Fractional Brownian Fields for Response Surface Metamodeling

Ning Zhang[1] and Daniel W. Apley[2]

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

ABSTRACT

Kriging, a widely used metamodeling method for computer simulation data, models the response surface as a realization of a random field. Stationary covariance functions such as the Gaussian, power exponential, or Matérn class are the most common choice for the underlying random field model. Nonstationary versions of these same covariance functions with scale parameters that vary spatially are also sometimes used. Fractional Brownian fields (FBFs) are a different form of nonstationary random field model having stationary increments, an example of more general intrinsic stationary processes. Although FBFs have been considered for intrinsic kriging in the spatial statistics literature, they have received little attention for computer simulation response surface metamodeling. For use in the latter context, we argue that they have some attractive (as well as some unattractive) properties that mitigate certain problems inherent to many stationary covariance models, such as reversion to the mean, numerical issues due to near-singularity of covariance matrices, and difficulties in handling abrupt response surface features.

KEY WORDS: Computer experiments, Gaussian random fields, Interpolation, Intrinsic Stationary, Kriging, Prediction intervals

[1] Dr. Zhang is an Analyst at Credit Suisse. His email is NingZhang2013@u.northwestern.edu.
[2] Corresponding author. Dr. Apley is a Professor of Industrial Engineering & Management Sciences at Northwestern University. His email is apley@northwestern.edu.

1. Introduction

Kriging is a popular method for interpolating the value of a function of interest at unobserved locations, based on a set of values at observed locations. It originated in geostatistical analyses (Matheron 1963) but also found use in other applications of spatial statistics (Cressie 1991). The proliferation of computer experiments (aka simulation) as an analysis tool in virtually all branches of the physical sciences and engineering gave birth to one of the most widespread application of kriging, namely, metamodeling data from computer experiments. Introduced for this purpose by Currin et al. (1988) and Sacks, Welch, Mitchell and Wynn (1989), the kriging approach has found its way into many textbooks on the topic of metamodeling output of computer experiments (Santner, Williams and Notz 2003, Fang, Li and Sudjianto 2006).

Kriging entails viewing the response surface as a realization of a spatial random field and then using statistical methods to predict (usually the best linear unbiased predictor or the posterior mean predictor in a Bayesian framework) the response at unobserved locations, given the observed values. Although Gaussian random fields (GRFs) with stationary covariance are the most common choice of model (Santner, Williams and Notz 2003), they have certain characteristics that may be undesirable for many response surface metamodeling applications, which we illustrate with the following example.

Consider a simple example from Xiong et al. (2007), also considered in Ba and Joseph (2012). Suppose the unknown "computer simulation model" is the function $y(x) = \sin(30(x - 0.9)^4)\cos(2(x - 0.9)) + (x - 0.9)/2$, where x denotes an input (aka design) variable, and y denotes an output (aka response) variable. Suppose also that we have observed the simulation response at the 17 input sites shown in Figure 1, of which 13 are evenly spaced over the region [0, 0.36] and 4 are evenly spaced over [0.4, 1]. Consider the popular Gaussian covariance function $R(x, x') = Cov[Y(x), Y(x')] = \sigma^2\exp[-\theta(x-x')^2]$, where $\theta > 0$ and $\sigma^2 > 0$ are the prior correlation and variance parameters, respectively. Using a constant prior mean μ (see Section 2 for details on the model parameters and estimation), the maximum likelihood estimates (MLEs) of the parameters are

$\hat{\theta} = 519.9$, $\hat{\mu} = -0.216$, $\hat{\sigma}^2 = 0.0735$, and the resulting kriging predictor [$\hat{y}(x)$ from Equation (2) in Section 2] is plotted in Figure 1(a) with 95% prediction intervals (PIs) also shown. Notice that $\hat{y}(x)$ reverts to the mean $\hat{\mu}$ as x moves away from the input sites, either interpolating between points or extrapolating beyond the experimental range. Similar results are observed if we use the power exponential covariance function $R(x, x') = Cov[Y(x), Y(x')] = \sigma^2 \exp[-\theta|x-x'|^p]$ ($1 \le p \le 2$), which is another popular stationary covariance function that gives additional flexibility. Again, all parameters were estimated by MLE ($\hat{p} = 1.348$, $\hat{\theta} = 25.3$, $\hat{\mu} = -0.158$, and $\hat{\sigma}^2 = 0.0837$), and the results are shown in Figure 1(b). The problem of $\hat{y}(x)$ reverting to the mean is less severe than that shown in Figure 1(a), but still substantial.

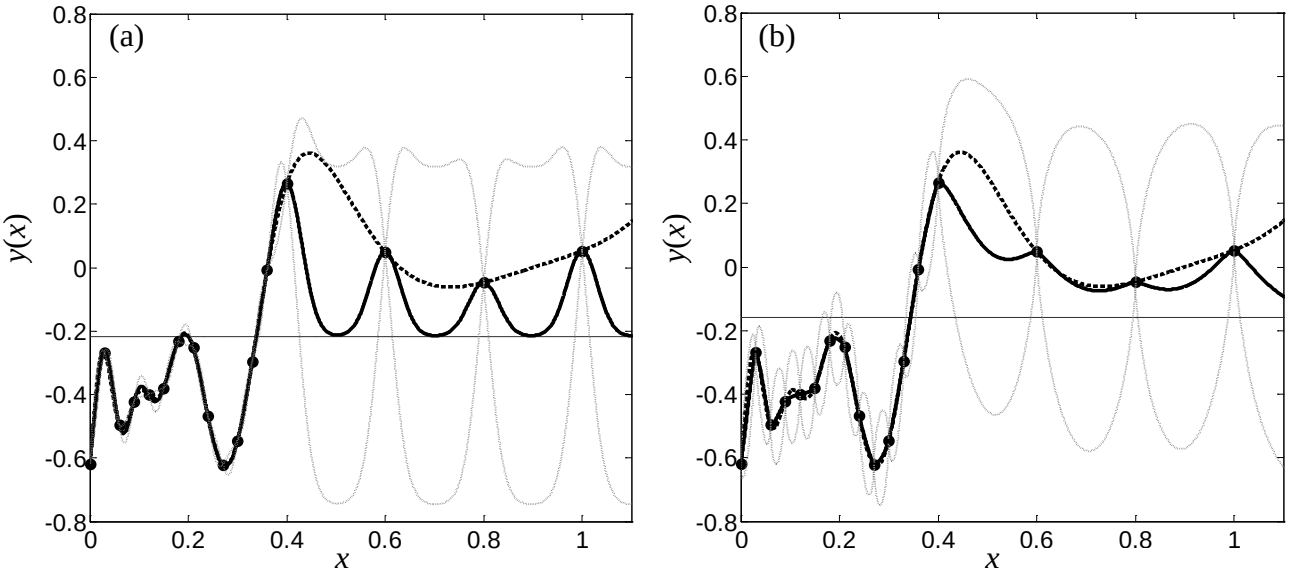


Figure 1. Plot of the function y(x) = sin(30(x – 0.9)^4)cos(2(x – 0.9)) + (x – 0.9)/2 (···), the kriging predictor $\hat{y}(x)$ (—) and 95% prediction intervals (gray dashed lines) with (a) Gaussian covariance function and (b) power exponential covariance function. The fitted mean $\hat{\mu}$ is the horizontal dashed line.

This undesirable feature of kriging with a covariance-stationary GRF has been noted by Joseph (2006) and others, and has been referred to as "reversion to the mean" by Staum (2009). Adding new input sites and collecting more observations in the region $x \in [0.4, 1]$ will certainly help to mitigate this problem. But one might argue that the input sites are already nearly dense enough to capture the overall trend of the response surface in the region $x \in [0.4, 1]$, and the prediction with the FBF model (shown later, in Figure 8) looks quite reasonable. Additionally, input sites are inherently sparser in higher dimensional problems, in which case it may be

difficult to avoid reversion to the mean by conducting simulations at a "sufficiently dense" set of input sites.

Although a $\hat{y}(\mathbf{x})$ (here $\mathbf{x}$ is a vector of input variables) that reverts to the mean may be reasonable for some applications, for many applications it is not. For example, Figure 2 shows a typical stress-strain curve for low carbon steel, which exhibits a linear trend in the elastic region (up to the yield point A), after which the relationship becomes nonlinear in the plastic deformation region. The dashed horizontal line is the MLE of $\hat{\mu}$ for a stationary Gaussian covariance function fitted to the nine evenly-spaced strain sites shown in Figure 2. A predictor of stress that reverts to the mean $\hat{\mu}$ for strain values beyond the right-most point in Figure 2 would be clearly inconsistent with the type of behavior one would expect.

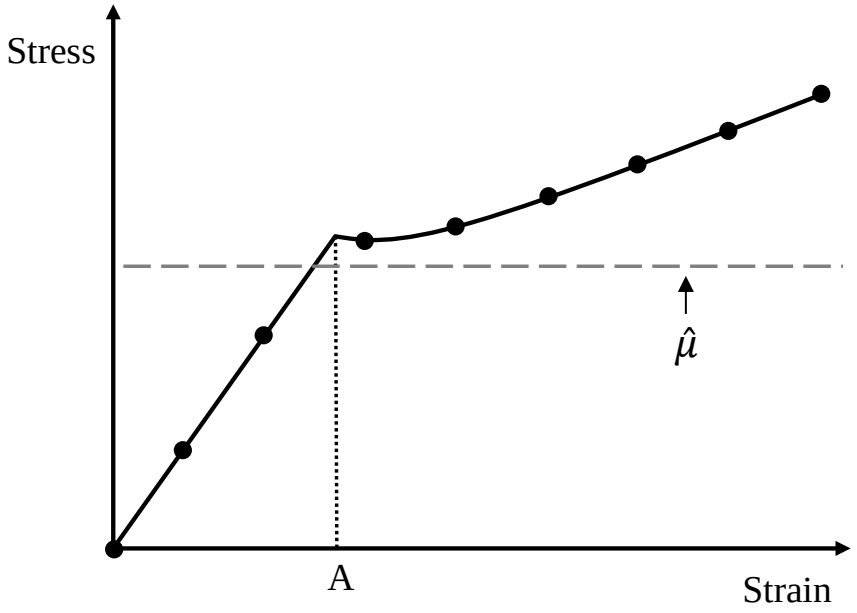


Figure2. A stress-strain curve that is typical of low-carbon steel.

