchanger simulator* from Qian, Seepersad, Joseph, Allen and Wu (2006) [d = 4 inputs; y = heat transfer rate; 64-run orthogonal-array-based Latin hypercube design (LHD)], the *borehole model* from Morris, Mitchell and Ylvisaker (1993) [d = 8 inputs; y = water flow rate through two aquafers; a 27-run, three-level orthogonal array design and 100 different 27-run LHDs (RMSEs are averaged over the latter)], the *G-protein simulator* from Yi, et al. (2005) [d = 4 inputs; y = concentration of a particular complex in a biochemical reaction; a 41-run, three-level orthogonal array design], and the *Nilson-Kuusk simulator* from Nilson and Kuusk (1989) with data from Bastos and O'Hagan (2009) [d = 5 inputs; y = reflectance of a plant canopy; a 100-run LHD and a 150-run LHD]. For the mathematical test functions, each example involved d = 2 inputs and either a 24-run or 48-run LHD (depending on how difficult the test function was to model). We averaged the test RMSEs across 100 replicates, where on each replicate we generated a different LHD design.

Across the real examples (Table 1), the BIPEXP had test RMSE that was between 0.6% and 63.7% better than the PEXP test RMSE. Zhang and Apley (2014) found that the FBF model performed worse than the PEXP model for the borehole example with LHDs, and they concluded that this was because the FBF model has difficulty in handling smooth response surfaces. The fact that the BIPEXP model performed even better than the PEXP model for this borehole example indicates that, unlike the FBF model, it is capable of handling smooth response surfaces. The results for the mathematical test functions (Table 2) are mixed, with the BIPEXP model performing better for half of the examples, and the PEPXP model performing better for the other half. The average percent improvement for the BIPEXP model across the ten examples in Table 3 is 3.97%, although this number is small relative to the range of variation in the percent improvement (from –31.2% for Function 7 to +39.8% for Function 6).

It may not be a coincidence that the BIPEXP model outperformed the PEXP model for all of the real examples in Table 1, whereas the results were mixed for the mathematical test functions

in Table 2. The mathematical test functions were typically more wiggly and less well-behaved than the response surfaces in the real examples. As discussed in Section 4.1, the sigmoidal basis functions associated with a BIPEXP model are generally better suited to handle well-behaved response surfaces.

4.5 Interpretation of a BI Gaussian Random Field as a Smoothed FBF

As discussed in Section 4.3 and in Zhang and Apley (2014), for smooth response surfaces the FBF model usually does not perform as well as a PEXP model (for the latter, the MLE of ν is usually close to its limiting value of 2, in which case it reduces to the very smooth Gaussian covariance model). We next show that for a certain class of covariance functions $R_c(\bullet)$ for the underlying random field $C(\bullet)$, a BI GRF can be viewed as a smoothed version of an FBF with an appropriate smoothing kernel.

Suppose the stationary covariance $R_c(\bullet)$ is such that $C(\bullet)$ has a white noise integral representation of the form

$$C(\mathbf{x}) = \int_{\mathbb{R}^d} \varphi(\mathbf{z}) W(\mathbf{x}-\mathbf{z}) d\mathbf{z} \tag{21}$$

where $W(\bullet)$ is white noise on $\mathbb{R}^d$ and $\varphi(\bullet)$ is a kernel smoothing function. Substituting (21) into (7) gives

$$\begin{aligned} B_p^C(\mathbf{x}) &= \int_{\mathbb{R}^d} k_{p,d}\left(\|\mathbf{x}-\mathbf{z}\|^{(p-d)/2} - \|\mathbf{z}\|^{(p-d)/2}\right)\left(\int_{\mathbb{R}^d} \varphi(\mathbf{y}) W(\mathbf{z}-\mathbf{y}) d\mathbf{y}\right) d\mathbf{z} \\ &= \int_{\mathbb{R}^d} \varphi(\mathbf{y}) \left[\int_{\mathbb{R}^d} k_{p,d}\left(\|\mathbf{x}-\mathbf{z}\|^{(p-d)/2} - \|\mathbf{z}\|^{(p-d)/2}\right) W(\mathbf{z}-\mathbf{y}) d\mathbf{z}\right] d\mathbf{y} \end{aligned}$$

