Statistica Sinica **12**(2002), 863-883

PARAMETER ESTIMATION FOR FRACTIONAL BROWNIAN SURFACES

Zhengyuan Zhu and Michael L. Stein

University of Chicago

Abstract: We study the use of increments to estimate the fractal dimension of a two-dimensional fractal Brownian surface observed on a regular grid. Linear filters are used to describe differencing of two-dimensional surfaces and generalized variograms are defined based on them. We examine the practical performance of ordinary and generalized least squares estimators based on different filters by numerically calculating their asymptotic variance and also by simulation using an exact simulation method for fractional Brownian surface proposed by Stein (2002). An extensive numerical and simulation study of the practical performance of estimators based on different collections of lags is presented for Gaussian and non-Gaussian random fields, and a comparison to the much more computationally intensive restricted maximum likelihood estimator is provided.

Key words and phrases: Fractal, fractional Gaussian random field, generalized least squares, self-similar, variogram, REML.

1. Introduction

Fractional Brownian fields are a class of Gaussian random fields that can be used to model many physical processes in space, see for example Peitgen and Saupe (1988) and Wilson (2000). A fractional Brownian field in $\mathbb{R}^d$ is a Gaussian random field with constant mean and covariance structure $\mathrm{Var}\{Z(\mathbf{x}) - Z(\mathbf{y})\} = 2C\|\mathbf{x}-\mathbf{y}\|^\alpha$, $\mathbf{x}$, $\mathbf{y}\in\mathbb{R}^d$, where $\alpha \in (0,2)$. The quantity $\gamma(\mathbf{x}-\mathbf{y}) = \frac{1}{2}\mathrm{Var}\{Z(\mathbf{x}) - Z(\mathbf{y})\}$ is known as the variogram (Cressie (1993, p.40)). The parameter C is a scale parameter and α is a smoothness parameter with larger values corresponding to smoother surfaces. There is a simple relation between α and the fractal dimension D of a fractional Brownian surface: $D = d + 1 - \alpha/2$ (Cressie (1993, p.311)). A number of authors in recent years have investigated the statistical properties of estimators of α using the empirical variogram based on a finite number of observations of Z on a regular grid, see for example Constantine and Hall (1994), Kent and Wood (1997), Davies and Hall (1999) and Chan and Wood (2000). The key idea can be easily explained in the one-dimensional case, where a fractional Brownian field is called fractional Brownian motion. Suppose we observe a fractional Brownian motion at n equally spaced locations with spacing n^{-1}. Let $Y_{k,n} = \sum_{i=1}^{n-k} \{Z((i+k)/n) - Z(i/n)\}^2/(n-k)$ be the empirical

variogram at lag k, which has the property

$$\log(n^{\alpha}Y_{k,n}) = \log Y_{k,n} + \alpha \log n \xrightarrow{p} \log C + \alpha \log k \tag{1.1}$$

as $n \to \infty$. We can calculate the empirical variogram at several different values of k and use ordinary least squares (OLS) or generalized least squares (GLS) to get an estimator of α.

In this paper we focus on estimators of α for fractional Brownian surfaces in $\mathbb{R}^2$ observed on a regular lattice, but we also discuss estimating C. We assume the grid is square and spacing is equal in both directions. Because of the self-similarity of fractional Brownian surface, without loss of generality we can assume observations are made on an $n \times n$ regular lattice with spacing n^{-1} between neighboring observations, so the sample size is $N = n^2$.

Kent and Wood (1997) used the empirical variogram of increments to estimate fractal dimension of one-dimensional locally self-similar Gaussian processes. Chan and Wood (2000) applied this method to two-dimensional Gaussian surfaces and derived the rate of convergence for OLS and GLS estimators of α. They also compared the performance of several estimators based on different filters (increments) using simulation, but they used an approximate simulation procedure for Gaussian random fields with α near 2. Our results show that their approximation yields seriously erroneous conclusions about the properties of some estimators when α is near 2.

We expand on the work of Chan and Wood (2000) on estimating α in three ways. First, for two-dimensional surfaces there are far more ways of differencing than for one-dimensional processes, and we consider a broader spectrum of filters than Chan and Wood (2000). Furthermore, we use rotations of filters in our estimators, whereas Chan and Wood (2000) only use estimators based on scale transformations of a filter. Asymptotic theory and simulations demonstrate that our estimators are substantially more efficient than those of Chan and Wood. See Section 2.3 and 4.4 for more details. Second, in our study we take advantage of a new exact simulation method for fractional Brownian surfaces proposed by Stein (2002), yielding meaningful and accurate comparisons between estimators for α near 2. Third, in those cases in which $N^{1/2}\hat{\alpha}$ converges in distribution, we calculate its asymptotic variance numerically. For fractional Brownian surfaces, the variogram-based estimators of α are asymptotically unbiased. Thus, for estimators that have the same convergence rate, we can compare them through their asymptotic variances.

It is known that the performance of estimators of α depends on the collection of lags on which the empirical variogram is calculated (we refer to this collection of lags as the lag set). Constantine and Hall (1994) proved that, for the OLS estimator, as $n \to \infty$, the asymptotically optimal number of lags, i.e., the size of the lag set, is bounded. Davies and Hall (1999) include one numerical example showing that the mean-squared error (MSE) of the OLS estimator of α for a

Gaussian process on the interval [0,1] is optimized at lag set size 2 and the MSE increases as the lag set size increases. No systematic study has been done on this matter, especially in two dimensions. We present rather extensive numerical and simulation studies to compare the performance of OLS and GLS estimators of α for fractional Brownian surfaces on different lag sets. We find that for the OLS estimator, increasing the number of lags beyond 2 always yields poorer estimators, which is consistent with Davies and Hall (1999). For the GLS estimator, the best choice of lag set is sensitive to the sample size, with bigger sample sizes corresponding to larger best lag sets.

Chan and Wood (2000) study theoretical properties of Gaussian processes that behave only locally like a fractional Brownian surface, as well as processes that are not isotropic and some that are not Gaussian. We consider only exactly self-similar fractional Brownian surfaces, allowing a more focused simulation study on the estimation of α without the need to consider the specific form for the variogram of the random field or the spacing of the observations. Our theoretical results are stated just for fractional Brownian surfaces, but these results can be extended to random fields whose behavior is sufficiently similar to a fractional Brownian surface: to those stationary Gaussian fields with covariance functions of the form (2.5) in Chan and Wood (2000) that satisfy certain regularity conditions, and to certain non-Gaussian random fields. In the simulation study we also apply our method to stationary Gaussian and non-Gaussian fields in addition to fractional Brownian surface.

Section 2 defines linear filters and generalized variograms based on them. Section 3 gives details of the estimation methodology and shows how to calculate the asymptotic variance of $N^{1/2}\hat{\alpha}$ when it converges in distribution. Results for estimating C are given as well. In Section 4 we give our simulation results for Gaussian and non-Gaussian random fields, including comparisons to the restricted maximum likelihood estimator and the estimator defined in Chan and Wood (2000).