A kriging predictor that does not revert to the mean (whether the mean is a constant or some global polynomial trend) is characteristic of a certain class of GRFs with nonstationary covariance. Here, we distinguish between two different types of covariance-nonstationarity. The first type, on which this paper focuses, is the form of nonstationarity associated with processes such as Brownian motion, in which the increments $y(x+\Delta_x) - y(x)$ are stationary in x for fixed Δ_x, but the variance of $y(x)$ may grow unbounded as x grows. Brownian motion is an example of

what is referred to as intrinsic stationarity in the spatial statistics literature (Cressie, 1991; Wackernagel, 2003). Readers familiar with time series forecasting may recognize this form of nonstationarity as that associated with the "integrator" component of an autoregressive integrated moving average (ARIMA) model, which has stationary increments. As one forecasts further into the future, ARMA forecasts revert asymptotically to the long-term mean, whereas the ARIMA forecasts do not. The second type of covariance nonstationarity is intended to model processes whose smoothness varies spatially by modifying standard stationary covariance functions to have parameters that vary spatially (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). Unlike the first (Brownian) type of nonstationarity models, which have stationary increments and parameters that do *not* vary spatially, the second type have nonstationary increments and are analogous to an ARMA model whose parameters vary over time, for which the predictions will still revert to the mean for the reasons discussed in Section 4.2.

The preceding suggests that a Brownian type of covariance-nonstationary model may be more appropriate than a covariance-stationary model for response surface applications in which reversion to the mean is undesirable. For a continuous, one-dimensional x (the spatial domain variable), the white noise integration that defines Brownian motion is the concept analogous to the integrator in an ARIMA time series model. For a multidimensional $\mathbf{x}$, a natural generalization that we consider in this paper is a fractional Brownian field (FBF) model, which further generalizes Brownian motion in terms of an index coefficient (p, defined in Section 3). FBFs have been widely studied in the past, including in the context of random field regression (e.g., kriging). Mandelbrot and Van Ness (1968) considered fractional Brownian motion (one-dimensional x) and advocated its use in time series prediction. They discussed a number of practical issues. However, the vast majority of the extensive literature on FBFs focuses on their mathematically rich properties (e.g., Lindstrom 1993). FBF models have also received some attention in the context of intrinsic kriging in the geostatistics literature, as an example of an

intrinsic stationary process for which the variogram (as opposed to the covariance function) is typically used as the basis for the predictor (see, e.g., Cressie, 1991). The term intrinsic kriging refers to a common implementation of the kriging predictor when an intrinsic stationary process is assumed (Matheron, 1973; Christensen, 1990). In intrinsic kriging, one works with weighted differences of the response observations, where the weights are chosen to cancel the effects of a deterministic trend in the data (analogous to how direct differencing cancels the effects of a linear trend with evenly spaced observations). This allows the covariance parameters to be estimated in a manner that is unaffected by the presence of the deterministic trend. If one assumes a (generalized) covariance function in intrinsic kriging that corresponds to a particular covariance function (e.g., Matérn, power exponential, FBF, etc) in regular kriging with the same parameters, then the intrinsic kriging predictor is the same as the regular kriging predictor, and the only difference lies in how the covariance parameters are estimated (Christensen, 1990; Cressie, 1991, Section 5.4). Although there has been prior work that has used FBF models in kriging or intrinsic kriging, to the best of our knowledge none of the prior work has explored the issues considered in this paper that are relevant to modeling deterministic computer experiment data, such as their inherent ability to avoid reversion to the mean and their good numerical conditioning.

The main objective of this paper is to call attention to FBFs as an attractive alternative for many response surface metamodeling applications. The format of the remainder of the paper is as follows. In Section 2 we give a brief review of kriging for metamodeling and discuss the problem of reversion to the mean in more detail. In Section 3 we review basic FBF concepts and discuss nuances of implementing kriging with FBF models. Section 4 discusses various characteristics of FBFs for kriging, and Section 5 concludes the paper.

2. Background on GRFs for Metamodeling and Reversion to the Mean

Let $\mathbf{x} = [x_1, x_2, \ldots, x_d]^T$ denote the set of input variables, and let $y(\mathbf{x})$ denote the scalar response function. Suppose we evaluate $y(\mathbf{x})$ at a set of n distinct input sites $\mathbf{S} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$

and observe a set of n response values $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T = [y(\mathbf{x}_1), y(\mathbf{x}_2), \ldots, y(\mathbf{x}_n)]^T$. The basic idea behind kriging is to model the response surface $y(\mathbf{x})$ as a realization of a random field $Y(\mathbf{x})$ over the d-dimensional input variable space. For covariance-stationary GRFs, which are perhaps the most commonly used random field models for kriging, one assumes some parametric mean function $E[Y(\mathbf{x})] = \mathbf{h}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}$ and covariance function $R(\mathbf{x}, \mathbf{x}') = Cov[Y(\mathbf{x}), Y(\mathbf{x}')]$. Here, $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_q]^T$ is a vector of unknown parameters, and $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_q(\mathbf{x})]^T$ is a vector of known basis functions (e.g., linear or quadratic). One common choice of covariance function parameterization is the so-called Gaussian covariance

$$R(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{i=1}^{d} \exp\{-\theta_i (x_i - x_i')^2\}, \tag{1}$$

where σ^2 is the prior variance of $Y(\bullet)$, and $\{\theta_i: i = 1, 2, \ldots, d\}$ are the correlation parameters that dictate the smoothness of the GRF. The kriging predictor $\hat{y}(\mathbf{x})$ of $y(\mathbf{x})$ is defined as the best linear unbiased predictor of $Y(\mathbf{x})$ given $\mathbf{y}$. For GRFs, if the covariance parameters are treated as known, the kriging predictor is equivalent to the posterior mean of $Y(\mathbf{x})$ given $\mathbf{y}$ (marginalized with respect to a noninformative prior on $\boldsymbol{\beta}$), which is (Currin, Mitchell, Morris and Ylvisaker, 1991)

$$\hat{y}(\mathbf{x}) = \mathbf{h}^T(\mathbf{x})\widehat{\boldsymbol{\beta}} + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{H}\widehat{\boldsymbol{\beta}}\right). \tag{2}$$

Here, $\mathbf{H}$ is an $n \times q$ matrix whose ith row is $\mathbf{h}^T(\mathbf{x}_i)$, $\mathbf{R}$ is an $n \times n$ matrix whose (i,j)-th element is $R(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{r}(\mathbf{x})$ is an $n \times 1$ vector whose ith element is $R(\mathbf{x}, \mathbf{x}_i)$, and $\widehat{\boldsymbol{\beta}} = [\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}$ is a point estimate of $\boldsymbol{\beta}$. The preceding approach is referred to as ordinary kriging for $\mathbf{h}(\mathbf{x}) = 1$ and universal kriging for general $\mathbf{h}(\mathbf{x})$ (Cressie 1990).

The reversion to the mean problem described in Section 1 is inherent to ordinary or universal kriging with stationary covariance. To see this, rewrite Equation (2) as

$$\hat{y}(\mathbf{x}) = \mathbf{h}^T(\mathbf{x})\widehat{\boldsymbol{\beta}} + \sum_{i=1}^{n} c_i \cdot g_i(\mathbf{x}) \tag{3}$$

where $g_i(\mathbf{x}) = Cov(Y(\mathbf{x}), Y(\mathbf{x}_i)) = R(\mathbf{x}, \mathbf{x}_i)$ $(i = 1, 2, \ldots, n)$ are additional basis functions with coefficients c_i (determined by Eq. 2) that are not functions of $\mathbf{x}$. From Equation (3) it is clear why we have reversion to the mean. For a stationary covariance, each $Cov(Y(\mathbf{x}), Y(\mathbf{x}_i))$ term decays to

zero as $\mathbf{x}$ strays further from the input sites $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, in which case $\hat{y}(\mathbf{x})$ reverts to the estimated mean function $\mathbf{h}^T(\mathbf{x})\hat{\boldsymbol{\beta}}$ (which is $\hat{\mu}$ in Figures 1 and 2).

With a stationary GRF model and a fixed set of input sites, the only way to mitigate reversion to the mean is to use small correlation parameters $\{\theta_i: i = 1, 2, \ldots, d\}$. With small correlation parameters, $Cov(Y(\mathbf{x}), Y(\mathbf{x}_i))$ decays to zero more slowly, in which case $\hat{y}(\mathbf{x})$ reverts to the estimated mean more slowly. However, steering the correlation parameters to values smaller than the MLE (e.g., using the approach of Li and Sudjianto, 2005, or a Bayesian approach with an informative prior) may have the following undesirable side effects. In addition to providing a predicted response surface, kriging also provides a convenient quantification of the prediction uncertainty via the prediction mean square error (MSE):

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

which coincides with the posterior variance in a Bayesian framework (Currin, Mitchell, Morris and Ylvisaker, 1991). Choosing overly small correlation parameters to mitigate reversion to the mean may also result in substantial underestimation of the prediction uncertainty. It may also result in numerical problems due to a nearly singular $\mathbf{R}$.

Other potential solutions to the reversion to the mean problem have been proposed in the literature. The method proposed by Joseph (2006) iteratively fits $\hat{y}(\mathbf{x})$, where on each iteration the prior mean function is taken to be $\hat{y}(\mathbf{x})$ from the previous iteration. Although this approach effectively avoids reversion to the mean, a drawback is that it can give an odd sigmoidal shaped $\hat{y}(\mathbf{x})$ when interpolating between two input sites. Although the method of Joseph, Hung and Sudjianto (2008) does not mitigate reversion to the mean, it does mitigate its adverse effects by using a more extensive set of basis functions $\{h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots\}$ for the prior mean, with variable selection used to select a subset of these. Hence, $\hat{y}(\mathbf{x})$ will revert to a more complex fitted mean. The use of variable selection invalidates the standard kriging MSE expression, so that calculation of prediction intervals (PIs) becomes unclear. Moreover, one needs to choose an appropriate class of sufficiently flexible basis functions. We believe the FBF kriging approach of this paper

is a more attractive solution to the reversion to the mean problem, a solution that is more parsimonious, as well as more flexible, and that operates within the standard kriging paradigm, thereby allowing standard kriging expressions for quantifying prediction uncertainty. Moreover, it exhibits a number of other desirable characteristics that we discuss in Section 4.

Chipman, Ranjan and Wang (2012) developed a quite different approach for nonparametric metamodeling of deterministic computer simulations via Bayesian additive regression tree (BART) models. Although the BART based surrogate model of Chipman et al. (2012) generally avoids reversion to the mean, it is not a kriging predictor, and it can give a rough, irregular predicted response surface.

