Lifted Brownian Kriging Models

Matthew Plumlee and Daniel W. Apley

ABSTRACT
Gaussian processes have become a standard framework for modeling deterministic computer simulations and producing predictions of the response surface. This article investigates a new covariance function that is shown to offer superior prediction compared to the more common covariances for computer simulations of real physical systems. This is demonstrated via a gamut of realistic examples. A simple, closed-form expression for the covariance is derived as a limiting form of a Brownian-like covariance model as it is extended to some hypothetical higher-dimensional input domain, and so we term it a lifted Brownian covariance. This covariance has connections with the multiquadric kernel. Through analysis of the kriging model, this article offers some theoretical comparisons between the proposed covariance model and existing covariance models. The major emphasis of the theory is explaining why the proposed covariance is superior to its traditional counterparts for many computer simulations of real physical systems. Supplementary materials for this article are available online.

1. Introduction

Originally created for geostatistics (Matheron 1963), kriging models are used to construct predictions of unevaluated points on a response surface based on a finite number of evaluations. Kriging has become the de facto method for prediction of a computer simulation’s response after an experiment (Sacks et al. 1989; Santner, Williams, and Notz 2003). The kriging model is defined by the covariance function, which controls the covariance of the evaluations. The kriging predictor is the best linear unbiased predictor under the kriging model (Stein 1999). With the increased usage of kriging for computer simulations, problems have emerged when directly using the tools designed for geostatistics. In Section 6 of this article, a broad array of realistic prediction examples are analyzed to compare the standard covariance (i.e., covariances like Gaussian, power exponential, or Matérn) to a significantly different covariance function that we introduce in this work, which, for reasons that will become apparent, we term the lifted Brownian covariance. The results show consistently superior performance of the lifted Brownian covariance function and strongly support its use on computer simulations over the traditional covariance model.

An explanation for this result is computer simulations of physical systems have properties that can be represented with the lifted Brownian covariance but not with the standard covariance. The lifted Brownian covariance has close connections with the popular multiquadric kernel used for surface smoothing (Hardy 1971). In this article, we take a kriging perspective and are able to derive formal results on the contrast between this covariance and existing covariances. More specifically, this article demonstrates the standard covariance model implies a response surface of peaks and valleys; much like a geological surface. In contrast, computer simulations often do not have negatively correlated increments that imply peak and valley behavior (Higdon et al. 2004; Loepky, Sacks, and Welch 2009), and the lifted Brownian covariance is more consistent with this type of behavior. The peak and valley behavior is also known as mean reverting behavior. One way to observe an absence of negatively correlated increments pictorially is to plot the change in the response while changing only a single coordinate and holding all other coordinates constant. This is shown in Figure 1 for some physical models of real systems often used in the literature (Morris, Mitchell, and Ylvisaker 1993; Kenett and Zacks 1998; Ben-Ari and Steinberg 2007; Moon, Dean, and Santner 2012). Section 6.2 discusses the details of these functions. The conclusion from these plots is that the responses are not composed of peaks and valleys.

The goal of this article is to build a more effective kriging covariance function for analysis of computer simulations after experiments. This work derives and studies the properties of a novel covariance function that is appealing in terms of ease of use, computational tractability, and representation of both rough and smooth surfaces. This work builds on Zhang and Apley (2014, 2015), who first discussed the idea of using formulations of Brownian fields (Mandelbrot and Van Ness 1968) as kriging models for computer response modeling to avoid the negatively correlated increments associated with more traditional covariance functions. Section 2 derives the proposed covariance as a particular limiting form of the Brownian integrated covariance of Zhang and Apley (2015). We also discuss how this limiting form inherits the desirable properties of the
Brownian integrated covariance that make it a good choice for computer response surfaces that do not have negatively correlated increments, while also having a simple closed-form expression. The Brownian integrated covariance of Zhang and Apley (2015) requires numerical integration and is more difficult to implement.

Section 3 is a theoretical investigation of the properties of the proposed covariance function, specifically contrasting its properties with standard covariance functions used for kriging. Section 4 discusses the implementation details of prediction with the proposed covariance function. Sections 5 and 6 report on numerical comparisons of the resulting kriging prediction using the standard covariance and the lifted Brownian covariance. Twelve examples are used that are all models of real engineering and scientific systems. We focus exclusively on examples that are models of real physical systems, as opposed to purely mathematical test functions, because the latter often do not exhibit behavior that is representative of real physical systems. The predictions when using the proposed covariance model were, on average, better than when using the standard covariance model for every example. We support and explain this finding via the theoretical arguments in Section 3.

2. Lifted Brownian Covariance

Let \( y(\cdot) \) represent the computer model with inputs labeled \( x \) in \( \mathbb{R}^d \). This article will operate under the assumption that the kriging model is a Gaussian process. The Gaussian process is a valid probability distribution on continuous functions with easily computable conditional distributions. Under this model, the joint distribution of \( y(\cdot) \)'s evaluations is multivariate normal. For notational reasons that will become apparent later, the presentation in this section assumes the response surface is centered such that \( y(0) = 0 \).

Zhang and Apley (2014) proposed directly using the models of fractional Brownian fields as kriging models, but these models have very limited short range behavior in the responses, which does not represent many computer simulations. Zhang and Apley (2015) incorporated an integrator (of the fractional Brownian type) into a standard covariance model, which resulted in a covariance that performs better for smooth surfaces. However, their Brownian-integrated covariance function, which must be evaluated many times during kriging prediction, is expressed only as an integral. To approximate it numerically can be computationally expensive in practice, relative to existing covariances. Moreover, approximating these integrals with numerical methods can propagate numerical errors in the resulting kriging predictor. Thus, numerical integration may exacerbate the known numerical issues that can occur when using kriging models (Haaland and Qian 2011; Peng and Wu 2014). These two practical issues led us to an approach similar to Zhang and Apley (2015), but with the practical feasibility issue resolved. The core innovation of our article's approach is taking a certain limit in the Brownian-integrated covariance model.

Tabling the derivation, we now state the proposed covariance function's form, which we term the lifted Brownian covariance function. The reason for this name will be illustrated by Theorem 1. The response \( y(\cdot) \) is considered a realization of the stochastic process \( Z(\cdot) \) over the input space with \( \text{cov}(Z(x), Z(x')) = c(x, x') \). As for any kriging covariance model, the covariance between these observations defines the best linear unbiased predictor and the standard error of that predictor (Sacks et al. 1989). A scaling constant (which corresponds to a prior variance in standard covariance models) is typically used in practice such that the stochastic process is \( aZ(\cdot) \) with \( a > 0 \). This function \( c(\cdot, \cdot) \) must be positive definite so that the covariance matrix of a collection of evaluations is always positive definite.

The proposed covariance function is given by

\[
c(x, x') = \psi(x) + \psi(x') - \psi(x - x') - 1,
\]

where \( \psi(h) = \left(1 + ||h||_a^2\right)^\beta \).