2. Linear Filter and Empirical Variogram

2.1. Definition of linear filter

Suppose Z is a random field on $\mathbb{R}^d$, and $A = \{a_1, a_2, \ldots, a_I\}$ and $\mathbf{\Delta} = \{\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_I\}$ are ordered sets, $a_i \in \mathbb{R}$ and $\boldsymbol{\delta}_i \in \mathbb{R}^d$. We define a linear filter $L = \{A, \mathbf{\Delta}\}$ to be the operation on Z such that $L(Z(\mathbf{x})) = \sum a_i Z(\mathbf{x} + \boldsymbol{\delta}_i)$. A filter is called finite if it is based on a finite number of observations. We are mainly interested in linear filters on two dimensional regular lattices such that $\boldsymbol{\delta}_i \in \mathbb{Z}^2$ and either $\sum a_i = 0$ or

$$\sum a_i = 0 \text{ and } \sum a_i \boldsymbol{\delta}_i = \mathbf{0}. \tag{2.1}$$

For any $c \neq 0$, the difference between $\{A, \mathbf{\Delta}\}$ and $\{cA, \mathbf{\Delta}\}$ is of no importance for our purpose, so we identify those filters as one filter.

2.2. Classification of linear filters on regular lattices

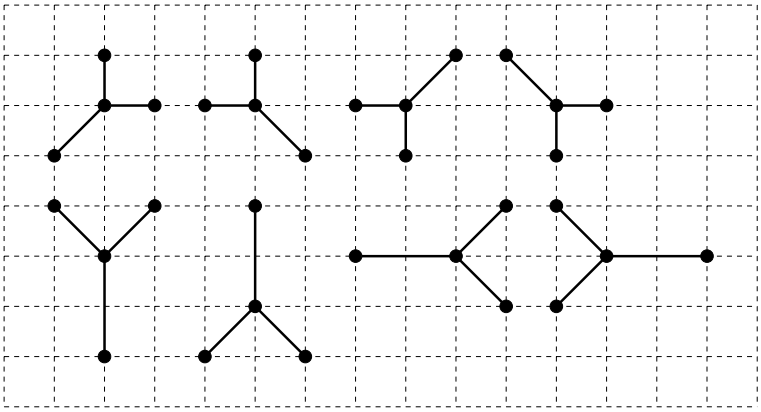

Figure 1. Transformations of the linear filter $\{A = \{1, 1, 1, -3\}, \boldsymbol{\Delta} = \{(1, 0), (0, 1), (-1, -1), (0, 0)\}\}$.

Unlike the one-dimensional case, different two-dimensional filters can have the same basic geometric structure, as shown in Figure 1. As we shall see in Section 3.1, by using filters having the same geometric structure, similar to (1.1), we obtain a linear regression for estimating α. To represent filters having the same geometric structure, we introduce the transformation of a linear filter T, defined on a linear filter L by $T \circ L = \{A, \mathbf{T}\boldsymbol{\Delta}\}$, where $\mathbf{T}$ is a 2×2 matrix and $\mathbf{T}\boldsymbol{\Delta} = \{\mathbf{T}\boldsymbol{\delta}_1, \mathbf{T}\boldsymbol{\delta}_2, \ldots, \mathbf{T}\boldsymbol{\delta}_I\}$. There is a trivial isomorphism between the transformation T and the matrix $\mathbf{T}$, so from now on we do not distinguish them.

Since we assume all the observations are on a regular grid, we are only interested in filters for which all $\boldsymbol{\delta}_i \in \mathbb{Z}^2$. We restrict our attention to the semigroup $\mathbf{G}$ of 2×2 matrices $\mathbf{T}$ generated by the matrices $\mathbf{R}$, $\mathbf{S}$, $\mathbf{D}$ and $k\mathbf{I}$, $k \in \mathbb{R}^+$, through matrix multiplication, where $\mathbf{I}$ is the 2×2 identity matrix and

$$\mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

$\mathbf{R}$ corresponds to 90° rotation, $\mathbf{S}$ is reflection with respect to the vertical axis and $\mathbf{D}$ is 45° rotation plus $2^{1/2}$ scaling. We could consider other transformations but we doubt there is much practical advantage to doing so.

The semigroup of transformations $\mathbf{G}$ partitions the space of all linear filters into categories; within each category one filter can be mapped into another using one of the transformations in $\mathbf{G}$. Each category can be further partitioned into subcategories using the subgroup of $\mathbf{G}$ generated by $k\mathbf{I}$, $k \in \mathbb{R}^+$. We use a filter to represent a category or a subcategory in different situations, and the meaning will be clear from the context.

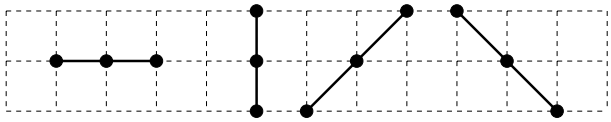

Figure 2. Representations of the four subcategories of $\{A = \{1, -2, 1\},\ \boldsymbol{\Delta} = \{(1,0),(0,0),(-1,0)\}\}$.

Figure 2 shows four filters that belong to the same category, $\{A = \{1, -2, 1\}$, $\boldsymbol{\Delta} = \{(1,0),(0,0),(-1,0)\}\}$, each of which represents a subcategory. The elements in every subcategory are all scalings of the representative filter of the subcategory.

We restrict our attention to categories of filters that have relatively simple forms: all have no more than 5 points and two distinct values for the a_i's. The categories of filters under study are as follows:

$$
\begin{aligned}
L^{(1)} &= \{A = \{1,1,-2\}, && \boldsymbol{\Delta} = \{(1,0),(-1,0),(0,0)\}\},\\
L^{(2)} &= \{A = \{1,1,1,-3\}, && \boldsymbol{\Delta} = \{(1,0),(0,1),(-1,-1),(0,0)\}\},\\
L^{(3)} &= \{A = \{1,1,-1,-1\}, && \boldsymbol{\Delta} = \{(1,0),(0,1),(1,1),(0,0)\}\},\\
L^{(4)} &= \{A = \{1,1,-1,-1\}, && \boldsymbol{\Delta} = \{(1,1),(0,-1),(1,0),(0,0)\}\},\\
L^{(5)} &= \{A = \{1,1,1,1,-4\}, && \boldsymbol{\Delta} = \{(1,0),(0,1),(-1,0),(0,-1),(0,0)\}\},\\
L^{(6)} &= \{A = \{1,1,1,1,-4\}, && \boldsymbol{\Delta} = \{(1,0),(-1,0),(1,1),(-1,-1),(0,0)\}\}.
\end{aligned}
$$

These filters satisfy (2.1) and are increments of order 1 as defined in Chan and Wood (2000). We also consider $L^{(0)} = \{A = \{1,-1\}, \boldsymbol{\Delta} = \{(1,0),(0,0)\}\}$, which only satisfies $\sum a_i = 0$ and gives increments of order 0.

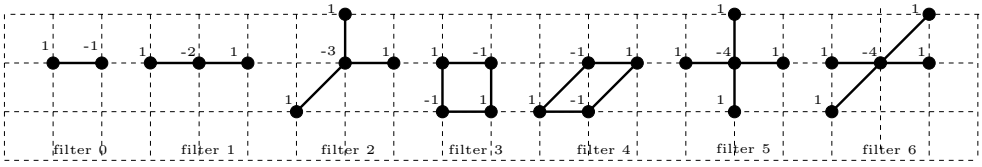

Figure 3. Filters $L^{(0)}$,$L^{(1)}$, ...,$L^{(6)}$

Figure 3 illustrates the shape of the representing filters for each category. Notice that $L^{(1)}$ has four subcategories (self, 90° rotation, and 45° and 135° rotation plus $2^{1/2}$ scaling, see Figure 2), $L^{(2)}$ has eight subcategories (self, 90°, 180°, 270° rotations and 45°, 135°, 225°, 315° rotations plus $2^{1/2}$ scaling, see Figure (1). Similarly there are two subcategories for $L^{(3)}$ and $L^{(5)}$, and eight for $L^{(4)}$ and $L^{(6)}$.