3. FBFs and their use for Kriging

A standard FBF of index $0 < p < 2$ is a random field $Y: R^d \rightarrow R$ defined via (Lindstrom, 1993)

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

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

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

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

(iv) sample paths are continuous with probability one.

Note that Lévy-Brownian motion is a special case of a FBF with $p = 1$, and Brownian motion is a special case of a FBF with $p = 1$ and $d = 1$.

Aside from a few nuances, discussed below, using an FBF model for kriging is straightforward. The covariance function $R(\mathbf{x}, \mathbf{x}') = Cov\left(Y(\mathbf{x}), Y(\mathbf{x}')\right)$ from Equation (5) can be substituted directly into Equations (2) and (4). We use a constant prior mean function, i.e., we model $Y(\mathbf{x})$ as an FBF plus an unknown constant $\mathbf{h}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta} = \mu$. Any more complex prior mean function seems entirely unnecessary, because there is no reversion to the mean with an FBF model (also see Section 4).

Although the covariance function (5) depends on choice of origin, it is well known (see Cressie, 1991, or Christensen, 1990) that the kriging predictor can be expressed equivalently in terms of the variogram of $Y(\mathbf{x})$, which does not depend on choice of origin. Rather than work with the variogram $\|\mathbf{x}_1 - \mathbf{x}_2\|^p$, as is typically done in intrinsic kriging, we prefer to work within the more familiar framework of kriging based on the covariance function. To this end, we choose an arbitrary input site (denoted by $\mathbf{x}_0$) from the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and consider the corresponding response y_0 to be perfect "prior" knowledge of μ. We then represent the response surface as a translated FBF $Y(\mathbf{x}) = Z(\mathbf{x} - \mathbf{x}_0) + y_0$ with $Z(\cdot)$ a standard FBF.

Although the prior distribution of $Y(\mathbf{x})$ clearly depends on which input site $\mathbf{x}_0$ is chosen as the origin, this choice is arbitrary in terms of its effect on the kriging predictor, because of the following. Harville (1974) showed that for $Y(\mathbf{x}) = \mu + Z(\mathbf{x})$ with $Z(\cdot)$ any zero-mean Gaussian process, the joint distribution of the contrasts $\{y_i - y_j: \ 1 \le i \le n, i \ne j\}$, as a function of $\mathbf{y}$ and the covariance parameters, is independent of μ and depends on j only up to constant multiplicative factor. Furthermore, because the increments of an FBF model are stationary, the joint distribution of $\{y_i - y_j: \ 1 \le i \le n, i \ne j\}$ is independent of any translation of the origin or any translation of $Y(\mathbf{x})$ by adding to it a constant. Combining these two facts, it follows that the joint distribution of $\{y_i - y_j: \ 1 \le i \le n, i \ne j\}$ under the translated FBF model $Y(\mathbf{x}) = Z(\mathbf{x} - \mathbf{x}_j) + y_j$ only depends on j up to constant multiplicative factor. Hence, the MLE of the covariance parameters does not depend on which input site is chosen as the origin. Similarly, if we augment the set $\{y_i - y_j: \ 1 \le i \le n, i \ne j\}$ by $Y(\mathbf{x}) - y_j$ for any $\mathbf{x}$ of interest and apply the preceding arguments, it follows that the posterior distribution of $Y(\mathbf{x})|\{y_1, y_2, \ldots, y_n\}$ with fixed covariance parameters does not depend on which input site is chosen as the origin. Consequently, in our approach the kriging predictor does not depend on choice of origin.

As notational convention, n will denote the total number of input sites (including the one chosen for the origin), $\mathbf{y}$ will denote the vector of $n-1$ remaining observations, $\mathbf{R}$ will denote the $(n-1)\times(n-1)$ covariance of $\mathbf{y}$, etc. Note that with y_0 treated as prior knowledge, the $n \times n$ covariance matrix of all n observations is singular.

Although the parameter p also has implications regarding the roughness of Y (which we discuss further in Section 4), it would be useful to have a separate parameter for each input coordinate to represent different levels of roughness in different coordinate directions. This is an important issue in kriging because the roughness, which is dictated by the covariance function for GRFs, plays a key role in the response prediction. Intuitively, the rougher the response surface, the more quickly their correlation decays as the distance between two points grows, and the less predictable the surface will be. It is also important that the estimation of the roughness parameters is tractable. For FBF models, anisotropic roughness can be handled by rescaling the spatial coordinates via

$$Y(\mathbf{x}) = Z[\mathbf{\Phi}(\mathbf{x} - \mathbf{x}_0)] + y_0 \tag{6}$$

with $Z(\bullet)$ a standard FBF and $\mathbf{\Phi} = \text{diag}\{\phi_1, \phi_2, \ldots, \phi_d\}$ with each $\phi_i > 0$. A larger ϕ_i means the response surface is rougher in ith spatial coordinate. We henceforth refer to $\mathbf{\Phi}$ as the scale parameters, which play a role that is somewhere between the roles of $\{\theta_i: i = 1, 2, \ldots, d\}$ and of σ in the Gaussian covariance function of Equation (1).

With the preceding scaling and translation convention, the FBF kriging predictor is

$$\hat{y}(\mathbf{x}) = y_0 + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \mathrm{y}_0 \cdot \mathbf{1}), \tag{7}$$

where $\mathbf{1}$ denotes a column vector of n–1 ones. The covariance function that comprises the elements of $\mathbf{r}(\mathbf{x})$ and $\mathbf{R}$ is

$$Cov[Y(\mathbf{x}), Y(\mathbf{x}')] = \frac{1}{2}\{\|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}_0)\|^p + \|\mathbf{\Phi}(\mathbf{x}' - \mathbf{x}_0)\|^p - \|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}')\|^p\}. \tag{8}$$

To use any GRF model for kriging, one must either choose or estimate each parameter. For FBF models, the parameters are the diagonal $\mathbf{\Phi}$ and p. The likelihood function for estimating $\{\mathbf{\Phi}, p\}$ is

$$L(\mathbf{\Phi}, p; \mathbf{y}) = \frac{1}{(2\pi)^{(n-1)/2}|\mathbf{R}|^{1/2}} \exp\left\{-\frac{1}{2}[\mathbf{y} - y_0 \cdot \mathbf{1}]^T\mathbf{R}^{-1}[\mathbf{y} - y_0 \cdot \mathbf{1}]\right\} \tag{9}$$

where $[\mathbf{R}]_{i,j} = Cov\left[Y(\mathbf{x}_i), Y\left(\mathbf{x}_j\right)\right]$, which is quite tractable and straightforward to maximize over $\mathbf{\Phi}$ and p. As we discuss in Section 4.4, it is also less affected by the numerical problems that plague many kriging analyses.

From the definition of FBFs, p is restricted to lie in the open interval (0, 2). However, we recommend further restricting p to the interval [1,2). For $p < 1$, $\hat{y}(\mathbf{x})$ has an undesirable peaked shape, which we illustrate in Section 4.1. The strict inequality $p < 2$ is not just a mathematical curiosity. Although for $d = 1$ FBF processes are defined for $p = 2$, for $d > 1$ they are not. Unlike for the power exponential covariance function, which remains a positive definite covariance function for $p = 2$ (Ranjan et al., 2011), using $p = 2$ in the FBF covariance function will result in covariance matrices that are not positive definite if $n-1 > d$, which is usually the case. This follows by noticing that the standard FBF covariance structure is $R(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T\mathbf{x}'$ for $p = 2$, for which $\mathbf{R}$ is the Gramm matrix, under the inner product $\mathbf{x}^T\mathbf{x}'$, of the set of d-dimensional input sites $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-1}\}$. Therefore, $\mathbf{R}$ cannot have rank greater than d and will be extremely rank-deficient for $n >> d$. Moreover, the entire covariance function $R(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T\mathbf{x}'$ is d-dimensional for $p = 2$, which implies that the process $Y(\bullet)$ can be represented as a linear combination of the d eigenfunctions (which are each linear in $\mathbf{x}$) associated with the d nonzero eigenvalues. This creates a number of practical implementation difficulties: $\hat{y}(\mathbf{x})$ will no longer interpolate all the observations; and, unless the response truly is a linear function of $\mathbf{x}$, the data will be inconsistent with the model, so that likelihood-based estimation cannot be implemented.

Because $R(\mathbf{x}, \mathbf{x}')$ is continuous in p, numerical problems may also surface if p is too close to 2. Hence, we recommend placing an upper bound on p that is slightly less than 2, the most appropriate value of which may depend on the numerical precision of the computations. Setting the upper bound at 1.9999 did not cause numerical problems in any of the examples that we tested, but we also observed no significant difference in the kriging results when we set the upper bound at 1.99 versus 1.9999. In light of this, for all examples in this paper in which p is estimated, we set an upper bound of 1.99 on p.

4. Discussion: Characteristics of FBFs for Kriging

We first note that kriging with FBFs is still kriging, and, as such, it possesses the attractive characteristics inherent to kriging: It interpolates perfectly in the sense that $\hat{y}(\mathbf{x})$ passes through

every observation, which is usually a desirable feature for deterministic simulations; and it allows prediction uncertainty to be conveniently quantified via the MSE calculation that is inherent to kriging. In the remainder of this section, we discuss characteristics that are more specific to kriging with FBFs.

4.1 The nature of the basis functions and fitted surfaces

A fundamental consideration in any metamodeling approach is the nature of the basis functions associated with the fitted model, which dictates the nature of the fitted surfaces. For FBFs, from Equation (7), the basis function representation analogous to Equation (3) is

$$\hat{y}(\mathbf{x}) = y_0 + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \mathrm{y}_0 \cdot \mathbf{1}) = y_0 + \sum_{i=1}^{n-1} c_i \cdot g_i(\mathbf{x}), \qquad (10)$$

where the basis functions are of the form $g_i(\mathbf{x}) = [\mathbf{r}(\mathbf{x})]_i = Cov\big(Y(\mathbf{x}), Y(\mathbf{x}_i)\big)$. From the covariance expression in Equation (8), the smoothness of $\hat{y}(\mathbf{x})$ for FBFs is similar to the smoothness of $\hat{y}(\mathbf{x})$ for a stationary covariance model of the form $Cov\big(Y(\mathbf{x}), Y(\mathbf{x}_i)\big) = \exp\{-\|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}_i)\|^p\}$ or for the power exponential covariance model with the same p. The smoothness depends on choice of p: For $1 < p < 2$, the FBF basis functions have continuous first-order derivatives; for $p = 1$, the first derivatives are discontinuous at $\{\mathbf{x}_0, \mathbf{x}_i\}$, i.e., $\hat{y}(\mathbf{x})$ is discontinuous at every input site; and for $p = 2$, if this were allowed, the basis functions would be analytic. However, as discussed in Section 3, one should not use a value of p that is too close to 2. Consequently, for very smooth response surfaces for which the MLE of p for the power exponential model is 2, the result will be an FBF $\hat{y}(\mathbf{x})$ that is less smooth than the power exponential $\hat{y}(\mathbf{x})$. For $p < 1$, the directional derivatives of the basis functions are infinite at the input sites. Hence, $\hat{y}(\mathbf{x})$ will have very sharp peaks at the input sites (see Figure 4(c)), and for this reason we again do not recommend using $p < 1$.