(1)

where \( a = (a_1, \ldots, a_d) \) is a vector of strictly positive lengthscale parameters,

\[
||h||_a^2 = h^T \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_d \end{pmatrix} h,
\]

and \( \beta \) is a parameter that controls the long range behavior of the model. In the sense described in detail in Section 3, this parameter \( \beta \) is the key component missing in standard kriging models that use standard covariance structures. A larger value of \( \beta \) corresponds to a surface with less negatively correlated increments, as illustrated in examples in Figure 2. Importantly, the

---

**Figure 1.** Plot of response while only changing a single coordinate for several functions described in Section 6.2.
value of \( \beta \) is restricted to be strictly positive and less than unity to ensure the Gaussian process has a Karhunen–Loève expansion with an infinite number of nonzero terms (see, e.g., Hofmann, Schölkopf, and Smola 2008). When \( \beta \in (0, 1) \), the response surface has an infinite degrees of freedom. We also consider the more general formulation

\[
c(x, x') = \psi(x) + \psi(x') - \psi(x - x') - 1 \quad \text{where} \quad \psi(h) = \left(1 + (\|h\|^2)^{\beta/2}\right)^{\beta/2}.
\]

(2)

As will be described in detail in Section 3, the parameter \( \gamma \in (0, 1] \) controls the short range behavior (related to local roughness/smoothness) of the response, and in (2), both \( \beta \) and \( \gamma \) control the long range behavior together. Discussions at the end of Section 4 give a practical guide when estimating or setting these parameters for prediction.

Historically, the knowledge that (2) is positive definite, a requirement for kriging models, dates back to the work of Micchelli (1986), who demonstrated that \( (1 + |x|^2)^{\beta} \) is a conditionally negative definite function, where \( \|\cdot\| \) is the standard Euclidian norm. Micchelli (1986) had important implications for the multiquadric kernel of Hardy (1971), which was widely used in practice prior to the proof of its positive definiteness. As shown in the Appendix of this article, the results from Micchelli (1986) demonstrate the positive definiteness of the proposed covariance function using a common derivation. As another example, Gneiting (2002) implicitly demonstrated that \( c(\cdot, \cdot) \) is positive definite. Despite these derivations, we are unaware of any mentions of it in the computer simulation literature, likely because the underlying properties have not been thoroughly explored until this work.

We now show the connection between the covariance in (1) and the Brownian-integrated covariance of Zhang and Apley (2015). Zhang and Apley resolved the short range modeling problem for fractional Brownian covariances by revisiting the definition of fractional Brownian motion as integrals over white noise. If \( W(\cdot) \) is a white-noise process, then a fractional Brownian field can be viewed as a white-noise integral of the form (Lindstrøm 1993):

\[
Z(x) = \int_{\mathbb{R}^d} \left(\|x - h\|^\beta - 2 \|h\|^\beta dh \right) W(h) dh.
\]

Zhang and Apley (2015) chose to replace the rough white-noise process with any stationary Gaussian process, \( S(\cdot) \), and defined a Brownian-integrated process as

\[
Z(x) = \frac{1}{2} \int_{\mathbb{R}^d} \left(\|x - h\|^\beta - 2 \|h\|^\beta dh \right) S(h) dh.
\]

Denote \( \text{cov}(S(x), S(x')) \) by \( r(x - x') \), which depends only on the distance \( |x - x'| \). Then \( Z(\cdot) \) has covariance function (up to a multiplicative constant)

\[
\text{cov}(Z(x), Z(x')) = \zeta(x) + \zeta(x') - \zeta(x - x') - \zeta(0),
\]

where \( \zeta(x) = \int_{\mathbb{R}^d} \left(\|x - h\|^\beta - 2 \|h\|^\beta dh \right) r(h) dh \).

(3)

We will show that as the dimension of the input grows in a certain manner, the Brownian integrated covariance approaches (1) if \( r(\cdot) \) is the Gaussian covariance. To mathematically lift a function is to extend it to some higher-dimensional domain, in a manner that preserves its evaluations on its original lower-dimensional domain. In the following result, we consider a lifted version of the Brownian integrated process \( Z(\cdot) \), which we evaluate at \( k \)-dimensional inputs (for each \( k > d \)) of the form \( x_k = (x, 0, 0, \ldots, 0)^T \in \mathbb{R}^k \) for \( x \in \mathbb{R}^d \). Note that this preserves the Euclidean norm of \( x \) and its padded \( k \)-dimensional version. The intuitive justification for letting the input dimension artificially approach infinity with the extra input components all identically zero is as follows: imagine that a function has a very large number of possible inputs, but in our experiment we only altered \( d \) of them, with the rest held constant at the origin. Given that many experiments leave many potential inputs unchanged for various reasons, this logic may align with many implementation scenarios. The real benefit of lifting the Brownian integrated process in this manner is that it results in a closed-form expression for the covariance, as stated in the following theorem, the proof of which is in the online supplementary materials.

**Theorem 1.** Consider any \( x \in \mathbb{R}^d \). For \( k > d \), let

- \( a_k \) be the vector of the first \( k \) elements of a positive sequence \( a_1, a_2, \ldots, a_d, a_{d+1}, a_{d+2}, \ldots \),

that is, the lengthscales are as before for the first \( d \) elements and can be anything afterward,

- \( x_k = (x, 0, 0, \ldots, 0)^T \), that is, \( x_k \) is the vector in \( \mathbb{R}^k \) with the first \( d \) entries \( x \) and the last \( k - d \) entries 0, and
\[ r_k(h) = \frac{k^k}{(2\pi)^k} \prod_{i=1}^{k} a_i \exp \left( -\frac{k}{2} \|h\|_{a_i}^2 \right), \]

Then as \( k \to \infty \)

\[ \zeta_k(x_k) \to (1 + \|x\|_a^2)^\beta - 1. \]

This result shows that the covariance given in (1) is the covariance of a Brownian integrated process lifted up to infinite-dimensional space but then evaluated only on the \( d \)-dimensional subspace that corresponds to its first \( d \) input coordinates, with the remaining input coordinates fixed at zero.

There is the remaining extension of this covariance that offers even more flexibility, which is given in (2). This has connections with the former version of lifted Brownian covariance, as illustrated in the next section.

The key importance of Theorem 1 is that we have arrived at a covariance that is of the Brownian integrated type proposed by Zhang and Apley (2015). But our proposed covariance can be as fast to evaluate as the Gaussian covariance in most application packages. Using the author's desktop computer, experiments with the computing package Matlab found that 10⁸ random evaluations of the Gaussian covariance function on the unit segment took 0.50 sec, and if \( \beta = 1/2 \) and \( \gamma = 1 \) the proposed covariance look 0.35 sec. If \( \beta \) is some other fraction, it took considerably longer, around 10 sec. Evaluating the power exponential covariance, for example, \( \exp(-|x - x'|^{1.9}) \), had similar computational time as when \( \beta \neq 1/2 \), taking around 5 sec. It should be noted that this number of evaluations is uncommon in experiments on computer simulations, it requires \( \sqrt{2} \cdot 10^8 \approx 14,000 \) observations to have a covariance matrix with \( 10^9 \) unique entries.