Defining $\mathbf{z}' = \mathbf{z} - \mathbf{y}$, and introducing a change of variables into the preceding gives

$$\begin{aligned} B_p^C(\mathbf{x}) &= \int_{\mathbb{R}^d} \varphi(\mathbf{y}) \left[\int_{\mathbb{R}^d} k_{p,d}\left(\|\mathbf{x}-\mathbf{y}-\mathbf{z}'\|^{(p-d)/2} - \|\mathbf{y}+\mathbf{z}'\|^{(p-d)/2}\right) W(\mathbf{z}') d\mathbf{z}'\right] d\mathbf{y} \\ &= \int_{\mathbb{R}^d} \varphi(\mathbf{y}) \left[\int_{\mathbb{R}^d} k_{p,d}\left(\|\mathbf{x}-\mathbf{y}-\mathbf{z}'\|^{\frac{p-d}{2}} - \|\mathbf{z}'\|^{\frac{p-d}{2}} + \|\mathbf{z}'\|^{\frac{p-d}{2}} - \|\mathbf{y}+\mathbf{z}'\|^{\frac{p-d}{2}}\right) W(\mathbf{z}') d\mathbf{z}'\right] d\mathbf{y} \\ &= \int_{\mathbb{R}^d} \varphi(\mathbf{y}) B_p^W(\mathbf{x}-\mathbf{y}) d\mathbf{y} - \int_{\mathbb{R}^d} \varphi(\mathbf{y}) B_p^W(-\mathbf{y}) d\mathbf{y}, \end{aligned} \tag{22}$$

where $B_p^W(\bullet)$ is an FBF.

The first term on the right hand side of (22) represents a smoothed version of the FBF $B_p^W(\bullet)$ using the smoothing kernel $\varphi(\bullet)$. The second term is a random variable that is constant over **x** and that can be viewed as ensuring $B_p^C(\mathbf{0}) = 0$. In this regard, a BI GRF can be viewed as a smoothed version of an FBF. The process of smoothing the FBF also smooths the basis functions, which explains why the BIPEXP is better able to represent smooth response surfaces.

4.6 Computational Expense

One drawback of the BI covariance model is that, even though we have reduced the *d*-dimensional integral of (11) for computing the BI covariance to the one-dimensional integral of (14), there is still no general closed-form expression for the BI covariance. The need for one-dimensional integration when computing the covariance substantially increases the computational expense of the BI covariance approach for the MLE model fitting stage (the computational expense of predicting the response at new sites is the same as for a PEXP model if one interpolates a lookup table, as described below). The following are some computational expense comparisons between BIPEXP and PEXP models using a Matlab implementation on Windows laptop with an Intel® Core™ i5-2520M CPU @ 2.50GHz and 2.88 GB RAM. A single evaluation (i.e., for a single **x** and $\mathbf{x}'$ and a single set of parameters) of the BIPEXP covariance $R_B(\mathbf{x}, \mathbf{x}')$ using (12) and (14) usually took between 100—200 times longer than a single evaluation of the PEXP covariance, and this was roughly independent of *d*. A single evaluation of the likelihood function usually took between 30—80 times longer for the BIPEXP model than for the PEXP model, some results for which are shown in Table 3. The relative computational expense did not depend strongly on *n* and *d* for the ranges shown in Table 3. Latin hypercube designs were used for all cases in Table 3.

	n		
d	10	100	1000
1	32.3	45.4	52.7
3	39.6	30.6	36.3
9	43.6	74.8	76.1

Table 3. Computational expense for the BIPEXP model, relative to the PEXP model, for a single evaluation of the likelihood function for various *n* and *d*. The numbers shown are the ratios of the two computation times.