We illustrate these points in the following. Figure 3(a) shows an example of two basis functions [$g_1(x)$ and $g_2(x)$ of Eq. (10)] for $d = 1$ and $p = 1$, i.e., for standard Brownian motion. The origin was taken to be $x_0 = 0$, and the two additional input sites corresponding to the two basis functions are $x_1 = 1.0$, and $x_2 = 3.1$. The ith basis function is piecewise linear with knots at

x_0 and the corresponding input site x_i, and it is constant outside of the two sites. Figure 3(b) shows analogous results for p = 1.9. The basis functions for p = 1.9 grow unbounded as $x \to \pm\infty$. Notice that for d = 1, the basis functions do not depend on the scale parameter, other than via a constant vertical scale factor ϕ^p that multiplies the basis function.

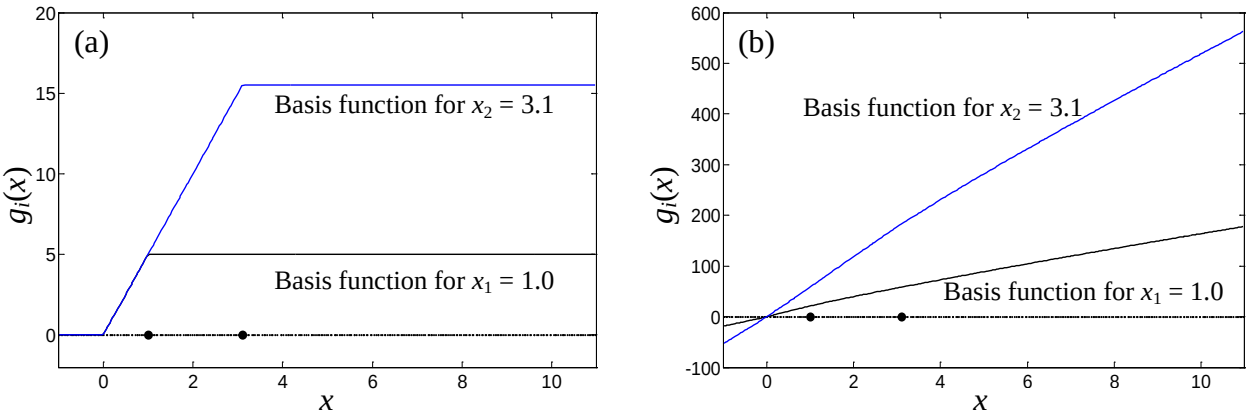


Figure 3. Example of basis functions $g_i(x)$ for d = 1 and (a) $p = 1$ and (b) $p = 1.9$

To illustrate the $\hat{y}(x)$ that can result from combining the basis functions via Equation (10) consider the response surface $y(x) = 0.00005x^5 - 0.0017x^4 + 0.005x^3 + 0.12x^2 + 0.7x$. The observed responses at seven input sites and the resulting $\hat{y}(x)$ for three different values of p are shown in Figure 4. Although the basis functions do not depend on the observed responses, their coefficients $\{c_i: i = 1, 2, \ldots, n-1\}$ do. It can be shown that for d = 1, $\hat{y}(x)$ is completely independent of ϕ. Figure 4(a) shows $\hat{y}(x)$ for p = 1, which in this case is piecewise linear with knots at every input site and constant beyond the range of input sites [as evident from Figure 3(a)]. Consequently, for p = 1 and d = 1, extrapolation of the response beyond the leftmost and rightmost input sites is flat with extrapolated value equal to the response at the closest input site. This characteristic of Brownian motion for extrapolation/interpolation is well known (Mandelbrot and Van Ness, 1968).

Although there is clearly no reversion to the mean in Figure 4(a), the piecewise linear nature of $\hat{y}(x)$ for p = 1 may be unappealing. For $p > 1$, $\hat{y}(x)$ is smoother, as is seen in Figure 4(b) for the case p = 1.9. Although the basis functions in Figure 3(b) appear nearly linear, when pieced together to form $\hat{y}(x)$ they are able to capture substantial curvature as seen in

Figure 4(b). For any $1 < p < 2$, interpolation between input sites is a curve with continuous first derivative everywhere. And trends are extrapolated beyond the range of input sites in the sense that $\hat{y}(x)$ is unbounded as $|x| \to \infty$. For $p < 1$, interpolation between input sites has undesirable peaked shape as shown in Figure 4(c). For $d = 1$, Mandelbrot and Van Ness (1968) discuss the characteristics of the FBF $\hat{y}(x)$ in more detail.

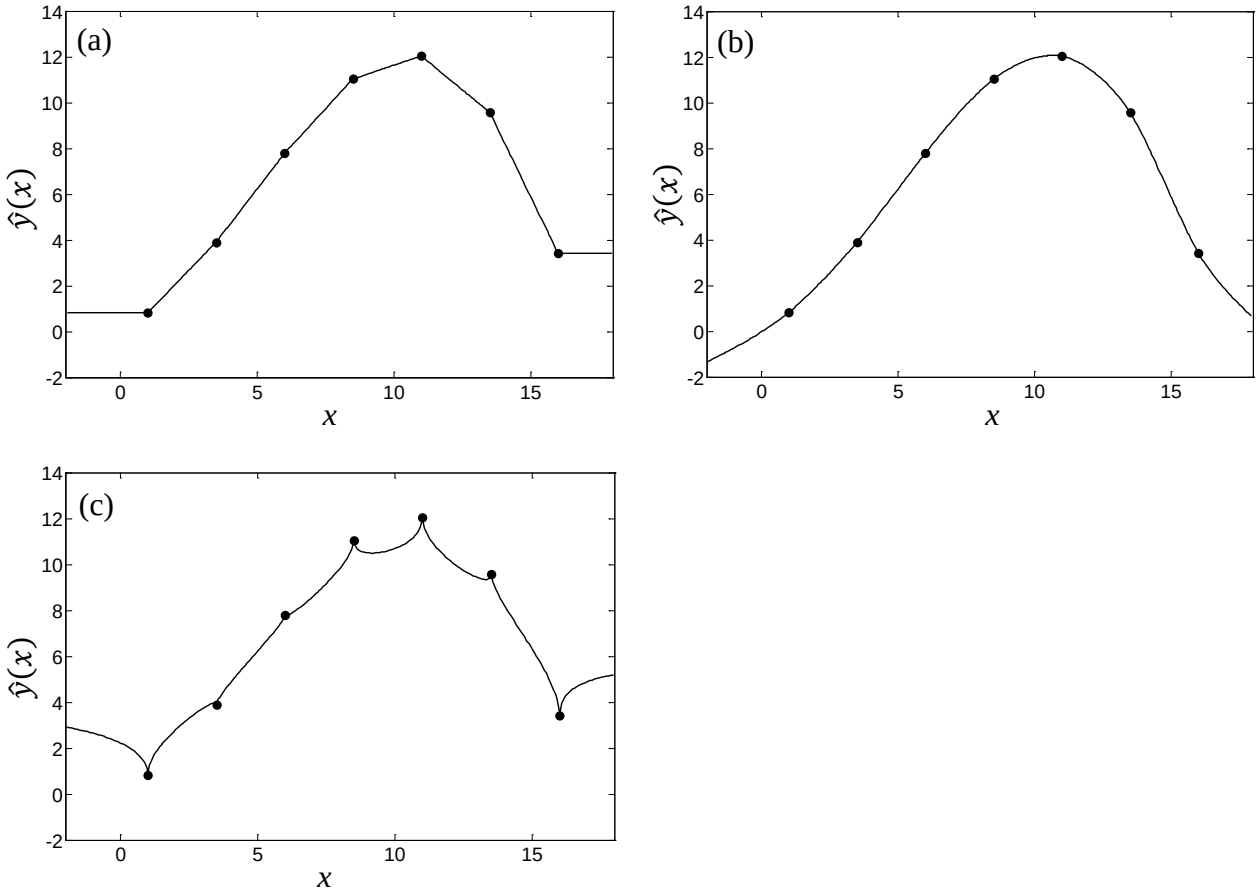


Figure 4. FBF $\hat{y}(x)$ for (a) $p = 1$, (b) $p = 1.9$ and (c) $p = 0.5$.

4.2 Avoiding reversion to the mean

As illustrated in Figure 1, the kriging predictor with a stationary covariance function will eventually revert to the mean as one extrapolates; and reversion to the mean may also occur when one interpolates between input sites, if there are gaps between the input sites that are large relative to the scale parameters. This is more likely to happen with sparse designs. From Equation (3) it is clear why $\hat{y}(\mathbf{x})$ reverts to the mean for kriging with stationary covariance

functions, for which the basis functions $g_i(\mathbf{x}) = Cov(Y(\mathbf{x}), Y(\mathbf{x}_i))$ decay to zero as $\mathbf{x}$ strays further from an input site $\mathbf{x}_i$. In contrast, the FBF basis functions in Equation (10) do not decay to zero. Having a covariance function that does not decay to zero as $\mathbf{x}$ diverges from $\mathbf{x}_i$ may seem like an undesirable property on first thought. But $Var(Y(\mathbf{x}))$ also depends on $\mathbf{x}$ for the nonstationary FBF model, and the dependence is such that the correlation $Corr(Y(\mathbf{x}), Y(\mathbf{x}_i)) = Cov(Y(\mathbf{x}), Y(\mathbf{x}_i))/\{Var(Y(\mathbf{x}))Var(Y(\mathbf{x}_i))\}^{1/2}$ between $Y(\mathbf{x})$ and $Y(\mathbf{x}_i)$ does indeed decay to zero as $\mathbf{x}$ diverges from $\mathbf{x}_i$.