### 3. Properties of Lifted Brownian Covariances Versus Stationary Covariances

This section outlines the properties of the proposed covariance, with an emphasis on computer simulation models. There are two points of discussion. We first show that unlike fractional Brownian field models of covariances, the proposed covariances have flexible short range behavior. Short range behavior can be thought of as the roughness of a surface. This implies the effect of small perturbations of the inputs on a computer model response can be flexibly modeled. Second, it is shown that the proposed covariances also have flexible long range behavior, which accounts for general trends in the response over large perturbations of inputs. Put in the simplest terms, all stationary covariances operate under the assumption that “what goes up is expected come down” in the long run. This creates the peak and valley (mean reversion) behavior discussed in the introduction. As will be shown, this is also related to the negative correlation between increments. While this may be logical for many geostatistical applications, the intuition for many computer simulations is the counter to this: if there is an increase along a direction, one might expect it to more likely continue to increase along that direction. This will, of course, not be true for all simulations, but it appears true for computer simulations of many physical systems. Unfortunately, the stationary covariance is the only one discussed in the majorities of articles designed for experiments on computer simulations (Santner, Williams, and Notz 2003). These properties are detailed to help explain the gains in predictive performance noted in Section 6.

Like the last section, the presentation in this section assumes the response surface is centered such that \( y(0) = 0 \).

#### 3.1 Short-Range Dependency

Consider now the popular measure of short-range dependency: the fractal dimension (Gneiting and Schlather 2004). In the above context, let \( x \) and \( h \) be in \( \mathbb{R}^d \), and let \( \kappa \) be some scalar. If

\[
\frac{|\mathbb{E}[Z(x)[Z(x) - Z(x + \kappa h)]]|}{|\kappa|^b} = \frac{|c(x, x) - c(x, x + \kappa h)|}{|\kappa|^b} = \text{constant},
\]

it is said to have fractal dimension \( D = \lim_{\kappa \to 0} d + 1 - b/2 \). The above is often written as

\[
|c(x, x) - c(x, x + \kappa h)| \sim |\kappa|^{2\beta} \text{ as } \kappa \to 0.
\]

The problem with the classical fractional Brownian covariance (Zhang and Apley 2014) is that it has limited flexibility in the short-range behavior of the response surfaces. For fractional Brownian fields with parameter \( \beta \), the fractal dimension is governed by the relationship

\[
|c(x, x) - c(x, x + \kappa h)| \sim |\kappa|^2 \text{ as } \kappa \to 0.
\]

But \( \beta \) also controls the long-range behavior (Gneiting and Schlather 2004). Thus, it is impossible to create a surface with both a certain short range and long range behavior since \( \beta \) controls both. Also, there is no fractional Brownian field such that

\[
|c(x, x) - c(x, x + \kappa h)| \sim |\kappa|^{2\gamma} \text{ as } \kappa \to 0,
\]

which is needed for very smooth response surfaces. When \( c \) is the most common covariance function, the Gaussian, \( c(x, x) - c(x, x + \kappa h) \) behaves as \( |\kappa|^2 \) for small \( \kappa \).

A random field with the covariance given in (2) has short range behavior of

\[
|c(x, x) - c(x, x + \kappa h)| \sim |\kappa|^{2\gamma} \text{ as } \kappa \to 0.
\]

Thus, we have separate control of the long range dependency, which will be shown to be controlled by the product \( \beta \gamma \), and the short range dependency of the surface, controlled by \( \gamma \). This is illustrated in Figure 3. When \( \gamma = 1 \),

\[
|c(x, x) - c(x, x + \kappa h)| \sim |\kappa|^2 \text{ as } \kappa \to 0,
\]

and the response surface is analytic. This may be undesirable for many applications if there is the potential for a discontinuity in the derivative at any point. On the other hand, \( \gamma = 1 \) matches the behavior of the popular Gaussian covariance function. Thus
Figure 3. Plot of sample paths under the kriging model with covariance (2) with various values of $\gamma$, the short range behavior parameter, with a constant value for $\beta \gamma$, which controls the long range behavior. These are conditional on an observation at the origin being zero.

$\gamma = 1$ may prove widely applicable. In the case studies presented in this article, we fixed $\gamma$ to 1.

3.2 Long-Range Dependency

We now show that all stationary covariances, which are dominant for analysis of experiments on computer simulations, imply that the surface exhibits negatively correlated increments, at least in the long range. To illustrate this, consider the difference along increments of our response as the input is changed,

$$S_1 = Z(h_1) - Z(0)$$

and

$$S_2 = Z(h_2 + h_1) - Z(h_1),$$

where $h_1$ and $h_2$ are any dimension $d$ vectors. Before establishing the mathematics, consider the practical meaning of $S_1$ and $S_2$ in a computer simulation. If we increase/decrease inputs by an amount $h_1$, $S_1$ is the amount of change in the response of a computer simulation. If we then increase/decrease the input by $h_2$, $S_2$ represents the marginal change in the computer simulation response. In general, the relationship between these increments varies based on the size of $h_1$ and $h_2$ and the angle between them.

For many computer simulations, similar changes to the inputs should cause similar marginal changes in the response. A way to measure the changes is to look at the correlation

$$\text{corr}(S_1, S_2) = \frac{\text{cov}(S_1, S_2)}{\sqrt{\text{var}(S_1) \text{var}(S_2)}}.$$

In general, if this correlation is negative, if a process rises on a segment from the origin to $h_1$, it will be expected to decrease in the following segment. Conversely, if this correlation is positive, if a process rises on a segment from the origin to $h_1$, it will expected to continue to increase. Colloquially, the sign of this correlation can be summarized as “what goes up is expected to come down” versus “what goes up is expected to go up.” A pictorial representation where the input space is two dimensional is displayed in Figure 4. As discussed in the supplementary material along with this article, it appears most simulations of physical models have the property of $\text{corr}(S_1, S_2)$ being positive.

Consider the following definition:

**Definition 1.** Let $Z(\cdot)$ be a Gaussian process defined on $\mathbb{R}^d$ with $Z(0) = 0$. For all $h \in \mathbb{R}^d$, define the long range correlation as

$$\tau = \lim_{\kappa \to \infty} \text{corr}\{Z(2\kappa h) - Z(\kappa h), Z(\kappa h)\}.$$

This can be considered a measure of the long range structure of the response surface. Thus, we will use this value to describe the long range behavior of the model. This property is in contrast to a coefficient discussed thoroughly in geostatistics: the Hurst coefficient (Gneiting and Schlather 2004). The Hurst coefficient is meant to only apply to stationary models of the covariance. The default covariance used in the literature is a strictly stationary covariance (Santner, Williams, and Notz 2003), where

$$c(x, x') = c(x + h, x' + h)$$

for all $h \in \mathbb{R}^d$.