Among the filters considered in Chan and Wood (2000), "horizontal", "vertical", "diagonal with positive gradient" and "diagonal with negative gradient" correspond to the four subcategories of $L^{(1)}$ (see Figure 2) and "square-increment" corresponds to one subcategory of $L^{(3)}$.

2.3. Definition of generalized variogram

The generalized variograms we consider are based on filters from a single category. We only consider isotropic random fields, so we include all filters in a particular category that have the same scale to form the generalized variogram. Formally, define $|T|$, the norm of T, as the square root of the determinant of the matrix $\mathbf{T}$, and let M be the number of distinct linear filters $T \circ L$ such that $|T| = 1$. As in Chan and Wood (2000), for notational convenience, we write our estimators assuming that we can observe $T \circ L(Z(\mathbf{x}))$ on an $n \times n$ grid for all T we are considering. This require observing Z on a slightly larger grid and then not making use of some increments for smaller values of $|T|$. In our simulations, we always use all available increments based on observations on an $n \times n$ grid, so that the number of increments changes with T and is always smaller than n^2.

Set $N = n^2$. Define the generalized variogram for the filter L on an $n \times n$ regular lattice at lag k to be

$$Y_{k,n} = \frac{1}{2MN} \sum_{\mathbf{x} \in D_n} \sum_{\mathcal{T}_{k,n}} \left\{ T_j \circ L \left(Z \left(\mathbf{x} \right) \right) \right\}^2,$$

where $D_n = \{1/n, 2/n, \ldots, 1\}^2$, $\mathcal{T}_{k,n} = \{T \circ L : |T| = k/n\}$ for $k \leq n$, and $k \in \mathbb{N}$ or $k/2^{1/2} \in \mathbb{N}$, $\mathbb{N}$ the set of all natural numbers.

To give a specific example of a $Y_{k,n}$, for Filter 1, $M = 2$ and the generalized variogram is

$$Y_{k,n} = \frac{1}{4N} \sum_{\mathbf{x} \in D_n} \left\{ \left(\frac{k}{n} \mathbf{I} \right) \circ L^{(1)} \left(Z \left(\mathbf{x} \right) \right)^2 + \left(\frac{k}{n} \mathbf{R} \right) \circ L^{(1)} \left(Z \left(\mathbf{x} \right) \right)^2 \right\}.$$

For $k \in \mathbb{N}$, $Y_{k,n}$ is a combination of variograms of "vertical" and "horizontal" increments, and for $k/2^{1/2} \in \mathbb{N}$, $Y_{k,n}$ is a combination of variograms of the "diagonal with positive gradient" and the "diagonal with negative gradient" increments, as defined in Chan and Wood (2000).

3. Theoretical Results

We first give asymptotic results for estimators of α and C and details of the estimation methodology. Then we show how to calculate the asymptotic variance of $N^{1/2}\hat{\alpha}$ when it converges in distribution. We numerically calculate the asymptotic variance of $N^{1/2}\hat{\alpha}$ for a group of estimators of α and find some interesting trends.

3.1. Asymptotic results

For a filter $L = (A, \mathbf{\Delta})$, define $f_L(\alpha) = -\sum a_s a_t \|\boldsymbol{\delta}_s - \boldsymbol{\delta}_t\|^\alpha / 2$, which we write as $f(\alpha)$ when the filter L is apparent from context. Setting $C' = Cf(\alpha)$,

for $Z_{k,n} = n^{\alpha} Y_{k,n}$, it follows that

$$\mu_k = \mathrm{E}\left(Z_{k,n}\right) = -\frac{n^{\alpha}}{2} \sum_{s,t} a_s a_t \gamma\left(\left\|\frac{k}{n}(\boldsymbol{\delta}_s - \boldsymbol{\delta}_t)\right\|\right) = C' k^{\alpha}. \tag{3.1}$$

Let $\sigma_{k,l,n} = \mathrm{Cov}\left(Z_{k,n}, Z_{l,n}\right)$. We have

$$\begin{aligned}
\sigma_{k,l,n} &= \frac{n^{2\alpha}}{4M^2 n^4} \sum_{j,j'} \sum_{\mathbf{x},\mathbf{y}} \mathrm{Cov}\left(\left\{T_j \circ L\left(Z\left(\mathbf{x}\right)\right)\right\}^2, \left\{T_{j'} \circ L\left(Z\left(\mathbf{y}\right)\right)\right\}^2\right) \\
&= \frac{n^{2\alpha}}{4M^2 n^4} \sum_{j,j'} \sum_{\mathbf{x},\mathbf{y}} 2[\mathrm{E}\left\{T_j \circ L\left(Z\left(\mathbf{x}\right)\right) T_{j'} \circ L\left(Z\left(\mathbf{y}\right)\right)\right\}]^2 \\
&= \frac{n^{2\alpha}}{2M^2 n^4} \sum_{j,j'} \sum_{\mathbf{x},\mathbf{y}} \left\{\sum a_s a_t \mathrm{E}\left(Z\left(\mathbf{x} + \mathbf{T}_j \boldsymbol{\delta}_s\right) Z\left(\mathbf{y} + \mathbf{T}_{j'} \boldsymbol{\delta}_t\right)\right)\right\}^2 \\
&= \frac{C n^{2\alpha}}{2M^2 n^4} \sum_{j,j'} \sum_{\mathbf{x},\mathbf{y}} \left\{\sum a_s a_t \|\mathbf{x} - \mathbf{y} + \mathbf{T}_j \boldsymbol{\delta}_s - \mathbf{T}_{j'} \boldsymbol{\delta}_t\|^{\alpha}\right\}^2,
\end{aligned}$$

where $\mathbf{x}, \mathbf{y}$ are summed over D_n, j is summed over $|T_j| = k/n$, and j' over $|T_{j'}| = l/n$.

If we can show $Z_{k,n} \xrightarrow{p} C f(\alpha) k^{\alpha}$ or, equivalently,

$$\log Z_{k,n} \xrightarrow{p} \log C + \log f(\alpha) + \alpha \log k, \tag{3.2}$$

then for any subset $\mathcal{K}_p = \{k_1, k_2, \ldots, k_p\} \subset \mathcal{K}$ where $\mathcal{K} = \{k : k < n, k \in \mathbb{N} \text{ or } k = \sqrt{2}k', k' \in \mathbb{N}\}$ and $p \geq 2$, we can use (3.2) to get a regression estimator of α and C.