Moreover, consider any two input locations $\mathbf{x}$ and $\mathbf{x}'$ having the same norm $||\mathbf{x}|| = ||\mathbf{x}'||$, but separated by an angle ω defined via $cos(\omega) = \mathbf{x}^T\mathbf{x}'/(||\mathbf{x}||\times||\mathbf{x}'||)$, as illustrated in Figure 5 for $d = 2$. For the standard FBF, it follows from (8) that $Corr(Y(\mathbf{x}), Y(\mathbf{x}')) = 1-[2-2cos(\omega)]^{p/2}/2$ depends only on ω, so that 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 $||\mathbf{x}||$. This may be a reasonable characteristic for many response surfaces. For example, suppose the origin in Figure 6 represents a geographic location, points $\mathbf{u}$ and $\mathbf{u}'$ are locations 100 meters from the origin, points $\mathbf{x}$ and $\mathbf{x}'$ are locations 1000 meters from the origin, and the response Y is the elevation relative to the elevation at the origin (i.e. $Y(\mathbf{0}) = 0$). We might expect the relative predictability (defined as the prediction error variance, divided by the prior variance) of the elevation at $\mathbf{x}$, given the elevation at $\mathbf{x}'$, to be similar to the relative predictability of the elevation at $\mathbf{u}$, given the elevation at $\mathbf{u}'$. This would imply that the correlation between $Y(\mathbf{u})$ and $Y(\mathbf{u}')$ is the same as the correlation between $Y(\mathbf{x})$ and $Y(\mathbf{x}')$.

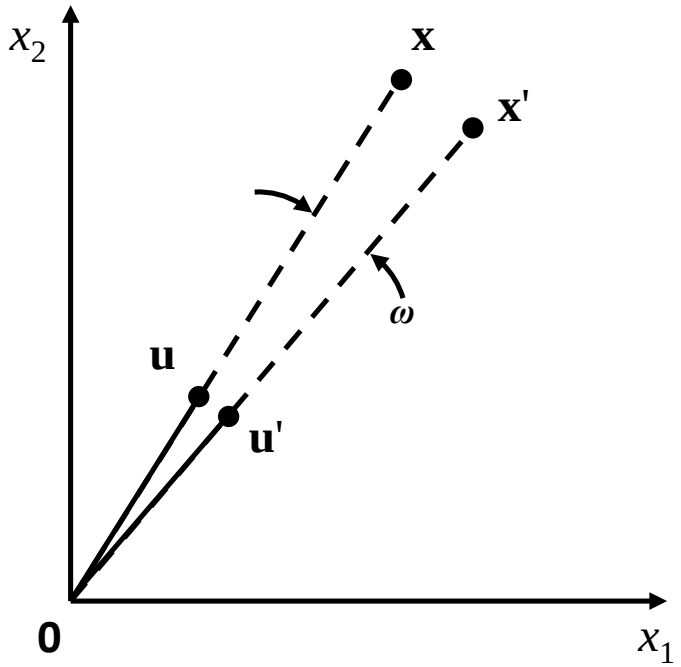


Figure 5. Illustration of how the correlation between the FBF response at **x** *and at* **x**′ *with* ||**x**|| = ||**x**′|| *depends only on the angle* ω *separating them.*

Figure 6(a) shows the FBF predictor $\hat{y}(x)$ for the same example of Figure 1 with a fixed p = 1.99 (the Gaussian correlation used in Figure 1 can be viewed as the power exponential with a fixed p = 2). The other parameter ϕ has no effect on $\hat{y}(x)$ in one dimension. Recall that in Figure 1, reversion to the mean is severe for the stationary covariance functions, whereas the FBF $\hat{y}(x)$ has no such problem. The root mean square error (RMSE) for 1,000 test sites evenly spaced over the interval [0,1] (i.e., for interpolation but no extrapolation) is 0.011 for the FBF model, versus 0.183 and 0.077 for the Gaussian and power exponential covariance models in Figure 1. Figure 6(b) also shows the FBF predictor $\hat{y}(x)$ with p = 1, which is the MLE of p when it is not held fixed at 1.99. There is clearly no reversion to the mean for the FBF precitor, but it might not be viewed as adequately smooth for this example, which we further discuss in Section 4.4.

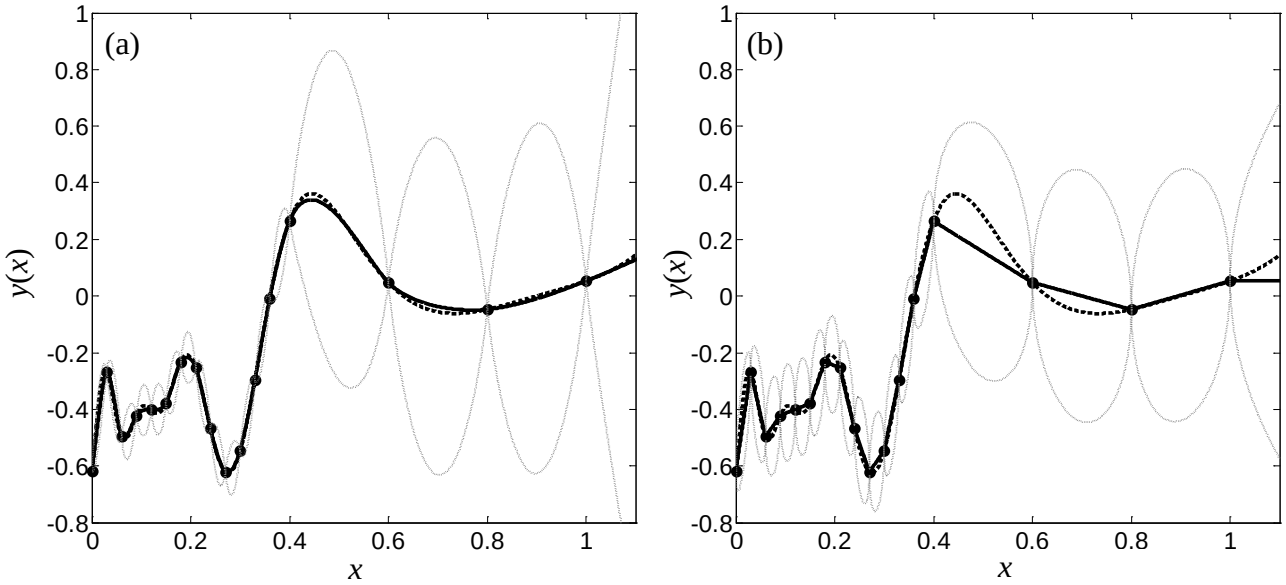


Figure 6. Plot of the same y(x) from Figure 1 (···), the FBF predictor $\hat{y}(x)$ (—),and 95% prediction intervals (gray dashed lines) with (a) p = 1.99 and (b) p = 1.

We next consider the same four-dimensional example considered in Qian, Seepersad, Joseph, Allen and Wu (2006), in which thermal conditions are simulated for a heat exchanger in an electronics cooling application. The simulated response is the heat transfer rate y_a, modeled as a function of the mass flow rate of entry air (x_1), temperature of entry air (x_2), temperature of the heat source (x_3), and solid material thermal conductivity (x_4). Details of the study and the range of the four input variables can be found in Qian et al. (2006). In their paper, a 64-run orthogonal array-based Latin hypercube design (LHD) was conducted to fit the kriging predictor, and 14 additional test runs were chosen randomly over a region slightly larger than the original design space to validate the predictor. Using the data from this, the MLEs for the Gaussian covariance model are $\hat{\theta}_1 = 0.220$, $\hat{\theta}_2 = 4.373$, $\hat{\theta}_3 = 0.143$, $\hat{\theta}_4 = 7.244$, and the RMSE is 5.148 (which agrees with the results in Qian, et al. 2006). For the power exponential covariance model, the results are very similar, because the MLE of p was 1.99. Qian et al. (2006) found the RMSE for the Gaussian covariance model with a linear prior mean function to be 2.588. For an FBF model fit to the same data the MLEs were $\hat{\phi}_1 = 21.31$, $\hat{\phi}_2 = 23.06$, $\hat{\phi}_3 = 15.86$, $\hat{\phi}_4 = 10.84$, and $\hat{p} = 1.93$. The test RMSE for the same 14-run test data was 1.727 for the FBF model, which is substantially better than for the Gaussian covariance or power exponential models. Notice that

for the Gaussian covariance model, the test RMSE was substantially better with a linear prior mean than with a constant prior mean, which indicates that there may have been substantial reversion to the mean. The FBF model avoids this.

As a higher-dimensional example with $d = 8$, consider the borehole example used in Morris, Mitchell and Ylvisaker (1993). The flow of water through a borehole that is drilled from the ground surface through two aquifers is modeled as

$$y = \frac{2\pi T_u (H_u - H_l)}{\ln\left(\frac{r}{r_w}\right)\left[1 + \frac{2LT_u}{\ln\left(\frac{r}{r_w}\right) r_w^2 K_w} + \frac{T_u}{T_l}\right]}$$

where y is the flow rate through the borehole, r_w (x_1) is the radius of borehole, r (x_2) is the radius of influence, T_u (x_3) is the transmissivity of the upper aquifer, H_u (x_4) is the potentiometric head of the upper aquifer, T_l (x_5) is the transmissivity of the lower aquifer, H_l (x_6) is the potentiometric head of the lower aquifer, L (x_7) is the length of the borehole, and K_w (x_8) is the hydraulic conductivity of the borehole. The details of the model and the ranges for the eight input variables can be found in Morris, et al. 1993.

Joseph, Hung and Sudjianto (2008) used a 27-run three-level orthogonal array design for the borehole example. Using this design (we also consider a LHD below), we fit three separate kriging models with Gaussian, power exponential, and FBF covariance. The MLEs were $\hat{\theta}_1 = 0.144$, $\hat{\theta}_2 = 9.59 \times 10^{-12}$, $\hat{\theta}_3 = 5.44 \times 10^{-8}$, $\hat{\theta}_4 = 0.00981$, $\hat{\theta}_5 = 6.37 \times 10^{-5}$, $\hat{\theta}_6 = 0.0141$, $\hat{\theta}_7 = 0.0161$, $\hat{\theta}_8 = 1.65 \times 10^{-3}$ for the Gaussian covariance (the power exponential covariance shared the same parameters because $\hat{p} = 2$), and $\hat{\phi}_1 = 188.542$, $\hat{\phi}_2 = 0.002$, $\hat{\phi}_3 = 0.002$, $\hat{\phi}_4 = 36.800$, $\hat{\phi}_5 = 0.005$, $\hat{\phi}_6 = 21.717$, $\hat{\phi}_7 = 27.549$, $\hat{\phi}_8 = 13.374$, and $\hat{p} = 1.87$ for the FBF model. To calculate the test RMSE, we generated 5,000 test points randomly over the rectangular input domain (no extrapolation). The test RMSE for the Gaussian and power exponential covariance models was 4.645 versus 3.303 for the FBF model. Because the sample standard deviation of the 5,000 test points was 45.18, these RMSEs correspond to r^2 values of

$1–4.645^2/45.18^2 = 0.9894$ and $1–3.303^2/45.18^2 = 0.9947$, respectively. We also considered a 27-run maximin distance LHD for the borehole example and refit the same three kriging models. After fitting to the data for 1,000 different LHDs, the average test RMSEs were 3.248 (r^2 = 0.9948) for the power exponential covariance model (which coincided with the Gaussian covariance model) and 5.341 (r^2 = 0.9860) for the FBF model.