These include the Gaussian, Matérn, power exponential, and Cauchy classes of covariances. The conclusion from Christakos (1984) is that a stationary covariance function will approach 0

![Figure 4](image-url)
as the distance between points grows. From this we can investigate the class of all stationary covariances. Returning to the concept of the correlation between directions (proof located in the Appendix and the supplementary material):

**Proposition 1.** Let \( Z(\cdot) \) be a Gaussian process with a strictly stationary covariance function defined on \( \mathbb{R}^d, d \geq 2 \). Then

\[
\tau = -\frac{1}{\sqrt{2}}.
\]

This demonstrates the lack of flexibility of stationary covariances discussed in the introduction of this article. A correlation coefficient of \( \approx -0.7 \) represents a strong negative relationship. All stationary covariance functions require some assumption of negatively correlated increments in the response surface. Even if this property is appropriate for some models, it is discouraging that there is no flexibility based on the response surface of interest. For many response surfaces from computer simulations of physical systems, this is not the expectation of the user. We can expand these results with a bit more information to reflect the rate at which a correlation would approach this negative value. Consider the power exponential covariance given by

\[
\exp(-\|x - x'\|^2/\alpha^2).
\]

We can then derive that

\[
\text{corr} [Z(2\kappa h) + Z(\kappa h), Z(\kappa h)] = \frac{1}{\sqrt{2}}
\]

\[
\sim \exp(-\text{constant} \cdot \kappa^p) \quad \text{as} \quad \kappa \to \infty,
\]

where the constant depends on the matrix \( A \). Thus, not only does it is the long-range correlation approach \(-0.7\), but it does so at an exponential rate.

The proposed lifted Brownian covariance behaves like a fractional Brownian covariance over long distances. We find the following general result (proof located in the Appendix and the supplementary material):

**Proposition 2.** If \( Z(\cdot) \) is a zero mean Gaussian process in \( \mathbb{R}^d \) with a covariance function given in (2), then

\[
\tau = 4^{\beta \gamma - 1} - 1.
\]

From this, if \( 0 < \gamma \beta < 1/2 \), then \( \tau < 0 \); if \( \gamma \beta = 1/2 \), then \( \tau = 0 \); and if \( 1/2 < \gamma \beta < 1 \), then \( \tau > 0 \). This flexibility in long-range behavior is not present in any current popular model of covariance. There are several implications of this, but most important is the flexibility gained when using the model. If our computer model response surfaces have negative long range correlation, thus \( \tau < 0 \), then \( \beta \gamma < 1/2 \) would be appropriate. But if our computer model response surfaces have positive long range correlation, thus \( \tau > 0 \), then \( \beta \gamma > 1/2 \) would be appropriate.

### 4. Implementation

Here, we explicitly describe the use of the lifted Brownian model for prediction based on data from a computer simulation. Our setup includes an initial evaluation at some location, which we denote by \( x_0 \) (which can be arbitrarily chosen among the simulated locations, as described by Zhang and Apley 2015). Then the lifted Brownian model of our response surface is \( y(x) = y(x_0) + \alpha Z(x - x_0) \), with \( y(x_0) \) treated as known and \( Z(\cdot) \) having the covariance in (2). The observed data (in addition to \( x_0, y(x_0) \)) are the inputs \( x_1, \ldots, x_n \) and corresponding responses \( y(x_1), \ldots, y(x_n) \). A point estimate for the evaluation at new point \( x \) is given by

\[
\hat{y}(x) = y(x_0) + c(x)^T C^{-1} (y - y(x_0)),
\]

where \( c(x) \) is the column vector \( c(x - x_0, x_1 - x_0), \ldots, c(x - x_0, x_n - x_0)^T \), \( C \) is a matrix with \( \gamma \)th element \( c(x_i - x_0, x_j - x_0) \) and \( y \) is the column vector \( (y(x_1), \ldots, y(x_n))^T \). Both \( c(\cdot) \) and \( C \) depend on the parameters \( \beta, \alpha, \text{and} \gamma \). The function \( \hat{y}(x) \) is called the kriging predictor. It is optimal in that it is the best linear unbiased estimate of \( y(x) \) under the kriging model. The prediction does not depend on the scaling value \( \alpha \). The expected value of the squared difference between \( y(x) \) and \( y(x) \) is given by

\[
\nu(x) = \alpha \{ c(x - x_0, x - x_0) - c(x)^T C^{-1} c(x) \}.
\]

When the kriging model is assumed to be a Gaussian process, the predictive distribution of the response at a point \( x \) is normal with mean \( \hat{y}(x) \) and variance \( \nu(x) \). Thus, a 90% prediction interval is (approximately) given by

\[
[\hat{y}(x) - 1.64 \sqrt{\nu(x)}, \hat{y}(x) + 1.64 \sqrt{\nu(x)}].
\]

The parameters \( \alpha, \beta, \gamma \) are often estimated from the data. A guide for the various parameters is now summarized. A reasonable decision would be to set

\[
\hat{\alpha} = \frac{(y - y(x_0))^T C^{-1} (y - y(x_0))}{n},
\]

which is its maximum likelihood value. The predominant method would be to find the maximum likelihood value of \( \alpha, \beta, \text{and} \gamma \) as

\[
(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (\hat{\alpha}_1, \ldots, \hat{\alpha}_n, \hat{\gamma}_1, \ldots, \hat{\gamma}_n)
\]

\[
= \arg\min_{\alpha, \beta, \gamma} \log \hat{\alpha} + \frac{1}{n} \log \det(C),
\]

where \( C \) is the covariance matrix given the values of \( \alpha, \beta, \text{and} \gamma \).

We recommend that \( \beta \) and \( \gamma \) are selected by maximizing the likelihood unless one has prior knowledge of the response that suggests a specific choice of \( \beta \) and/or \( \gamma \). For example, if one expects the response surface is analytic, then one can choose \( \gamma = 1 \) instead of estimating it. This is analogous to choosing a Gaussian covariance over a power exponential covariance, as the former is a special case of the latter with exponent parameter corresponding to an analytic surface. We chose \( \gamma = 1 \) in all of our examples, since the response surfaces are known to be smooth. Choosing an appropriate value for \( \beta \) a priori of data may be less straightforward. However, because the product \( \beta \gamma \) relates to long range behavior per Section 3, one could potentially choose this based on whether negatively correlated increments is expected. In summary of the previous section, the value of \( \beta \gamma \) could be chosen to be larger than \( 1/2 \) if one does not expect negatively correlated increments. If negatively correlated increments are expected, then \( \beta \gamma \) could be chosen to be less than \( 1/2 \). All things considered, our general recommendation is for a user to select parameters via a likelihood-based method.
We did not notice any significant problems finding maximum likelihood estimates via numerical packages (relative to existing covariances). Additionally, we have observed evidence that the parameters estimated via maximum likelihood resulted in good out-of-sample predictive accuracy for our examples. Identifiability issues may exist when estimating all of these parameters jointly, as for most covariance models, this potential issue evidently had no adverse effect on prediction in our numerical examples.