Theorem 1. *For any finite filter $L = (A, \boldsymbol{\Delta})$ satisfying* (2.1),

$$N\sigma_{k,l,n} \longrightarrow \sigma_{k,l} = \frac{C}{2M^2} \sum_{j,j'} \sum_{\mathbf{x}} \left\{\sum a_s a_t \|\mathbf{x} + T_j^* \boldsymbol{\delta}_s - T_{j'}^* \boldsymbol{\delta}_t\|^{\alpha}\right\}^2, \tag{3.3}$$

where $\mathbf{x} \in \mathcal{Z}^2$ and $|T_j^*| = |T_{j'}^*| = 1$. Moreover, letting $\mathbf{Z}_n = (Z_{k_1,n}, Z_{k_2,n}, \ldots, Z_{k_p,n})^T$,

$$N^{1/2}\left\{\mathbf{Z}_n - \mathrm{E}\left(\mathbf{Z}_n\right)\right\} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \boldsymbol{\Sigma}\right) \tag{3.4}$$

where $\boldsymbol{\Sigma} = (\sigma_{k,l})$. If $\alpha < 1$, then (3.3) *and* (3.4) *hold for any filter L satisfying $\sum a_i = 0$.*

Theorem 1 combined with (3.1) implies (3.2), which is the foundation of the variogram method for estimating α.

We next consider the estimator of α of the form

$$\hat{\alpha} = \sum_{l} g_l \log Z_{k_l,n}, \tag{3.5}$$

where $g_l,\ l = 1, \ldots, p$, is a set of numbers such that

$$\sum_l g_l = 0,\ \sum_l g_l \log(k_l) = 1. \tag{3.6}$$

Theorem 2. *For any finite filter L which satisfies* (2.1) *and $\hat{\alpha}$ defined as in* (3.5),

$$N^{1/2}\{\hat{\alpha} - \alpha\} \xrightarrow{\mathcal{L}} N\left(0, \sigma_\alpha^2\right), \tag{3.7}$$

$$\sigma_\alpha^2 = \sum_{s,t} \frac{g_s g_t}{\mu_{k_s}\mu_{k_t}} \sigma_{s,t}. \tag{3.8}$$

If $\alpha < 1$, then (3.7) *holds for any filter L satisfying $\sum a_i = 0$.*

Notice that although $Z_{k,n} = n^\alpha Y_{k,n}$ is not observable, for g_ls satisfying (3.6), $\hat{\alpha}$ is a statistic that only depends on the observable part $Y_{k,n}$ and does not depend on C or α.

Theorem 2 covers the OLS estimator and GLS estimator with nonstochastic weight matrix. However, when using GLS in practice, the weight matrix also has to be estimated from the data. The following theorem gives the asymptotic behavior for the GLS estimator with estimated weight matrix.

Theorem 3. *For any finite filter $L = (A, \boldsymbol{\Delta})$ satisfying* (2.1), *consider $\hat{\alpha} = \sum_l \hat{g}_{l,n} \log Z_{k_l,n}$, where $\hat{g}_{l,n} \xrightarrow{p} g_l$ and the g_ls are nonstochastic constants. If both $\hat{g}_{l,n}$ and g_l satisfy* (3.6), *then*

$$N^{1/2}\{\hat{\alpha} - \alpha\} \xrightarrow{\mathcal{L}} N\left(0, \sigma_\alpha^2\right), \tag{3.9}$$

where σ_α^2 is as defined in (3.8). *If $\alpha < 1$, then* (3.9) *holds for any filter L satisfying $\sum a_i = 0$.*

In practice, estimating the scale parameter C may also be of interest, for example, to obtain standard errors for kriging predictors. Consider the estimator of $C' = Cf(\alpha)$ of the form

$$\hat{C}' = n^{\hat{\alpha}-\alpha} \exp\left\{\sum_l h_l \log Z_{k_l,n}\right\}, \tag{3.10}$$

where $\hat{\alpha}$ is the same as (3.5) and $h_l,\ l = 1, \ldots, p$, is a set of numbers such that

$$\sum_l h_l = 1,\ \sum_l h_l \log(k_l) = 0. \tag{3.11}$$

Theorem 4. *For any finite filter $L = (A, \boldsymbol{\Delta})$ that satisfies* (2.1) *and $\hat{C}'$ defined as in* (3.10),

$$\frac{N^{1/2}}{\log N}\left\{\hat{C}' - C'\right\} \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{4}\sigma_\alpha^2\right), \tag{3.12}$$

where σ_α^2 *is as defined in* (3.8). *If* $\alpha < 1$, *then* (3.12) *holds if* $\sum a_i = 0$.

The following corollary of Theorem 4 is immediate.

Corollary 1. *For any finite filter* $L = (A, \mathbf{\Delta})$ *that satisfies* (2.1), $\hat{C}'$ *defined as in* (3.10), *and* $\hat{C} = \hat{C}'/f(\hat{\alpha})$,

$$\frac{N^{1/2}}{\log N}\left\{\hat{C} - C\right\} \xrightarrow{\mathcal{L}} N\left(0, \frac{\sigma_\alpha^2}{4f(\alpha)^2}\right), \tag{3.13}$$

where σ_α^2 *is as defined in* (3.8). *If* $\alpha < 1$, *then* (3.13) *holds if* $\sum a_i = 0$.

Theorem 1 follows directly from Theorem 3.2 in Chan and Wood (2000). The proofs of Theorems 2–4 can be found in Zhu and Stein (2001).

Notice that, because of (3.11), the estimators in Theorem 4 and Corollary 1 are statistics that only depend on $Y_{k_l,n}$ and do not depend on C or α. Both estimators defined by (3.17) and (3.18) satisfy (3.11). Thus, Theorem 4 and Corollary 1 can be applied to both the "OLS" and "GLS" estimators of C. The asymptotic results for the "GLS" estimators of C with estimated weight matrix can be obtained in a similar way as when estimating α (see Theorem 3). These are omitted.

3.2. Numerical calculation

Theorems 2 and 3 give us a way to compare estimators of α by numerically calculating their asymptotic variances. We proceed as follows. Suppose we use $\mathcal{K}_p = \{k_1, k_2, \ldots, k_p\}$ as the lag set to estimate α, and let $\hat{\alpha}_O$ and $\hat{\alpha}_G$ be the OLS and GLS estimators of α using filter L. We write $\hat{\alpha}$ to indicate either $\hat{\alpha}_O$ or $\hat{\alpha}_G$. Let $\mathbf{W}$ be the weight matrix of filter L used in GLS, $d_j = \log k_j$, $\mathbf{d} = (d_1, d_2, \ldots, d_p)^T$ and let $\bar{d}$ be the mean of $\mathbf{d}$. It follows that for $\hat{\alpha}_O$,

$$g_l = \frac{d_l - \bar{d}}{\sum (d_j - \bar{d})^2}, \tag{3.14}$$

and for $\hat{\alpha}_G$,

$$g_l = \frac{(\mathbf{1}^T\mathbf{W}\mathbf{1})(\mathbf{d}^T\mathbf{W})_l - (\mathbf{1}^T\mathbf{W}\mathbf{d})(\mathbf{1}^T\mathbf{W})_l}{(\mathbf{1}^T\mathbf{W}\mathbf{1})(\mathbf{d}^T\mathbf{W}\mathbf{d}) - (\mathbf{1}^T\mathbf{W}\mathbf{d})^2}, \tag{3.15}$$

where $\mathbf{1}$ is a vector of ones and $(\mathbf{V})_l$ is the lth component of a vector $\mathbf{V}$. It is easy to show that in both cases the g_ls satisfy (3.6), so Theorem 2 can be applied to OLS estimators and to GLS estimators with nonstochastic weight matrix. When the weight matrix is estimated by replacing α by $\hat{\alpha}_O$ in the expressions for the weights, since $\hat{\alpha}_O$ is a consistent estimator of α by Theorem 2, Theorem 3 applies

as long as the weight matrix is a continuous function of α. All weight matrices used here are continuous functions of α.