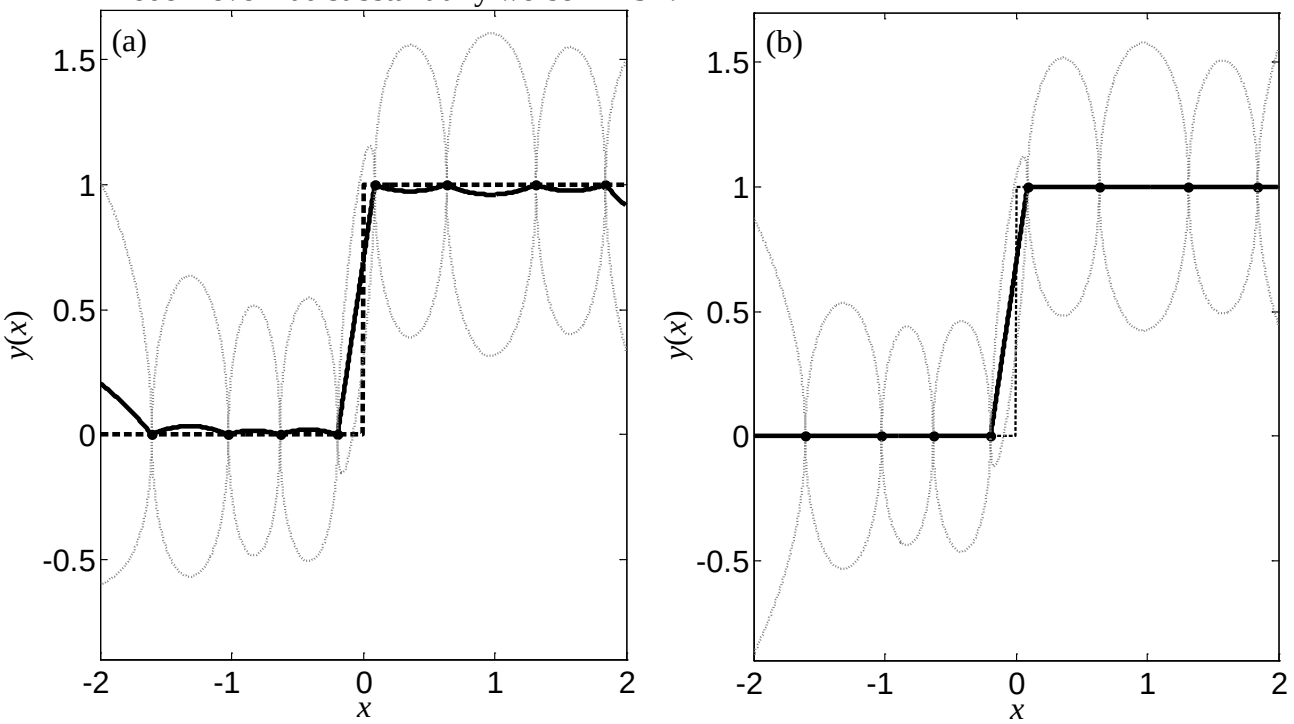
With such high r^2 values, both methods predicted the response surface quite well. For the orthogonal array design, the FBF model did a little better than the Gaussian (or power exponential) covariance model, and vice-versa for the LHDs. Part of the reason why the FBF predictor outperformed the Gaussian (or power exponential) covariance predictor for the orthogonal array design may be because the orthogonal array design was rather sparse in the interior of the design space. Such "holes" in the design space are where stationary covariance predictors are most subject to reversion to the mean, and there may have been mild reversion to the mean in these examples.

For the LHDs, the Gaussian covariance models predicted better than the FBF model. This is likely because the borehole response surface is quite smooth and predictable, and the LHD was a more appropriate space filling design than the orthogonal array. For other smooth response surfaces, we have observed results consistent with this: The Gaussian covariance model usually predicts better than the FBF model when a space filling design with a reasonably large number of input sites is used (barring severe numerical issues); whereas the FBF model may perform better if there are large holes in the design space. On the other hand, when the response surface has abrupt features, the FBF model can be much more robust and effective, as discussed in the following section.

4.3 Handling response surfaces with abrupt features

Figure 7 shows a one-dimensional example of a response surface with an abrupt feature, in this case an abrupt step at $x = 0$. Eight input sites were generated via a LHD. Figure 7(a) shows $\hat{y}(x)$ for the power exponential covariance model with MLEs of $\hat{\theta} = 1.56$ and $\hat{p} = 1$, for which

mean reversion is quite severe (mean reversion for the Gaussian covariance model was even more severe). Figure 7(b) shows analogous results for the FBF model with MLEs of $\hat{\phi} = 0.51$ and $\hat{p} = 1$, which clearly avoids reversion to the mean and provides better prediction. To further study this, we generated 1,000 different LHDs, each with 8 design points, and fitted separate FBF and power exponential models to each design. The test RMSEs (based on 5,000 test points) over the 1,000 replicates for the two models are plotted in Figure 8(a). As a point of reference, for the $\hat{y}(x)$ in Figures 7(a) and 7(b), the test RMSEs were 0.102 and 0.092, respectively. In spite of the better visual appearance of the fit in Figure 7(b) relative to the fit in Figure 8(a), the RMSEs are similar because most of the RMSE is attributed to the error in the vicinity of the step, which is similar for both methods. In Figure 8(a), the FBF model had smaller test RMSE for 70% of the replicates. Although the average test RMSE was only slightly better for the FBF model (0.123 vs. 0.126), it was substantially better for roughly 10% of the replicates. In contrast, the FBF model never had substantially worse RMSE.



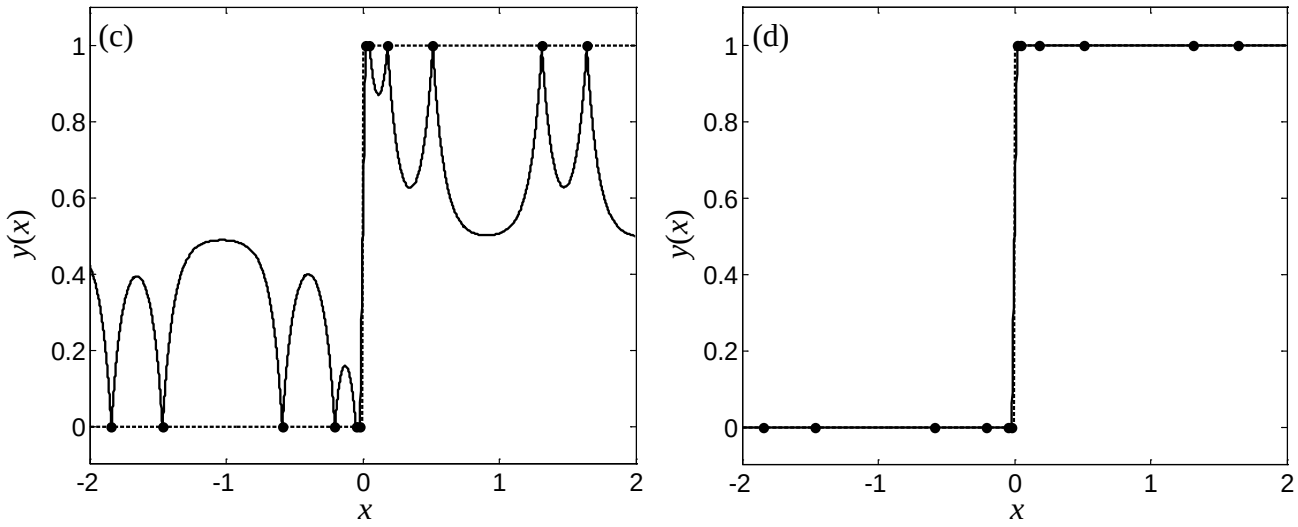


Figure 7. Plot of the step function response surface (---) and $\hat{y}(x)$ (—) with (a)(c) power exponential covariance function, (b)(d) FBF covariance function. (a) and (b) share the same LHD, and (c) and (d) share the same design, which is a LHD plus four extra points near the step.

The replicates in Figure 8(a) for which the power exponential model performed much worse than the FBF were those for which there happened to be two design points close to the step but on opposite sides of $x = 0$. In these situations, the MLE of θ for the power exponential model was quite large, resulting in more extreme reversion to the mean. To further investigate this phenomenon, we repeated the Monte Carlo simulation but with an augmented LHD. We generated an 8-point LHD as for the results shown Figure 8(a), but then added an additional 4 points at $x = -0.05, -0.02, 0.02$, and 0.05, to give a total of 12 points. For an example of the augmented design, Figures 7(c) and 7(d) show $\hat{y}(x)$ for the power exponential model ($\hat{p} = 1$, $\hat{\theta} = 148.3$) and FBF model ($\hat{p} = 1$, $\hat{\phi} = 2.273$), respectively. Figure 8(b) shows the test RMSEs for the two models over 1,000 replicates, each having a different 12-point augmented LHD as described above. The average test RMSEs were 0.393 for the power exponential, versus 0.029 for the FBF, which is more than an order of magnitude smaller.