5. Pedagogical Illustration

To illustrate the proposed benefits of the model, we will first study a one-dimensional function. Consider the average number of users in a queue when the service rate is one and the input is the arrival rate,

\[ y(x) = \frac{x}{1-x}. \]

This function was studied by Ankenman, Nelson, and Staum (2010) in the context of simulation experiments with stochastic responses. In keeping with the rest of this work, we consider only the deterministic average as the response function.

We now explicitly compare the predictive performance under a standard covariance model and the lifted Brownian covariance model after observing the responses at \( (0, 0.2, 0.4, 0.6, 0.8) \). We choose the Gaussian covariance as the representative standard covariance model,

\[ \exp \left\{ -a(x - x')^2 \right\}, \]

where \( a \) is a lengthscale parameter. We also estimate a mean parameter for the standard model. The details of this are described in the next section. This agrees with the model chosen by Ankenman, Nelson, and Staum (2010) in their original article. For our lifted Brownian model, the value of \( \gamma \) was selected to be 1 because we know the response is smooth. Then \( \beta \) was selected to be 0.75, a somewhat arbitrarily chosen value such that \( \beta \gamma > 1/2 \). Choosing \( x_0 = 0 \) arbitrarily from among the five simulation locations, the covariance model is then

\[ \{1 + a x^2\}^{0.75} + \{1 + a (x')^2\}^{0.75} - \{1 + a (x - x')^2\}^{0.75} - 1. \]

With some abuse of terminology, the predictor under the Gaussian covariance model is referred to as the standard predictor, and the predictor under the lifted Brownian covariance model is referred to as the lifted Brownian predictor.

An interesting feature of the predictions under each model is the behavior when \( a \) is misspecified too large. When this happens, the literature has claimed that the predictive surface will exhibit mean reversion, where the predictions will reduce to the estimated mean of the response surface (Joseph 2006; Lee and Owen 2015). While this effect is acceptable in many geostatistics applications, where a return to some long term average is expected, it can produce illogical conclusions for many computer simulations of physical systems. Five hundred inputs on the segment 0 to 0.8 are selected. The mean square prediction error is calculated by

\[ \frac{1}{500} \sum_{i=1}^{500} \{y(x_i) - \hat{y}(x_i)\}^2. \]

A lower mean square prediction error indicates that the prediction is better. The left panel of Figure 5 shows the mean square prediction error under different values of \( a \) using each covariance. The standard predictor suffers from the mean reversion effect and has a large increase in mean square prediction error as \( a \) grows, as shown in the middle panel of Figure 5. This is because by increasing \( a \) the model is dictated by the long range behavior, which is poorly modeled by the standard covariance. Even as \( a \) grows large, the mean square prediction error of the lifted Brownian predictor does not exhibit similar growth. This is a desirable characteristic for practitioners who use optimization packages to estimate the unknown parameters. Even with some misestimation via an optimizer converging to an incorrect point, the predictor remains well-behaved, because the long-range behavior is properly specified under the lifted Brownian covariance.

Another important consideration regarding the lengthscale parameter is the singularity of the matrix \( C \). \( C \) is inverted when finding the predictor. A large condition number of the matrix results in a numerically unstable prediction (Haaland and Qian 2011). This problem typically gets worse as \( a \) gets small for standard models. The right panel of Figure 5 shows this condition number of \( C \) for the standard and lifted Brownian predictors for small values of \( a \). The respective condition numbers of the matrices appear to be very similar as \( a \) gets small, with the lifted Brownian slightly below the standard for very small values of \( a \). An explicit expression for the condition number of the lifted

![Figure 5](image-url). The left panel shows the mean square prediction error versus the lengthscale parameter \( a \) for the standard predictor (solid) and lifted Brownian predictor (dashed). The middle panel shows an example of mean reversion. The lifted Brownian predictor is given by the dashed line and the standard predictor is given by the solid line when \( a = 20^2 \) along with the true function (thin line). The observations are marked with dots. The right panel shows the condition number of the covariance matrix that is inverted during prediction with various values of \( a \) for each predictor.
Brownian covariance matrix is difficult to derive (even in rate) without onerous assumptions on where the points are observed. But this gives some evidence that it is not a significant problem for this example. More discussion on this topic is in the next section.

6. Numerical Comparisons
The purpose of this section is to document the predictive advantages of the proposed covariance function over a broad class of examples of computer simulations.

6.1 Prediction Methods
This section compares prediction when the standard covariance is replaced with the lifted Brownian covariance. The prediction performance was also compared to the standard R packages mlegp and GPfit, which both default to a standard covariance, but the results using our software were superior.

For the standard predictor, we use a Gaussian covariance model with covariance function
\[ c(x, x') \propto \exp\left(-\|x - x'\|_2^2\right), \]
a variance term \( \alpha \) and a constant prior mean \( \mu \). The value of \( \mu, \alpha, \) and vector \( \mathbf{a} \) are estimated via their maximum likelihood value, \( \hat{\mu}, \hat{\alpha}, \) and \( \hat{\mathbf{a}} \), based on the observations (excluded for brevity).

The predictor under this model at a point \( x \) is
\[ \hat{y}(x) = \hat{\mu} + c(x)^T \mathbf{C}^{-1} (\mathbf{y} - \hat{\mu}). \]
The predictive distribution is given by a normal distribution with mean \( \hat{y}(x) \). The variance of prediction is given by
\[ \nu(x) = \hat{\alpha} \left[ 1 - c(x)^T \mathbf{C}^{-1} c(x) + \frac{(1 - c(x)^T \mathbf{C}^{-1} 1)^2}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \right]. \]
The 90% predictive interval is the same as for the lifted Brownian covariance in the preceding section.

The procedure for prediction under the lifted Brownian covariance model proceeds as in Section 4. Because the functions we are comparing are smooth, it is reasonable to fix \( \gamma \) at 1 and estimate \( \beta \) via maximum likelihood. The origin point was chosen randomly for each design from among the simulated sites. The choice is arbitrary, as the predictor is completely independent of which of the simulated sites are chosen as \( x_0 \) (Zhang and Apley 2015).

The software used to find the correlation parameters via maximum likelihood is of upmost importance to the overall performance of the prediction methodology. Because the likelihood is not convex, local minima must be avoided to ensure that we have found a good maximum likelihood estimate of the correlation parameters. For this comparison, we took the search mechanism from the R software package GPfit to serve as our search. This program was described by MacDonald, Ranjan, and Chipman (2015). The program was modified such that the "nugget effect," which can degrade prediction in this scenario, was fixed at a very small value (10^{-5}). This value can also be decided via cross-validation in lieu of using a fixed value if a practitioner desires (which produced similar conclusions to a fixed value in our examples). The same search mechanism is used for finding the maximum likelihood parameters for both correlation functions. The code used to generate the data for the figures and tables is located in the supplementary materials.