In order to calculate the asymptotic variance for the GLS estimator of α, we define the asymptotic weight matrix as

$$W_{s,t} = \lim_{n\to\infty} \frac{N\mathrm{Cov}\left(Z_{k_s,n}, Z_{k_t,n}\right)}{\mathrm{E}\left(Z_{k_s,n}\right)\mathrm{E}\left(Z_{k_t,n}\right)} = \frac{\sigma_{k_s,k_t}}{\mu_{k_s}\mu_{k_t}}. \tag{3.16}$$

This formula is obtained by using the first-order Taylor expansion to approximate $N\mathrm{Cov}\left(\log Z_{k_s,n}, \log Z_{k_t,n}\right)$ and then taking the limit as $n \to \infty$. Thus, using (3.8) and (3.14)—(3.16), we can numerically calculate the asymptotic variance of estimators of α based on different filters L and lag sets $\mathcal{K}_p$ for both OLS and GLS estimators.

Once we have the estimator of α, it is easy to get an estimator of C using OLS or GLS. We can estimate C' by

$$\hat{C}'_O = \frac{n^{\hat{\alpha}}}{n^{\alpha}} \exp\left\{p^{-1}\sum_{l=1}^{p} \log Z_{k_l,n} - \hat{\alpha}\bar{d}\right\} = n^{\hat{\alpha}_O} \exp\left\{p^{-1}\sum_{l=1}^{p} \log Y_{k_l,n} - \hat{\alpha}\bar{d}\right\}, \tag{3.17}$$

where p is the number of lags in the lag set. Or, using GLS we have

$$\begin{aligned}\hat{C}'_G &= \frac{n^{\hat{\alpha}_G}}{n^{\alpha}} \exp\left\{\frac{(\mathbf{d}^T\mathbf{W}\mathbf{d})(\mathbf{1}^T\mathbf{W}\log\mathbf{Z}_n) - (\mathbf{1}^T\mathbf{W}\mathbf{d})(\mathbf{d}^T\mathbf{W}\log\mathbf{Z}_n)}{(\mathbf{1}^T\mathbf{W}\mathbf{1})(\mathbf{d}^T\mathbf{W}\mathbf{d}) - (\mathbf{1}^T\mathbf{W}\mathbf{d})^2}\right\} \\ &= n^{\hat{\alpha}_G} \exp\left\{\frac{(\mathbf{d}^T\mathbf{W}\mathbf{d})(\mathbf{1}^T\mathbf{W}\log\mathbf{Y}_n) - (\mathbf{1}^T\mathbf{W}\mathbf{d})(\mathbf{d}^T\mathbf{W}\log\mathbf{Y}_n)}{(\mathbf{1}^T\mathbf{W}\mathbf{1})(\mathbf{d}^T\mathbf{W}\mathbf{d}) - (\mathbf{1}^T\mathbf{W}\mathbf{d})^2}\right\},\end{aligned} \tag{3.18}$$

where $\mathbf{Y}_n = (Y_{k_1,n}, Y_{k_2,n}, \ldots, Y_{k_p,n})^T$ and log of a vector just means the vector of logarithms of its components. For Filter 0, $f(\alpha) = 1$, so we can estimate C by (3.17) or (3.18) directly. For Filter 1, $f(\alpha) = 4 - 2^{\alpha}$, so we can estimate C by (3.17) or (3.18) divided by $4 - 2^{\hat{\alpha}}$. It is easy to check that Theorem 4 and Corollary 1 can be applied to both the OLS and GLS estimators of C, and the asymptotic variance can be obtained.

From Theorems 2–4, the OLS and GLS estimators of α and C satistfy $\hat{\alpha} - \alpha = O_p(N^{-1/2})$ and $\hat{C} - C = O_p(N^{-1/2}\log N)$ for appropriately chosen filters. Using a Taylor expansion argument we can also show that $\hat{\alpha}$ and $\hat{C}$ have asymptotic correlation 1. Theorem 1 in Section 6.7 of Stein (1999) suggests that the same rates of convergence hold for the restricted maximum likelihood estimators of C and α.

We calculate the asymptotic variance of estimators of α for each combination of the lag sets $\mathcal{K}_2 = \{1, 2\}$, $\mathcal{K}_4 = \{1, 2, 3, 4\}$ and $\mathcal{K}_8 = \{1, 2, 3, 4, 5, 6, 7, 8\}$, with filters $L^{(0)}$ through $L^{(6)}$ and $\alpha \in \{0.1, 0.7, 1, 1.3, 1.9\}$. The results are listed on the left side of Table 1. From the table we can see several interesting trends:

Table 1. Asymptotic variance and simulation results of the first study.

		Asymp. var. $\lim_{n\to\infty} \mathrm{Var}(n\hat{\alpha})$					Simulation results of $n^2\mathrm{Var}(\hat{\alpha})$				
Method: OLS		α					α				
Lag Set	Filter	0.1	0.7	1.0	1.3	1.9	0.1	0.7	1.0	1.3	1.9
$\mathcal{K}_2$	0	2.09	3.8	∞	∞	∞	2.5	3.8	6.3	15.2	38.1
	1	4.86	5.4	5.8	6.1	6.9	6.1	6.5	6.9	7.2	7.4
	2	2.63	4.4	5.3	6.1	7.6	3.1	5.1	6.1	7.1	(1)8.9
	3	10.17	9.3	9.0	8.7	8.4	(1)11.8	10.6	10.1	9.6	(1)8.6
	4	3.14	5.5	6.6	7.5	9.0	3.1	5.8	7.1	8.3	10.1
	5	6.31	6.2	6.2	6.2	6.5	8.0	7.7	7.7	7.7	7.4
	6	2.70	5.0	6.2	7.2	9.2	2.9	5.6	7.0	8.4	(1)10.8
$\mathcal{K}_4$	0	0.74	6.0	∞	∞	∞	0.79	5.3	10.6	24.4	44.5
	1	1.18	4.4	6.0	7.5	10.7	1.27	4.7	6.5	8.3	11.6
	2	0.88	4.6	6.6	8.5	12.5	1.07	5.5	8.0	10.4	(3)14.8
	3	2.65	6.0	7.6	9.0	11.6	2.87	5.9	7.6	9.2	(2)12.3
	4	1.02	5.6	8.0	10.2	14.6	1.18	6.6	9.4	12.0	(3)16.0
	5	1.58	4.5	5.9	7.2	9.6	1.89	5.1	6.7	8.4	(2)11.7
	6	0.85	5.4	7.9	10.3	15.4	1.05	6.6	9.7	12.7	(2)17.7
$\mathcal{K}_8$	0	0.66	13.2	∞	∞	∞	0.71	11.3	22.3	44.0	55.5
	1	0.72	7.7	11.5	15.1	23.1	0.82	8.7	12.8	16.6	(11)25.9
	2	0.67	8.6	13.2	17.6	27.5	1.01	11.2	16.8	22.1	(23)35.3
	3	1.31	9.3	13.4	17.0	24.3	1.45	11.1	16.3	21.2	(14)25.0
	4	0.78	10.6	16.0	21.2	32.0	1.03	11.9	17.5	22.3	(27)36.7
	5	0.87	7.6	11.0	14.1	20.7	1.22	10.2	14.9	19.3	(10)27.4
	6	0.71	10.5	16.0	21.5	33.9	1.08	13.0	19.5	25.5	(31)43.3
Method: GLS		α					α				
$\mathcal{K}_4$	0	0.74	3.6	∞	∞	∞	0.79	3.8	4.9	6.1	(47)30.4
	1	1.18	3.8	4.8	5.5	6.5	1.27	4.3	5.6	6.5	7.0
	2	0.87	3.7	4.8	5.6	6.9	1.06	4.4	5.7	6.8	8.1
	3	2.65	5.6	6.6	7.3	8.0	2.86	5.7	6.8	7.7	(1)8.6
	4	1.01	4.5	5.9	6.9	8.4	1.17	5.1	6.6	7.9	9.3
	5	1.58	4.1	5.0	5.5	6.2	1.89	4.7	5.9	6.7	7.3
	6	0.85	4.3	5.6	6.6	8.3	1.04	4.9	6.5	7.8	9.8
$\mathcal{K}_8$	0	0.57	3.5	∞	∞	∞	0.62	3.7	4.8	5.9	(49)35.6
	1	0.66	3.6	4.6	5.31	6.4	0.75	4.2	5.4	6.3	7.0
	2	0.59	3.5	4.6	5.46	6.8	0.84	4.3	5.6	6.7	8.0
	3	1.26	5.0	6.1	6.93	7.9	1.38	5.4	6.7	7.6	(1)8.5
	4	0.68	4.3	5.6	6.67	8.3	0.88	4.9	6.4	7.6	9.3
	5	0.81	3.8	4.7	5.29	6.1	1.12	4.6	5.8	6.6	7.3
	6	0.60	4.1	5.4	6.45	8.1	0.86	4.9	6.4	7.7	9.6