Table 4 shows the total computation time (in seconds) for calculating the MLEs for the PEXP and BIPEXP models in each of the examples in Table 1. For these examples, the BIPEXP model took anywhere between 9 and 176 times longer to fit than the PEXP model.

Example, Design	PEXP	BIPEXP	Ratio
Heat Exchanger	31.4	907.4	28.9
G-Protein	8.8	555.2	63.0
Nilson-Kuusk, n = 100	59.1	4388.1	74.3
Nilson-Kuusk, n= 150	224.7	2120.7	9.4
Borehole, n = 27, OA design	13.8	1232.1	89.6
Borehole, n = 27, LHD	11.3	2001.9	176.4

Table 4. Total computation time (in seconds) for fitting the BIPEXP and PEXP models for the examples in Table 1.

After the MLE stage, a brute force use of (14) when predicting new observations would result in a computational expense that is roughly 100—200 times higher for the BIPEXP model. This follows because a single evaluation of the covariance function typically takes between 100—200 times longer, and the term $\mathbf{R}^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1})$ in the expression for $\hat{y}(\mathbf{x})$ can be calculated in advance, during the MLE stage. However, the computational expense for predicting new observations can be reduced to a level comparable to that for a PEXP model via interpolation of a lookup table for $V_B(H)$, as a function of the scalar distance measure H, for the specific set of parameter MLEs.

More specifically, to calculate the prediction $\hat{y}(\mathbf{x})$ at a new site $\mathbf{x}$ using (17), the vector $\mathbf{R}^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1})$ is calculated in advance, at the last iteration of the MLE stage. For each new site $\mathbf{x}$, the elements of $\mathbf{r}(\mathbf{x})$ are calculated via $Cov[Y(\mathbf{x}), Y(\mathbf{x}_i)] = [V_B(\|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}_1)\|) + V_B(\|\mathbf{\Phi}(\mathbf{x}_i - \mathbf{x}_1)\|) - V_B(\|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}_i)\|)]/2$ with the $V_B(\cdot)$ values interpolated from a lookup table of $V_B(H)$ versus H. The lookup table is created in advance, by evaluating (14) for $H \in [0, H_{max}]$, where H_{max} denotes the largest value of $\|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}_1)\|$, $\|\mathbf{\Phi}(\mathbf{x}_i - \mathbf{x}_1)\|$, or $\|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}_i)\|$ that is of interest. Figure 6 plots $V_B(H)$ versus H for various combinations of $\{p, \nu\}$ for a BIPEXP model with d = 1. From this, it can be seen that $V_B(H)$ is a well-behaved function that is straightforward

to interpolate. It took roughly 0.5 seconds to calculate 1,000 lookup table values for $V_B(\cdot)$ via (14), so this is clearly not computationally prohibitive to do in advance.