6.2 Exemplar Functions
We first compare the prediction on a set of exemplar functions. Each one has an explicit form, and thus experiments on computer simulations are not needed. But these have many properties similar to simulations of interest in practice. The explicit form of the test functions adds flexibility in analysis that is not possible with a small dataset. Consider the classic beam bending problem, where a beam, operating with linear elasticity with a rectangular cross-section, is fixed on one end and a force is applied at the other end. If the force applied is 600N and the elastic modulus of the beam is 600GPa, the amount of deflection on the force bearing end can be calculated explicitly as
\[ \frac{4}{10^9} \frac{L^3}{bh^3}. \]
We consider this function, along with the following others:

- The midpoint voltage of a transformerless (OTL) circuit function is given by
\[ B \frac{(V_{b1} + 0.74)(R_{c2} + 9)}{B(R_{c2} + 9) + R_f} + 11.35 \frac{R_f}{B(R_{c2} + 9) + R_f} + 0.74 B \frac{R_f}{R_{c2} + 9} \frac{R_{c2} + 9}{B(R_{c2} + 9) + R_f}, \]
where \( V_{b1} = 12 R_{b2}/(R_{b1} + R_{b2}) \), the inputs are \( (R_{b1}, R_{b2}, R_f, R_{c1}, R_{c2}, B) \), and the dimension of the input is 6 (Ben-Ari and Steinberg 2007).
- The resulting cycle time for a piston moving within a cylinder is
\[ 2\pi \sqrt{\frac{M}{k + S^2 \frac{P_0}{T}}}, \]
where \( V = \frac{S}{2k} \sqrt{\frac{A^2 + 4kP_0}{T_0}} \)
and \( A = P_0 S + 19.62 M - \frac{k V_0}{S} \)
where the inputs are \( (M, S, V_0, k, P_0, T, T_0) \) and the dimension of the input is 7 (Kenett and Zacks 1998).
- An often used function to study computer simulation metamodeling is the borehole function,
\[ 2\pi T_u (H_u - H_l) \left\{ \log \left( \frac{r}{r_w} \right) \left( 1 + \frac{2L T_u}{\log(r/r_w) r_w^2 K_v + T_0/T} \right) \right\}^{-1}, \]
where the inputs are \( (T_u, r, r_w, H_u, T_l, H_l, L, K_v) \) and the dimension of the input is 8 (Morris, Mitchell, and Ylvisaker 1993).
- Lastly, consider a model of a light aircraft wing,
\[ 0.036 S_{A0}^{0.758} W_{L0}^{0.035} q^{0.006} A^{0.04} (N_x W_d)^{0.49} \left( \frac{A}{\cos(\Lambda)} \right)^{0.6} \times \left( \frac{100 k}{\cos(\Lambda)} \right)^{-0.3} + S_{A0} W_p, \]
where the inputs are \((S_w, W_{f,w}, A, q, \lambda, t_c, N_z, W_{d_g}, W_p)\) and the dimension of the input is 10 (Moon, Dean, and Santner 2012).

Each of the input regions discussed here is rectangular in a \(d\)-dimensional Euclidean space. The ranges for each of the inputs is given in Table 1.

For our comparison, each experiment consists of a randomly generated Latin hypercube design (Santner, Williams, and Notz 2003) with a sample size of 10\(d\) in keeping with the suggestion of Loepky, Sacks, and Welch (2009). Then 1000 points were randomly chosen uniformly over the input space. Two quantities were recorded. The first was the sample mean of the square prediction error over these 1000 test points (MSPE). We also chose a quantity that measures the adequacy of the uncertainty quantification of the prediction. The quantity used to evaluate the quality of the 90% predictive intervals was the interval score (Gneiting and Raftery 2007). For an observation \(y(x)\), the negatively oriented mean interval score (MIS) is given by the average of the 1000 values of

\[
3.28 \sqrt{\nu(x)} + 20 \left[ \frac{y(x) - \hat{y}(x) - 1.64 \sqrt{\nu(x)}}{\nu(x)} \right]_+ + 20 \left[ y(x) - \hat{y}(x) - 1.64 \sqrt{\nu(x)} \right]_+,
\]

where \([\cdot]_+\) stands for the positive part.

Table 2 displays the mean values over 25 replications for each function, where for each replicate we generated a different Latin hypercube designs and refit the models. The trend throughout Table 2 is that the MSPE and the MIS are improved by using the lifted Brownian covariance, and this is true for all five examples. This indicates that both the prediction and uncertainty quantification is superior using the proposed model. All differences between the standard predictor and the lifted Brownian predictor in Table 2 can be classified as statistically significant by a paired \(t\)-test at the 0.05 level with the exception of the MSPE for the borehole function.

Using a similar setup, we can also compare the lifted Brownian covariance to the Brownian integrated covariance of Zhang and Apley (2015). In that article, when the Brownian integrated covariance is used to predict the Borehole function after a 27-run Latin hypercube designed experiment, which is less than the 80 runs used above, the MSPE was 6.26 after 100 replicates. When using our code, we find an MSPE of 6.32 from 25 samples when the lifted Brownian covariance is used to predict. This indicates these approaches are comparable.

Figure 6 visually indicates the difference in results on a replicate-to-replicate comparison by looking at the percentage change in MSPE and the percentage change in MIS for each simulation. This is given by the standard predictor quantity minus the lifted Brownian quantity divided by the lifted Brownian quantity. All medians are located on the side above zero that corresponds to the lifted Brownian being superior. The entire box is located on the side that corresponds to the lifted Brownian being superior for 7 out of the 10 quantities, with the piston function being the only exception for both quantities.

We now revisit the concept of numerical stability of the prediction. This is dictated by the singularity of the covariance matrix. Since these examples are based on predictors computed numerically, the prediction numerical accuracy is built into the comparison. In these examples, the possible numerical issues of the lifted Brownian predictor appeared to be insignificant compared to the performance gains. This is indirect, but perhaps the most important, evidence of the lack of numerical issues when using the lifted Brownian predictor.

### 6.3 Dataset Analysis

The last subsection provided some evidence supporting the use of lifted Brownian models for some functions that should...
be representative of computer simulations of real physical systems. This subsection details analogous comparisons on known datasets from experiments on computer simulations. The following datasets were used by Zhang and Apley (2015): (i) a model of a G-protein described in Yi et al. (2005) and studied by Loeppky, Sacks, and Welch (2009); (ii) a model of a heat exchanger presented in Qian et al. (2006); (iii) a model of plant canopy introduced in Nilson and Kuusk (1989). In addition, this article will analyze the experiments on computer simulation data from the model of spot welding from Bayarri et al. (2007). We also look at the ocean circulation data from Gough and Welch (1994). Lastly, we consider the data obtained from simulating transport in saturated porous media by the MARTHE software (developed by BRGM, the French Geological Survey); retrieved at http://www благодарством.html. These six datasets have different sample sizes and input dimensions. The number of total simulated input sites were 10^3, 78, 240, 35, 36, 300. For the first four datasets, there is a single response. For the last two, we simply used the first column of responses as our single response for the sake of clarity during comparison.