Notes: OLS and GLS estimator are defined by (3.14) and (3.15). The asymptotic variance is computed using (3.8) and the weight matrix in the GLS estimators are the asymptotic weight matrix defined in (3.16). The simulation results are based on 500 simulated fractional Brownian surfaces on a 90×90 grid. When using GLS estimators, we use the weight matrix defined in (4.1) evaluated at the true value of α. Lag sets are $\mathcal{K}_2 = \{1, 2\}$, $\mathcal{K}_4 = \{1, 2, 3, 4\}$ and $\mathcal{K}_8 = \{1, 2, 3, 4, 5, 6, 7, 8\}$. If any estimates fall outside (0,2], the number of times this happens is given in brackets $(\cdot)$.

1. The GLS estimator for $\mathcal{K}_4$ and $\mathcal{K}_8$ is much better than the OLS estimator with the same lag set except for $\alpha = 0.1$, for which there is little improvement. In contrast, the OLS estimator for $\mathcal{K}_2$ is only moderately worse than the GLS estimator for $\mathcal{K}_4$ and $\mathcal{K}_8$ when $\alpha > 1$.

2. For $\alpha > 1$, increasing the size of the lag set makes the OLS estimator much worse. Although the GLS estimators get better for larger lag sets, the improvement is very limited from $\mathcal{K}_4$ to $\mathcal{K}_8$.

3. Filter 0 is often best for $\alpha < 1$; Filter 2 does well for $\alpha \leq 1$ and Filter 5 for $\alpha > 1$; Filter 1 performs well for all α and Filter 3 poorly for all α.

These results are confirmed by the simulation results listed on the right side of Table 1. The details of the simulation study are explained in the next section.

4. Simulation Studies

4.1. Weight matrix

In Section 3 we used the asymptotic weight matrix and evaluated it at the true α value when using GLS estimates. In the simulation study we use the weight matrix calculated as

$$\mathbf{W}_{s,t,n} = \frac{\mathrm{Cov}\left(Z_{k_s,n}, Z_{k_t,n}\right)}{\mathrm{E}\left(Z_{k_s,n}\right)\mathrm{E}\left(Z_{k_t,n}\right)} = \frac{\sigma_{k_s,k_t,n}}{\mu_{k_s}\mu_{k_t}}. \tag{4.1}$$

This formula is obtained by using first order Taylor expansion on $\mathrm{Cov}(\log Z_{k_s,n}, \log Z_{k_t,n})$. In the first study we compared the performance of seven filters and, to reduce the computational burden, we evaluated (4.1) at the true α value. The resulting $\hat{\alpha}_G$ is then not a statistic (it depends on the unknown α); however, for large sample sizes it should have similar behavior to a GLS estimator using an estimated value for α in the weight matrix.

The second study considers only two filters and uses the OLS estimator $\hat{\alpha}_O$ to calculate the estimated weight matrix $\hat{\mathbf{W}}_{s,t,n}$ and hence corresponds to an estimator that can be used in practice. Since it is time-intensive to calculate the weight matrix, we pre-calculated and stored the weight matrix for $\alpha \in \{0.02, 0.04, \ldots, 1.98\}$ and used linear interpolation to get the estimated weight matrix for any $\alpha \in (0, 2)$. The weight matrix itself is of independent interest. Figure 4 shows the variances and correlations of the logarithm of the empirical variogram, from which the weight matrix was obtained, for Filters 0, 1 and 2 with lag set $\mathcal{K}_4$ and observations on a 362×362 grid. For Filter 0, the plot shows that for $\alpha \in (0, 1)$ the variance of the logarithm of the empirical variogram of each lag increases as α increases and as the length of the lag increases. The correlation of lags that are close in length to each other (i.e., lags 2 and 3, lags 3 and 4) generally increases monotonically as α increases while the correlation of

lags that are not close in length (i.e., lags 1 and 4, lags 1 and 3) is decreasing for $\alpha \in (0, 0.5)$ and increasing for $\alpha \in (0.5, 1)$. For $\alpha \in (1, 2)$ the variances at all lags increase dramatically because of under-differencing. In addition, the correlations are near 1, which indicates that variograms based on longer lags do not offer much new information about α.