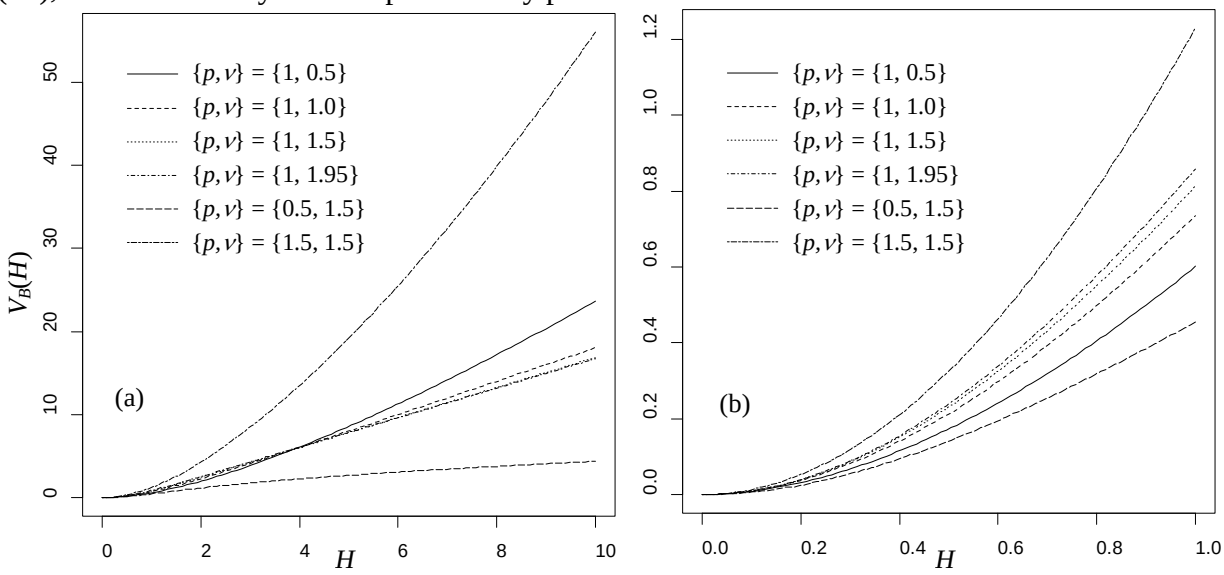

Figure 6. Plot of $V_B(H)$ vs. H for various combinations of $\{p, \nu\}$ for a BIPEXP model with d = 1, over the range (a) $H \in [0,10]$ and (b) $H \in [0,1]$.

5. CONCLUSIONS

In this article, we have developed a new random field model for use when modeling deterministic computer experiment response surfaces via kriging. The BI random field model can be viewed as incorporating an integrator (of the fractional Brownian motion type) into a conventional GRF model with stationary, isotropic covariance, analogous to incorporating the integrator in an ARIMA time series model. The presence of the integrator is what results in sigmoidal shaped basis functions, as opposed to the localized basis functions associated with more standard covariance models like PEXP. We have shown that the BI covariance can be evaluated via the one-dimensional integral of (14), regardless of the dimension of the input space and the covariance model chosen for the underlying random field $C(\bullet)$. Although we have focused on the BIPEXP model in our examples, the BI random field model can be used with any stationary, isotropic covariance model for $C(\bullet)$ (anisotropy is handled by incorporating lengthscale parameters into the isotropic BI covariance model).

For each of the real examples in Table 1, the BIPEXP model resulted in an improvement in test RMSE over the PEXP model. For the mathematical test functions in Table 2, the results were

mixed, with the BIPEXP model performing better on some examples, and the PEPX model performing better on others. This may relate to what is perhaps the most fundamental difference between a BI covariance model versus a covariance model without in integrator – sigmoidal versus localized basis functions. As mentioned in Section 4.1, a common perception in the machine learning field is that nicely-behaved response surfaces are usually more easily modeled with sigmoidal basis functions (as in a neural network) than with localized basis functions (as in a radial basis function network). The mathematical test functions in Table 2 were generally more wiggly and less nicely-behaved than the response surfaces for the real examples in Table 1. This could explain why the BIPEXP model performed better than the PEXP model for the real examples, whereas the results were mixed for the mathematical test functions.

Without prior knowledge of how nicely-behaved the response surface is, it is difficult to know in advance which model is more appropriate. In machine learning, the standard strategy is to try different models and use appropriate model selection methods (e.g., AIC, cross-validation, etc.) to select the best model. In time series modeling, this is also the standard strategy for determining whether an ARMA or ARIMA model is more appropriate. For computer response surface metamodeling, a similar strategy of trying both models (a BI covariance model and its non-integrated counterpart) and selecting the more appropriate one may be a quite reasonable approach. We are currently investigating model selection methods for accomplishing this reliably and efficiently.