Our simulation drew a random sample from the existing observations to serve as the experimental data. Then the methods previously described were used to produce predictions and prediction intervals under the standard covariance model and the lifted Brownian covariance model. The remaining data, termed the test data, were used to compute the MSPE and MIS quantities discussed in the previous subsection. The sample size of the experiment was set to be 40, 40, 50, 20, 28, and 100 for the six datasets. This was done to ensure that remaining the data were sufficient to serve as measures of the two quantities of interest.

Table 3 displays the mean values from 25 replications for each function, where on each replicate we generated a different set of experimental observations through random sampling of the dataset. As in the last section, there is a trend of the lifted Brownian predictor outperforming the standard predictor across all six examples. Looking at the sample mean of the MSPEs and MISs, the lifted Brownian predictor outcomes are better than the standard predictor outcomes. This indicates that both the prediction and uncertainty quantification when using the lifted Brownian model are usually better than standard predictor.

Figure 7 visually indicates the difference in results on a replicate-to-replicate comparison, like Figure 6. All medians are located on the side that indicates the lifted Brownian predictor is superior. The entire box is located on the side that indicates

| Data             | LB  | S     | LB−S  | Scale | LB   | S     | LB−S  | Scale |
|------------------|-----|-------|-------|-------|------|-------|-------|-------|
| G-protein        | 5.39| 6.42  | −1.03(0.66) | 10^−4 | 1.02 | 1.16  | −0.14(0.09) | 10^−1 |
| Heat exchanger   | 3.70| 5.82  | −2.11(0.76) | 10^0  | 0.93 | 1.10  | −0.18(0.11) | 10^1  |
| Nilson-Kuusk     | 0.79| 1.24  | −0.45(0.12) | 10^−3 | 1.39 | 1.88  | −0.49(0.11) | 10^−1 |
| Spot welding     | 3.56| 5.12  | −1.57(0.40) | 10^−1 | 3.77 | 4.81  | −1.04(0.32) | 10^0  |
| Ocean circulation| 3.12| 3.37  | −0.25(0.24) | 10^−2 | 1.18 | 1.34  | −0.16(0.08) | 10^0  |
| Marthe           | 7.39| 8.34  | −0.95(0.51) | 10^−2 | 1.01 | 1.17  | −0.16(0.05) | 10^0  |
Figure 7. Boxplots of the simulations described in Section 6.3. The labels for the vertical axis are the same as Figure 6. Some outliers are not plotted for clarity.

lifted Brownian is superior for half of the reported outcomes, although this is nearly the case for all 12 boxes. The variation in these samples comes both from the changes in the experimental design, as well as the changes to the composition of the test set. In some cases, the test sets were very small. Thus, there is a significant amount of replicate-to-replicate variation. However, this simulation study confirms the results from that previous section that predictive ability is improved by using the lifted Brownian model over the standard model. As in the last section, the numerical instability, if it exists, appears to be minor compared to the performance gains.

7. Conclusions

This article has investigated the covariance function of (1) and (2) for Gaussian-process-based response surface modeling of deterministic computer simulations. In particular, we have argued that it is a very fitting choice for many simulation models of real physical systems for which Gaussian process modeling is commonly used. To support this argument, we have demonstrated substantially and consistently superior predictive performance across a variety of real examples (see Figures 6 and 7). We have also derived the covariance function (1) as a limiting form of a Brownian-like covariance function (hence the term lifted Brownian covariance) that is known to have good predictive properties but that has no closed-form expression and is computationally expensive to implement. The lifted Brownian covariance can be evaluated via the closed-form expressions (1) and (2), like any of the standard covariance functions commonly used for kriging. To further explain our empirical findings on the predictive superiority of the lifted Brownian covariance model, we have introduced a new measure of long range behavior and argued that its characteristics for the lifted Brownian covariance are more consistent with many real physical systems. In light of our findings, we recommend that the lifted Brownian covariance be used as an alternative to, or at least considered along with, the standard covariance functions used for kriging with deterministic simulation models.

Appendix: More Investigations of Long-Range Dependence

A.1 Conditional Negative Definiteness of \((1 + \| h \|^2 \gamma) \beta\) Implies Positive Definiteness of (2)

For any vector \(b\) of length \(n\) and sequence of inputs \(x_1, \ldots, x_n\), we have

\[
q = \sum_{i,j=1}^{n} b_i b_j c(x_i, x_j) = \sum_{i,j=1}^{n} b_i b_j \left\{ (1 + \|x_i\|^2 \gamma)^\beta + (1 + \|x_j\|^2 \gamma)^\beta - (1 + \|x_i - x_j\|^2 \gamma)^\beta \right\},
\]

where \(\| \cdot \|\) is the standard Euclidian norm. Let \(b_0 = -\sum_{i=1}^{n} b_i\) and consider a sequences of inputs \(x_0 = 0, x_1, \ldots, x_n\).

\[
q = -\sum_{i,j=1}^{n} b_i b_j (1 + \|x_i - x_j\|^2 \gamma)^\beta - b_0 \sum_{i=1}^{n} b_i (1 + \|x_i\|^2 \gamma)^\beta - b_0 \sum_{j=1}^{n} b_j (1 + \|x_j\|^2 \gamma)^\beta - b_0^2 = -\sum_{i,j=0}^{n} b_i b_j (1 + \|x_i - x_j\|^2 \gamma)^\beta.
\]

From the conditional negative definiteness of \((1 + \| h \|^2 \gamma) \beta\) on \(\mathbb{R}^d\) and the fact that \(\sum_{i=0}^{n} b_i = 0\),

\[
q = -\sum_{i,j=0}^{n} b_i b_j (1 + \|x_i - x_j\|^2 \gamma)^\beta > 0.
\]

A.2 Proof of Theorem 1

We first note that \(\exp(-\cdot)\), which is a completely monotone function, and by Schoenberg’s result we have that \(r(\cdot) = \exp(-\| \cdot \|^2)\) is a positive definite function on \(\mathbb{R}^k\) for all \(k\) (Fasshauer 2005). Clearly, this covariance function is strictly stationary and
isotropic and thus stands as a reasonable covariance function for \( r \). Multiplying by a constant does not affect the positive definite property of this function.