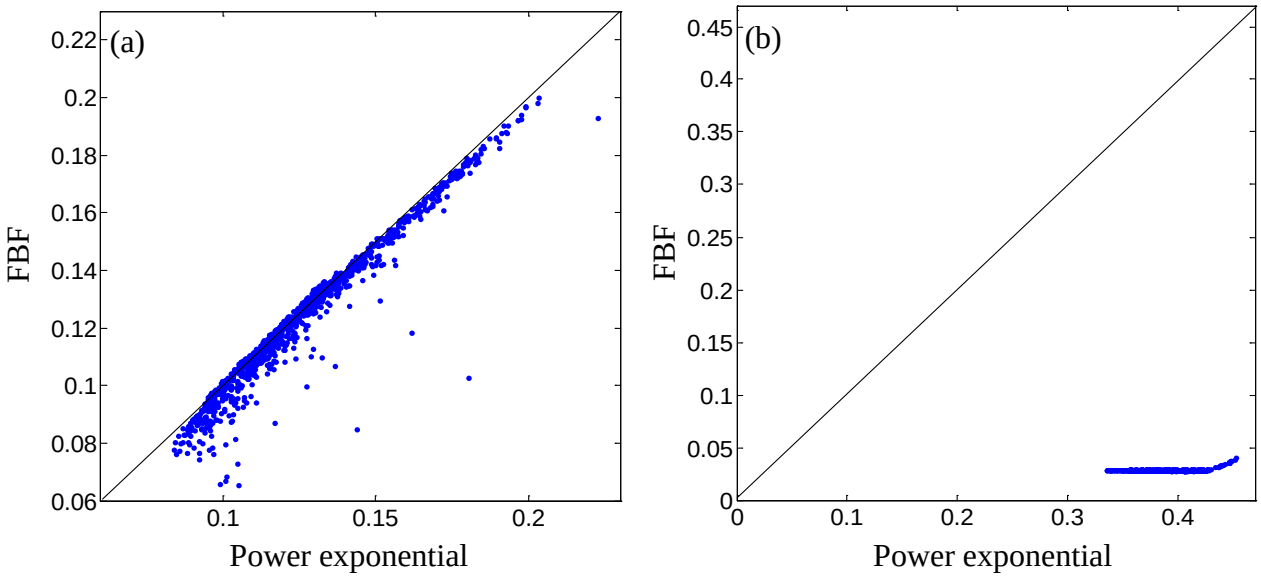


Figure 8. Plot of RMSE for power exponential vs. FBF covariance over 1,000 replicates for (a) LHD, (b) augmented LHD with four extra points near the step.

If an user observed response values similar to those shown in Figure 7(a,b) for an initial 8-point LHD, a natural next step would be to augment the design by adding points in the vicinity of what appears to be an abrupt change in the response surface, similar to what we did for Figure 7(c,d). But having denser input sites in the vicinity of the abrupt feature will inevitably result in larger (rougher) MLEs of the correlation/scale parameters. This greatly exacerbates the mean reversion for the power exponential model, as seen in Figure 7(c), but has little adverse effect on the predicted response for the FBF model.

As a smoother one-dimensional example, consider the function $y(x) = 1 - \exp(-x)\left(1 - \sqrt{x}\right)$, $x \in [0,10]$ shown in Figure 9. We used 9 design points, of which 4 are evenly spaced in the region [0, 0.15], and 5 are evenly, but more sparsely, spaced in the region [2, 10]. The MLEs were $\hat{\theta} = 5.425$ for the Gaussian covariance model, $\hat{\theta} = 10.428$ and $\hat{p} = 1$ for the power exponential model, and $\hat{\phi} = 0.178$ and $\hat{p} = 1$ for the FBF model. Figure 9(a) shows $\hat{y}(x)$ for the power exponential model, from which we see that mean reversion is quite severe (mean reversion for the Gaussian covariance model was even more severe). As in the preceding example, the reason is that the abrupt decrease in y near $x = 0$ resulted in a large (rough) estimate of θ, for which reversion to the mean is more severe. Figure 9(b) shows analogous results for the

FBF model, which clearly avoids reversion to the mean and provides better prediction. Both figures also show 95% PIs.

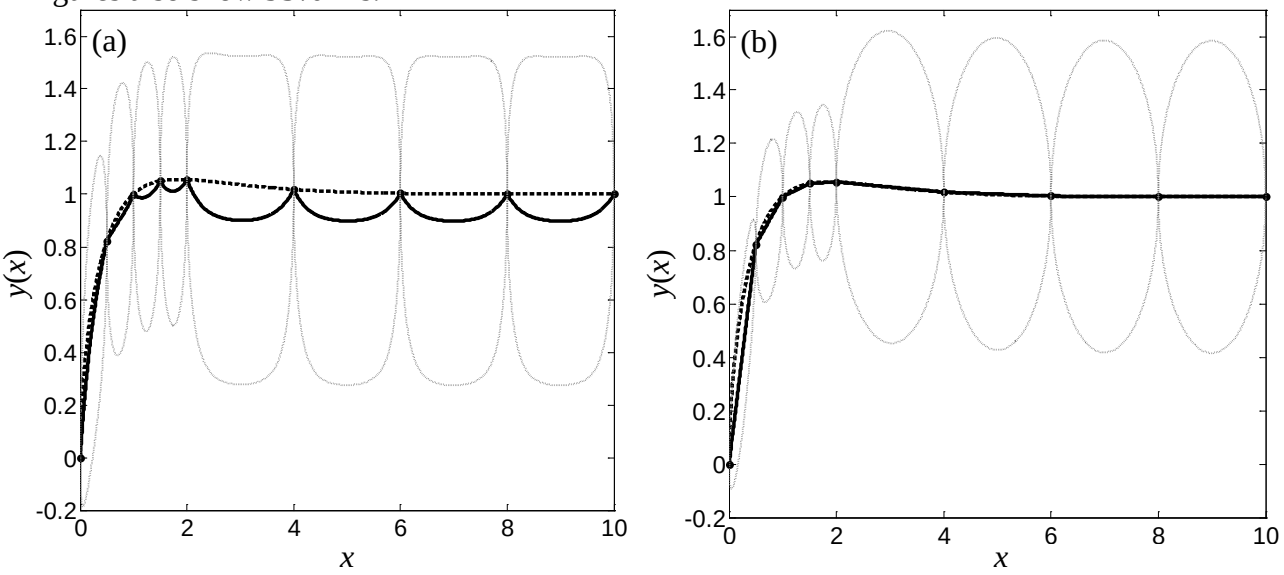


Figure 9. The function $y(x) = 1 - exp(-x)\left(1 - \sqrt{x}\right)$ (---), kriging predictor $\hat{y}(x)$ (—) and the 95% prediction interval (gray dashed lines) with (a) power exponential covariance model and (b) FBF covariance model.

Similar conclusions are drawn from the following two-dimensional example, which uses the response surface considered in Xiong et al (2007):

$$y(x_1, x_2) = \cos\big(6(x_1 - 0.5)\big) + 3.1|x_1 - 0.7| + 2(x_1 - 0.5) + 7\sin\left(\frac{1}{|x_1 - 0.5| + 0.31}\right) + 0.5x_2$$

for $x_1, x_2 \in [0,1]$. The plot of this function in Figure 10(a) demonstrates an abrupt feature around the x_1 = 0.5 line, at which the function is minimized. Figure 10(b) shows 24 design points generated via a Latin hyper cube design. Based on this design, $\hat{y}(x)$ for the power exponential model (which was the same as the Gaussian covariance model because $\hat{p} = 2$) and FBF model are shown in Figures 10(b) and 10(c), respectively. The MLEs were $\hat{\theta}_1 = 0.14$, $\hat{\theta}_2 = 71.89$, and $\hat{p} = 2$ in Figure 10(b) and $\hat{\phi}_1 = 0.58$, $\hat{\phi}_2 = 34.12$, and $\hat{p} = 1.91$ in Figure 10(c). Some level of

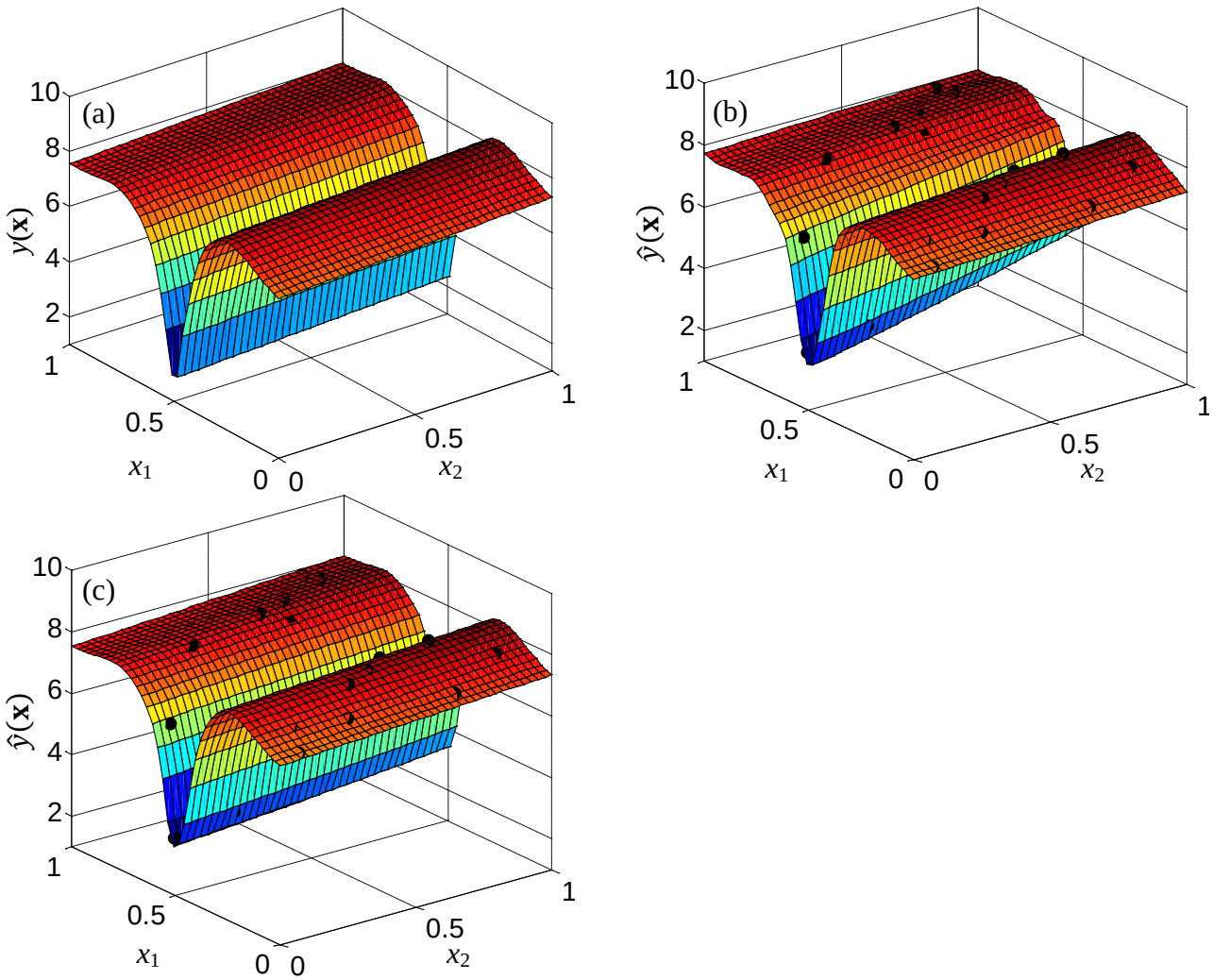


Figure 10. Example surface from Xiong et al (2007): (a) true response surface; (b) $\hat{y}(\mathbf{x})$ for the power-exponential model; and (c) $\hat{y}(\mathbf{x})$ for the FBF model.