Because there is no closed-form expression for the BIPEXP covariance function, and numerical integration is required, the computational expense of the MLE model fitting stage is much higher than when using a PEXP model (roughly 50 times higher on average for our implementation). However, using the simple lookup table interpolation procedure described in Section 4.6, the computational expense when predicting the response at new input sites is essentially the same as for a PEXP model. We are currently investigating methods for reducing the computational expense for the MLE stage to something that is more on par with a PEXP model. For example, if one is content with allowing only a discrete set of choices for $\{p, \nu\}$, then

one could calculate lookup tables in advance for $V_B(H)$ versus H for each combination $\{p, \nu\}$. Subsequently, when fitting the model in the MLE stage, these functions could be interpolated to quickly calculate the BIPEXP covariance without numerical integration. We are also investigating whether (14) reduces to a closed-form expression for any special cases of underlying covariance models for $R_C(\bullet)$.

SUPPLEMENTARY MATERIALS

The online supplementary materials contain code for implementing the BI-covariance calculations, some discussion on the relationship between the PEXP and BIPEXP models, a recursive expression for the hypergeometric function needed in (14), details on the examples in Sec. 4.4, and additional basis functions and GRF realizations analogous to Figures 4 and 5.

ACKNOWLEDGEMENT

This work was supported by the National Science Foundation under grants #XXX and #XXX.

REFERENCES

Ba, S. and V. R. Joseph (2012), "Composite Gaussian Process Models for Emulating Expensive Functions," *Annals of Applied Statistics*, **6**(4), 1838-1860.

Barron, A. R., (1993), "Universal Approximation Bounds for Superpositions of a Sigmoidal Function," *IEEE Transactions on Information Theory*, 39(3), pp. 930—944.

Bastos, L. S. and O'Hagan, A. (2009), "Diagnostics for Gaussian Process Emulators," *Technometrics*, **51**, 425-438.

Booker, A. J., Dennis, J. E., Jr., Frank, P. D., Serafini, D. B., Torczon, V., and Trosset, M. W. (1999), "A Rigorous Framework for Optimization of Expensive Function by Surrogates," *Structural Optimization*, **17**, 1-13.

Cressie, N. A. (1990), "The Origins of Kriging," *Mathematical Geology*, **22**(3), 239-252.

Cressie, N. A. (1991), *Statistics for Spatial Data*, New York: Wiley.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D., (1988), "A Bayesian Approach to the Design and Analysis of Computer Experiments", ORNL-6498, available from http://www.ornl.gov/~webworks/cpr/rpt/6863.pdf, National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

Currin, C., Mitchell, T., Morris, M. and Ylvisaker, D. (1991), "Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer

Experiments," *Journal of the American Statistical Association*, **86**(416), 953-963.

Fang, K. T., Li, R. and Sudjianto, A. (2006), *Design and Modeling for Computer Experiments*, Boca Raton, FL: Chapman & Hall/CRC.

Flake, G. W. (2012), "Square Unit Augmented, Radially Extended, Multilayer Perceptrons," in *Neural Networks: Tricks of the Trade, 2nd ed.*, Montavon, G., Orr, G. B., and Muller, K. (Eds.), LNCS 7700, pp. 143–161, 2012, Springer-Verlag Berlin Heidelberg.

Fuentes, M., Chen, L. and Davis, J. (2008), "A Class of Nonseparable and Nonstationary Spatial Temporal Covariance Functions," *Environmetrics*, **19**, 487-507.

Gramacy, R. B. and Lee, H. K. H. (2008), "Bayesian Treed Gaussian Process Models with an Application to Computer Modeling," *Journal of the American Statistical Association*, **103**, 1119-1130.

Gramacy, R. B. and Lee, H. K. H. (2012), "Cases for the Nugget in Modeling Computer Experiments," *Statistics and Computing*, **22**, 713-722.

Huang, W., Wang, K., Jay Breidt, F. and Davis, R. A. (2011), "A Class of Stochastic Volatility Models for Environmental Applications," *Journal of Time Series Analysis*, **32**, 364–377.

Jin, R., Chen, W. and Sudjianto, A. (2002), "On Sequential Sampling for Global Metamodelling in Engineering Design," *Proceedings of the ASME Design Automation Conference*, Paper Number DETC2002/DAC-34092, Montreal, Canada, September-October 2002.