Moreover, \( \zeta_k(x_k) \) can be written as

\[
\zeta_k(x_k) = \int_{\mathbb{R}^d} \left( \| \tilde{x}_k - h \|^p - \| h \|^p \right) \tilde{r}(h) \, dh,
\]

where

\[
\tilde{x}_k = \begin{pmatrix}
\sqrt{\alpha_1} & 0 & \cdots & 0 \\
0 & \sqrt{\alpha_2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sqrt{\alpha_d}
\end{pmatrix} x_k,
\]

and

\[
\tilde{r}_k(h) = \left( \frac{k}{2\pi} \right)^{k/2} \exp(-k\|h\|^2/2).
\]

As shown in Zhang and Apley (2015), \( \zeta_k(x_k) \) only depends on \( \lambda = \| \tilde{x}_k \| = \| \tilde{x}_d \| = \| x \|_a \) and thus we redefine \( \tilde{x}_k \) to be \( \lambda / \sqrt{k} \). Thus,

\[
\left( \frac{k}{2\pi} \right)^{k/2} \int_{\mathbb{R}^d} \| \tilde{x}_k + h \|^{2\beta} - \| h \|^{2\beta} \exp(-k\|h\|^2/2) \, dh
\]

\[
= \int_{\mathbb{R}^d} \| \tilde{x}_k + \frac{h}{\sqrt{k}} \|^{2\beta} - \frac{h}{\sqrt{k}} \|^{2\beta} \left(2\pi \right)^{-k/2} \exp\left(-\frac{\|h\|^2}{2}\right) \, dh.
\]

We can interpret \( (2\pi)^{-k/2} \exp(-\|h\|^2/2) \) as the density function of \( k \) independent Gaussian variables. Let \( Q_1, \ldots, Q_k \) be independent random variables with a Gaussian distributions. Then

\[
\zeta_k(\lambda) = \mathbb{E} \left\{ \left( \lambda^2 + 2\lambda \overline{Q} + Q^2 \right)^\beta - \left( Q^2 \right)^\beta \right\}.
\]

where \( \overline{Q} \) is the sample mean of all \( Q_i \)'s and \( \overline{Q}^2 \) is the sample mean of all \( Q_i^2 \)'s. Now employing the law of large numbers, we have that \( \overline{Q} \to \mathbb{E}(Q_k) = 0 \) and \( \overline{Q}^2 \to \mathbb{E}(Q_k^2) = 1 \) as \( k \to \infty \), and thus we have that

\[
\left( \frac{k}{2\pi} \right)^{k/2} \int_{\mathbb{R}^d} \| \tilde{x}_k + h \|^{2\beta} - \| h \|^{2\beta} \exp(-k\|h\|^2/2) \, dh
\]

\[
\xrightarrow{k \to \infty} (1 + \lambda^2)^\beta - 1 = (1 + \| x \|_a^2)^\beta - 1.
\]

### A.3 Proof of Proposition 1 and Proposition 2

For \( h_1 \) and \( h_2 \) such that \( \| h_1 \_1 = 1, \| h_2 \_2 = 1 \) and \( h_1^T h_2 > 0 \), define the long range correlation between the pair \( h_1 \) and \( h_2 \) as

\[
T(h_1, h_2) = \lim_{k \to \infty} \text{corr} \{ Z(k h_2 + x h_1) - Z(k h_1), Z(x h_1) \}.
\]

Under conditions for Proposition 1, Supplemental Theorem 1 yields

\[
\tau = T(h, h) = -\frac{1}{\sqrt{2}}.
\]
Gough, W. A., and Welch, W. J. (1994), "Parameter Space Exploration of an Ocean General Circulation Model Using an Isopycnal Mixing Parameterization," Journal of Marine Research, 52, 773–796. [174]

Haaland, B., and Qian, P. Z. G. (2011), "Accurate Emulators for Large-Scale Computer Experiments," The Annals of Statistics, 39, 2974–3002. [166,171]

Hardy, R. L. (1971), "Multiquadratic Equations of Topography and Other Irregular Surfaces," Journal of Geophysical Research, 76, 1905–1915. [165]

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004), "Combining Field Data and Computer Simulations for Calibration and Prediction," SIAM Journal on Scientific Computing, 26, 448–466. [165]

Hofmann, T., Schölkopf, B., and Smola, A. J. (2008), "Kernel Methods in Machine Learning," The Annals of Statistics, 36, 1171–1220. [167]

Hofmann, T., Schölkopf, B., and Smola, A. J. (2008), "Kernel Methods in Machine Learning," The Annals of Statistics, 36, 1171–1220. [167]

Joseph, V. R. (2006), "Limit Kriging," Technometrics, 48, 458–466. [171]

Kenett, R., and Zacks, S. (1998), Modern Industrial Statistics: Design and Control of Quality and Reliability, Pacific Grove, CA: Duxbury Press. [165]

Lee, M. R., and Owen, A. B. (2015), "Single Nugget Kriging." arXiv preprint arXiv:1507.05128. [171]

Lindstrom, T. (1993), "Fractional Brownian Fields as Integrals of White Noise," Bulletin of the London Mathematical Society, 25, 83–88. [167]

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. [165,173]

MacDonald, B., Ranjan, P., and Chipman, H. (2015), "GPfit: An R Package for Fitting a Gaussian Process Model to Deterministic Simulator Outputs," Journal of Statistical Software, 64, 1–23. [172]

Mandelbrot, B. B., and Van Ness, J. W. (1968), "Fractional Brownian Motions, Fractional Noises and Applications," SIAM Review, 10, 422–437. [165]

Matheron, G. (1963), "Principles of Geostatistics," Economic Geology, 58, 1246–1266. [165]

Mitchelli, C. A. (1986), "Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions," Constructive Approximation, 2, 11–22. [167]

Moon, H., Dean, A. M., and Santner, T. J. (2012), "Two-Stage Sensitivity-Based Group Screening in Computer Experiments," Technometrics, 54, 376–387. [165]

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, 243–255. [165]

Nilson, T., and Kuusk, A. (1989), "A Reflectance Model for the Homogeneous Plant Canopy and its Inversion," Remote Sensing of Environment, 27, 157–167. [174]

Peng, C.-Y., and Wu, C. J. (2014), "On the Choice of Nugget in Kriging Modeling for Deterministic Computer Experiments," Journal of Computational and Graphical Statistics, 23, 151–168. [166]

Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K., and Wu, C. F. J. (2006), "Building Surrogate Models Based on Detailed and Approximate Simulations," Journal of Mechanical Design, 128, 668–677. [174]

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments," Statistical Science, 4, 409–423. [165,166]

Santner, T. J., Williams, B. J., and Notz, W. (2003), The Design and Analysis of Computer Experiments, New York: Springer Science & Business Media. [165,168,169,173]

Stein, M. L. (1999), Interpolation of Spatial Data: Some Theory for Kriging, New York: Springer Science & Business Media. [165]

Yi, T.-M., Fazel, M., Liu, X., Otitoju, T., Goncalves, J., Papachristodoulou, A., Prajna, S., and Doyle, J. (2005), "Application of Robust Model Validation Using SOSTOOLS to the Study of G-Protein Signaling in Yeast," Proceedings of the First Conference on Foundations of Systems Biology in Engineering. Available: https://faculty.washington.edu/mfazel/FOSBE.html [174]

Zhang, N., and Apley, D. W. (2014), "Fractional Brownian Fields for Response Surface Metamodeling," Journal of Quality Technology, 46, 285–301. [165,166,168]

——— (2015), "Brownian Integrated Covariance Functions for Gaussian Process Modeling: Sigmoidal Versus Localized Basis Functions," Journal of the American Statistical Association, DOI: 10.1080/01621459.2015.1077711. [165,166,167,168,170,172,173,176]