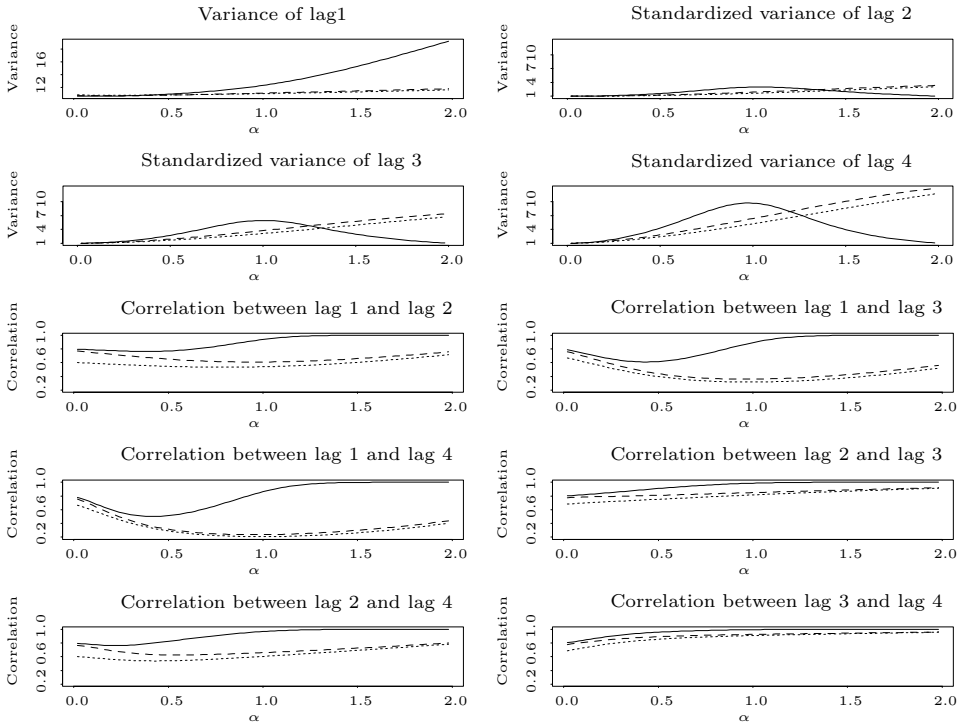

Figure 4. The variance and correlation of the logarithm of the empirical variogram of lag set $\mathcal{K}_4$. Solid line: Filter 0; dotted line: Filter 1; broken line: Filter 2. The variance of lag 1 empirical variogram is plotted in log scale. The standardized variance of lag i is the ratio between the variance of the empirical variogram on that lag and the variance of lag 1. The variogram is based on observations on a 362×362 grid.

For Filter 1, the variance of the logarithm of the empirical variogram of each lag changes much less with α than for Filter 0, and the variance itself is much smaller for larger α. The variance increases as the length of the lag increases in a similar manner as Filter 0 for $\alpha \in (0, 1)$. The correlation of lags also follows a pattern similar to Filter 0 for $\alpha \in (0, 1)$. If we take $|l_1 - l_2|/(l_1 + l_2)$ as the relative distance between lag l_1 and lag l_2 then, as the relative distance decreases, the correlation curves change from convex to concave. The plots for Filter 2 are very

similar to those for Filter 1 except that the variances and correlations are slightly higher. Intuitively, as Filter 1 is more geographically compact than Filters 2−6, we would expect the empirical variograms at different lags to be less correlated with each other and have smaller variance as well, which in turn would give a better estimator of α. This is confirmed by Figure 4 and simulation results in Section 4.2. The general patterns for Filter 1 and 2 are also observed in plots for Filters 3−6 and other lag sets, which we omit.

4.2. Simulation procedure and results

We use the embedding procedure described in Stein (2002) to get exact simulations of fractional Brownian surfaces. Surfaces with fractal dimension ranging from 2.05 to 2.95 are considered, with corresponding α ranging from 0.1 to 1.9.

In the first study, we simulated the fractional Brownian surfaces on a 90×90 regular lattice for $\alpha \in \{0.1, 0.7, 1, 1.3, 1.9\}$ and used all combinations of the seven filters and the three lag sets used in Section 3 to estimate α. For the GLS estimates, as noted in Section 4.1, we used the weight matrix defined in (4.1) and evaluated it at the true value of α. The results are listed on the right side of Table 1. From this table we can see that the simulation results match the numerically calculated asymptotic constants closely and preserve every trend we found from the numerical calculation. Both the asymptotic and simulation results show that the OLS estimators using more than two lags are much worse than using just $\mathcal{K}_2$. Thus, from now on we only consider OLS estimators using two lags, in which case OLS and GLS are identical. In addition, we only consider the zeroth order increment filter $L^{(0)}$, which works well for $\alpha < 1$, and Filter 1, which is the best overall performer among the six filters that are increments of order 1.

We also calculated the sample bias and sample MSE for each filter and lag set combination. It turns out that the sample bias was never large compared to the simulation error and the sample MSE was almost identical to the sample variance, so we do not list these results here.

In the second study we simulated the fractional Brownian surfaces on a 362×362 regular lattice for $\alpha \in \{0.1, 0.4, 0.7, 1, 1.3, 1.6, 1.9\}$ and studied the estimator based on all twelve combinations of filters $L^{(0)}$ and $L^{(1)}$ and six different lag sets. To compare asymptotic and finite-sample results, we used every other point in each coordinate on the 362×362 grid to get samples of fractional Brownian surfaces on a 181×181 grid, and applied the same procedure to get fractional Brownian surfaces on 90×90 and 45×45 grids. Using the same set of simulations for the different samples sizes made comparisons across sample sizes more precise.

The first simulation study showed that increasing the size of lag sets from four to eight only improves the GLS estimator by a small amount, so we included six relatively small lag sets, $\mathcal{K}_2^d = \{1, 2^{1/2}\}$, $\mathcal{K}_3^d = \{1, 2^{1/2}, 2\}$, $\mathcal{K}_4^d =$

$\{1, 2^{1/2}, 2, 2(2)^{1/2}\}$, $\mathcal{K}_2$, $\mathcal{K}_4$ and $\mathcal{K}_6 = \{1, 2, 3, 4, 5, 6\}$, in the second study to draw a more detailed picture about the performance of different lag sets. Unlike the first study, we evaluated the weight matrix at $\hat{\alpha}_O$ based on $\mathcal{K}_2$ instead of the true value of α. For purposes of comparison, we also calculated the asymptotic variance for the corresponding filters and lag sets. The results are listed in Table 2. We listed variances and not biases for the same reason as in the first study.

From Table 2 we can draw the following conclusions.

1. For $\alpha > 1$, Filter 1 is almost always better than Filter 0, and the difference increases as the sample size increases. These simulation results are consistent with the asymptotic result that estimators based on Filter 1 have a higher rate of convergence than those based on Filter 0 for $\alpha \geq 1$. We see that for sufficiently large sample size, it is essential to use Filter 1 rather than Filter 0.

2. For $\alpha = 1$, the estimators based on both filters have very similar performance for the sample sizes we considered, and for $\alpha < 1$ the estimator from Filter 0 is always better than the corresponding estimator using Filter 1. This is consistent with the asymptotic results in Table 2 since for $\alpha < 1$ the estimators based on Filters 0 and 1 have the same convergence rate. Part of the problem can also be attributed to the fact that while Filter 0 estimators are based on order 0 differencing of the original data, Filter 1 estimators are based on order 1 differencing of the original data, yielding a smaller number of increments for each $Y_{k,n}$. When the sample size is finite and relatively small, this loss of information from over-differencing may not be negligible and can be part of the reason that Filter 1 estimators are inferior.

3. For α near 2, increasing the lag size may not necessarily improve the performance of the GLS estimator. In fact, for $\alpha = 1.6$ or 1.9, the best lag set at sample size 362×362 is $\mathcal{K}_4^d$, at sample size 181×181 both $\mathcal{K}_3^d$ and $\mathcal{K}_4^d$ have the best performance, but at sample size 90×90 and 45×45 the best lag set is $\mathcal{K}_3^d$. We can observe a similar pattern between $\mathcal{K}_4$ and $\mathcal{K}_6$, with $\mathcal{K}_6$ performing better at sample size 362×362 and $\mathcal{K}_4$ better at sample size 45×45. Apparently one may need fairly large sample sizes for $\hat{\alpha}_O$ to yield a sufficiently accurate estimator for the weight matrix to justify using a large lag set.