Joseph, V. R. (2006), "Limit Kriging," *Technometrics*, **48**(4), 458-466.

Joseph, V. R., Hung, Y. and Sudjianto, A. (2008), "Blind Kriging: A New Method for Developing Metamodels," *Journal of Mechanical Design*, **130**, 031102-1-8.

Lee, C. C., Chung, P. C., Tsai, J. R., and Chang, C. I. (1999), "Robust Radial Basis Function Neural Networks," *IEEE Transactions on Systems, Man, and Cybernetics*, 29(6), pp. 674—685.

Li, R. and Sudjianto, A. (2005), "Analysis of Computer Experiments Using Penalized Likelihood in Gaussian Kriging Models," *Technometrics*, **47**, 111-120.

Lindstrom, T. (1993), "Fractional Brownian Fields as Integrals of White Noise," *Bull. London. Math. Soc.*, **25**, 83-88.

Loeppky, J. L., Sacks, J. and Welch, W. J. (2009), "Choosing the Sample Size of a Computer Experiment: A Practical Guide," *Technometrics*, **51**, 366-376.

Morris, M. D., Mitchell, T. J. and Ylvisaker, D. (1993), "Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction," *Technometrics*, **35**, 41-47.

Nilson, T. and Kuusk, A. (1989), "A reflectance Model for the Homogeneous Plant Canopy and Its Inversion," *Remote Sensing of Environment*, **27**, 157-167.

Paciorek, C. J. (2003), “Nonstationary Gaussian Process for Regression and Spatial Modelling,” *Ph.D. Dissertation*, Carnegie Mellon University, Pittsburgh, PA, U.S.A.

Paciorek, C. and Schervish, M. (2006), “Spatial Modelling Using a New Class of Nonstationary Covariance Function,” *Environmetrics*, **17**, 483-506.

Qian, Z., Seepersad, C., Joseph, R., Allen, J. and Wu, C. F. J. (2006), “Building Surrogate Models with Detailed and Approximate Simullations,” *ASME Journal of Mechanical Design*, **128**, 668-677.

Ranjan, P., Haynes, R. and Karsten, R. (2011), “A Computationally Stable Approach to Gaussian Process Interpolation of Deterministic Computer Simulation Data,” *Technometrics*, **53**, 366-378.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989), “Design and Analysis of Computer Experiments,” *Statistical Science*, **4**(4), 409-435.

Sampson, P. D. and Guttorp, P. (1992), “Nonparametric Estimation of Nonstationary Spatial Covariance Structure,” *Journal of the American Statistical Association*, **87**, 108-119.

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

Sasena, M. J., Papalambros, P. Y. and Goovaerts P. (2002), "Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization," *Engineering Optimization*, 34(3), 263–278.

Schmidt, A. M. and O’Hagan, A. (2003), “Bayesian Inference for Nonstationary Spatial Covariance Structure via Spatial Deformations,” *Journal of the Royal Statistical Society, Series B*, **65**, 745-758.

Staum, J. (2009), "Better Simulation Metamodeling: The Why, What, and How of Stochastic Kriging," *Proceedings of the 2009 Winter Simulation Conference*, ed. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Wiley.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J. and Morris, M. D. (1992), “Screening, Predicting, and Computer Experiments,” *Technometrics*, **34**, 15-25.

Wolfram (2014), The symbolic integration tool at integrals.wolfram.com or the Mathematica 10.0 symbolic manipulation software from Wolfram Research, Inc.

Xiong, Y., Chen, W., Apley, D. W. and Ding, X. (2007), “A Nonstationary Covariance Based Kriging Method for Metamodeling in Engineering Design,” *International Journal for Numerical Methods in Engineering*, **71**(6), 733-756.

Zhang, N. and Apley, D. W. (2014), “Fractional Brownian Fields for Response Surface Metamodeling,” *Journal of Quality Technology*, 46(4), 285—301.