mean reversion for the power exponential model can be observed in Figure 10(b), for example along the $x_1 = 0.5$ line and also the wavy behavior along the $x_1 = 0.3$ line. In contrast, the FBF predictor in Figure 10(c) does not exhibit this undesirable trait. We also generated 1,000 different LHDs, similar to what was described for Figure 8. Figure 11(a) compares the test RMSEs (based on 12,100 test points) for the two models over 1,000 24–point LHDs, and Figure 11(b) shows analogous results for an augmented design with an additional 8 points around the abrupt feature generated via a LHD over the region $0.45 \le x_1 \le 0.55$, $0 \le x_2 \le 1$. The conclusions are similar to those from Figure 8: For the 24-point design (Figure 11a), the FBF model provides a smaller RMSE in 76.8% of the replicates, and the average test RMSEs are 0.235 for the power exponential versus 0.181 for the FBF model. The differences are more pronounced for the

augmented design (Figure 11b) with an additional 8 points added near the abrupt feature, for which the average test RMSEs are 0.139 for the power exponential versus 0.052 for the FBF model. And the FBF model never did substantially worse than the power exponential model.

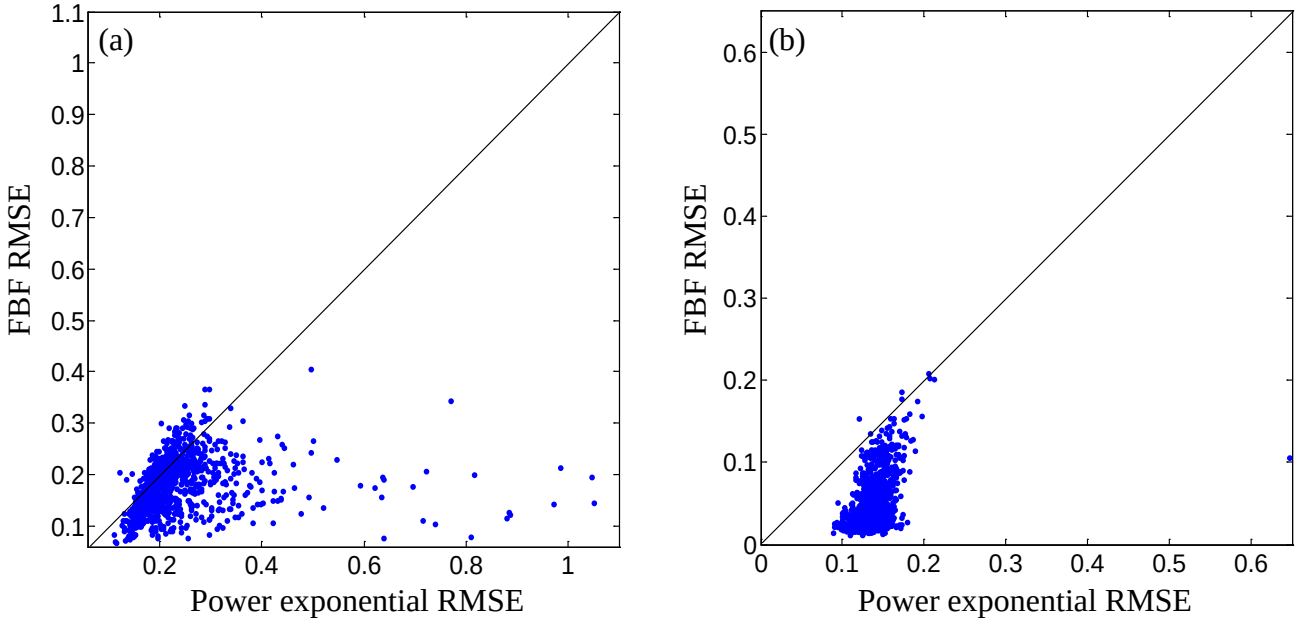


Figure 11. Plot of RMSE for kriging with power exponential covariance function vs. kriging with FBF for the response surface in Figure 10(a). (a) Latin Hyper Cube design, (b) augmented Latin Hyper Cube design with extra points near the abrupt feature.

4.4 Numerical stability and local roughness of FBFs

An additional advantage of FBF models for kriging is numerical stability. Numerical problems resulting from a nearly singular covariance matrix (**R** must be inverted) plague kriging with many stationary covariance models when n is large or when design points are added sequentially in a manner that places them more densely in some regions of the input space (e.g., in the context of optimizing the response). The smaller the correlation parameters $\{\theta_i: i = 1, 2, \ldots, d\}$, the worse the singularity. Hence, singularity problems can be especially problematic when searching over a range of values of $\{\theta_i: i = 1, 2, \ldots, d\}$ within an MLE algorithm, if small values are tried.

For FBF models, the scale parameters $\mathbf{\Phi}$ have much less effect on the condition number of **R** than do the scale (correlation) parameters for the Gaussian covariance model. For example, rescaling $\mathbf{\Phi} \to \alpha\mathbf{\Phi}$ for some scalar α has no effect at all on the condition number of **R** for the

FBF model. We have encountered almost no numerical problems when fitting FBF models, but we are rarely able to fit a large data set using a Gaussian covariance model without taking some precautions to avoid singularity issues. An explanation for the better numerical stability is that an FBF stochastic process $Y(\mathbf{x})$ has continuous sample paths but is nowhere differentiable. The nondifferentiability translates to local behavior that is somewhat rough, which results in closely-spaced design points not being as highly correlated as with a Gaussian covariance model.

This relates to a potential drawback of FBFs. Their local behavior may be rougher than what one would ideally expect for some response surfaces. Although $\hat{y}(\mathbf{x})$ has continuous first derivatives everywhere for $1 < p < 2$, its second derivative blows up at the input sites. This could be problematic if one uses second derivatives in an optimization routine to minimize or maximize $\hat{y}(\mathbf{x})$. And as discussed in Section 4.2, the Gaussian covariance model usually predicts better than the FBF model for a very smooth response surface when a space filling design with a reasonably large number of input sites is used.

It should be noted that the power exponential $\hat{y}(\mathbf{x})$ has the same smoothness characteristics (continuous first derivatives but second derivatives that blow up at the input sites) as the FBF $\hat{y}(\mathbf{x})$ for the same value of $1 < p < 2$. In light of this, one may wonder why the power exponential model is better able to accommodate smooth response surfaces. Perhaps the reason is that the limiting form of the power exponential model as $p \to 2$ is the valid Gaussian covariance model, which results in an analytic $\hat{y}(\mathbf{x})$. In contrast, the FBF model has no valid limiting form as $p \to 2$, at least not for $d > 1$.

5. Conclusions

We have argued that an FBF model may be an attractive choice for kriging covariance model in many response surface metamodeling applications, even though it has seldom been considered for these purposes. From a practical implementation perspective, we have discussed a number of characteristics of kriging with FBFs, the attractive ones being: 1) They inherently avoid reversion to the mean in an appealing way, without introducing numerical problems and

excessively narrow PIs (as is the case if one enforces overly small correlation parameters to mitigate reversion to the mean with Gaussian or power exponential covariance models). 2) Numerical problems due to poorly conditioned covariance matrices are less severe than with Gaussian and power exponential covariance functions. 3) They provide quite robust and effective prediction when the true response surface exhibits abrupt features (for Gaussian and power exponential covariance functions, abrupt features result in rough correlation parameter estimates, which exacerbates reversion to the mean).

A downside of using FBFs for response surface metamodeling is that for smooth response surfaces, the Gaussian covariance model usually predicts better than the FBF model when a space filling design with a reasonably large number of input sites is used (and barring severe numerical problems). However, when one uses a design for which there are holes in the design space (over which $\hat{y}(\mathbf{x})$ is more subject to reversion to the mean), the FBF model may result in better prediction even for smooth surfaces. This was the case when the orthogonal array designs were used for the borehole examples. Holes in the design space, or more generally designs that have points that are more densely spaced in some regions and sparsely spaced in others, can also result from sequential design procedures, especially in the context of optimization of the response surface.

Although the power exponential model has similar smoothness characteristics as the FBF model for the same value of p, it has a valid limiting form as $p \to 2$, and this limiting form is the Gaussian covariance model with analytic $\hat{y}(\mathbf{x})$. In contrast, the FBF model has no valid limiting form as $p \to 2$. We have also developed smoothed versions of FBF models that perform better for smooth response surfaces, but that retain most of the desirable characteristics of FBFs (Zhang and Apley, 2014).

Acknowledgement

This work was supported by the National Science Foundation under grant CMMI-1233403. The authors also gratefully acknowledge a number of helpful comments from the anonymous referees.

REFERENCES

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

Chipman, H., Ranjan, P. and Wang, W. (2012). "Sequential Design for Computer Experiments with a Flexible Bayesian Additive Model," *Canadian Journal of Statistics*, **40**, 663-678.

Christensen, R. (1990). "The equivalence of Predictions from Universal Kriging and Intrinsic Random Function-Kriging," *Mathematical Geology*, **22**(6), 655-664.

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.

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.

Harville, D. (1974). "Bayesian Inference for Variance Components Using Only Error Contrasts," *Biometrika*, **61**(2), 383-385.

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.

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.

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.

Mandelbrot, B. B. and Van Ness, J. W. (1968). "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, **10**(4), 422-437.

Matheron, G. (1963). "Principles of Geostatistics," *Economic Geology*, 58, 1246-1266.

Matheron, G. (1973). "The Intrinsic Random Functions and Their Applications," *Advances in Applied Probability*, Vol. 5, No. 3, pp. 439-468.

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.

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.

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.

Wackernagel, H. (2003). Multivariate Geostatistics, 3rd Ed., Springer, New York.

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). "Brownian Integrated Covariance Functions for Gaussian Process Modeling of Computer Experiments." Submitted for publication. Also available as part of the unpublished Ph.D. Dissertation "Fractional Brownian Fields for Engineering Response Surface Metamodeling" of N. Zhang, Northwestern University, 2013.