4. For $\alpha \leq 1.3$, increasing the lag size does improve the performance of the estimator and the improvement is most dramatic for $\alpha = 0.1$. $\mathcal{K}_6$ almost always yields the best estimator for $\alpha \leq 1.3$, which suggests that even larger lag sets are justified in these cases.

5. For α near 2, $\mathcal{K}_4^d$ is substantially better than $\mathcal{K}_4$, and the difference is larger for smaller sample sizes. The opposite is true for α small. These results suggest that if we know α is likely to be large, we should include subcategories of 45° rotation in our estimation.

Table 2. Asymptotic variance and simulation result of the second study.

	Filter 0							Filter 1						
	α							α						
Lag Set	0.1	0.4	0.7	1.0	1.3	1.6	1.9	0.1	0.4	0.7	1.0	1.3	1.6	1.9
Asymptotic variance $\lim_{n\to\infty} \mathrm{Var}(n\hat{\alpha})$														
$\mathcal{K}_2^d$	7.59	5.9	5.2	∞	∞	∞	∞	14.86	12.6	10.8	9.5	8.5	8.1	8.2
$\mathcal{K}_3^d$	2.09	2.5	3.7	∞	∞	∞	∞	4.87	5.1	5.4	5.7	6.0	6.2	6.4
$\mathcal{K}_4^d$	1.04	2.2	3.7	∞	∞	∞	∞	1.98	3.0	4.0	4.9	5.5	6.0	6.4
$\mathcal{K}_2$	2.09	2.5	3.8	∞	∞	∞	∞	4.87	5.1	5.4	5.8	6.1	6.5	6.9
$\mathcal{K}_4$	0.74	2.2	3.6	∞	∞	∞	∞	1.19	2.6	3.8	4.8	5.5	6.0	6.5
$\mathcal{K}_6$	0.60	2.2	3.5	∞	∞	∞	∞	0.77	2.3	3.7	4.6	5.4	5.9	6.4
Simulation results of $n^2\mathrm{Var}(\hat{\alpha})$, Sample size: 362×362														
$\mathcal{K}_2^d$	6.46	5.0	4.3	7.0	44.9	325.2	378.7	13.33	11.6	10.0	8.8	7.9	8.9	8.4
$\mathcal{K}_3^d$	2.29	2.2	3.0	7.1	20.5	157.1	325.3	5.71	5.7	5.7	5.7	5.6	5.0	5.0
$\mathcal{K}_4^d$	1.10	1.9	3.0	4.7	6.1	14.4	107.6	2.30	3.2	4.1	4.8	5.4	4.8	5.0
$\mathcal{K}_2$	2.29	2.2	3.1	8.6	57.7	361.1	397.2	5.71	5.7	5.7	5.7	5.7	5.0	5.3
$\mathcal{K}_4$	0.71	1.8	2.9	4.0	5.2	8.0	114.5	1.17	2.4	3.6	4.4	5.1	5.0	5.4
$\mathcal{K}_6$	0.58	1.8	2.9	4.0	5.0	7.7	113.3	0.81	2.2	3.3	4.0	4.6	4.8	5.2
Simulation results of $n^2\mathrm{Var}(\hat{\alpha})$, Sample size: 181×181														
$\mathcal{K}_2^d$	7.23	5.8	5.0	7.5	30.4	125.6	117.3	13.21	11.7	10.8	10.5	10.5	7.0	7.0
$\mathcal{K}_3^d$	2.09	3.0	4.2	7.8	20.5	89.9	106.4	4.88	5.6	6.5	7.2	7.8	5.8	6.1
$\mathcal{K}_4^d$	1.08	2.5	4.3	6.4	8.7	9.4	29.7	1.90	3.5	5.0	6.3	7.3	5.8	6.1
$\mathcal{K}_2$	2.09	3.0	4.5	9.5	37.8	139.1	123.1	4.88	5.6	6.5	7.3	7.9	6.5	7.1
$\mathcal{K}_4$	0.87	2.7	4.5	6.1	7.5	11.3	31.5	1.29	3.1	4.7	5.9	6.7	6.2	6.8
$\mathcal{K}_6$	0.73	2.7	4.3	5.6	6.9	11.0	32.1	0.97	3.0	4.6	5.7	6.5	6.2	6.8
Simulation results of $n^2\mathrm{Var}(\hat{\alpha})$, Sample size: 90×90														
$\mathcal{K}_2^d$	7.77	6.2	5.3	6.9	18.6	48.6	37.8	15.91	13.8	11.9	10.4	9.3	8.3	8.3
$\mathcal{K}_3^d$	2.14	2.7	3.9	6.9	16.2	46.2	37.2	4.91	5.3	5.8	6.2	6.5	5.8	6.0
$\mathcal{K}_4^d$	1.10	2.3	3.8	5.6	7.2	10.1	10.2	2.20	3.2	4.3	5.2	6.0	5.9	6.0
$\mathcal{K}_2$	2.14	2.7	4.0	7.7	22.5	52.9	39.2	4.91	5.3	5.8	6.4	6.9	6.0	6.4
$\mathcal{K}_4$	0.93	2.4	3.9	5.3	6.7	10.8	11.3	1.56	2.9	4.3	5.5	6.2	5.9	6.5
$\mathcal{K}_6$	0.64	2.4	3.8	5.1	6.5	10.8	11.8	0.81	2.5	4.0	5.1	6.0	5.9	6.5
Simulation results of $n^2\mathrm{Var}(\hat{\alpha})$, Sample size: 45×45														
$\mathcal{K}_2^d$	6.98	5.1	4.2	5.1	10.6	19.6	13.3	14.09	12.3	11.0	10.2	10.1	8.2	8.3
$\mathcal{K}_3^d$	1.52	1.7	2.4	4.6	10.3	20.2	13.4	3.63	4.0	4.7	5.4	6.0	4.9	5.1
$\mathcal{K}_4^d$	0.77	1.4	2.4	3.9	5.6	7.7	4.1	1.75	2.8	4.0	5.0	5.7	5.2	5.3
$\mathcal{K}_2$	1.52	1.7	2.5	5.0	11.9	20.9	13.4	3.62	4.0	4.7	5.4	5.9	5.2	5.6
$\mathcal{K}_4$	0.59	1.4	2.5	3.6	5.0	9.4	5.2	1.10	2.0	3.0	4.0	4.8	5.6	6.0
$\mathcal{K}_6$	0.43	1.4	2.4	3.4	4.8	9.3	5.6	0.64	1.8	2.8	3.7	4.6	5.6	6.1

Notes: $\hat{\alpha}$s are GLS estimators defined by (3.15). For $\mathcal{K}_2^d$ and $\mathcal{K}_2$ the GLS and OLS estimators are identical. The asymptotic variances are computed using (3.8) and the weight matrices in the GLS estimators are the asymptotic weight matrices defined in (3.16). The simulation results are based on 100 simulated fractional Brownian surfaces on a 362×362 grid. For the GLS estimators, we use the weight matrix in (4.1) evaluated at $\alpha = \hat{\alpha}_O$. Lag sets are $\mathcal{K}_2^d = \{1, \sqrt{2}\}$, $\mathcal{K}_3^d = \{1, \sqrt{2}, 2\}$, $\mathcal{K}_4^d = \{1, \sqrt{2}, 2, 2\sqrt{2}\}$, $\mathcal{K}_2 = \{1, 2\}$, $\mathcal{K}_4 = \{1, 2, 3, 4\}$ and $\mathcal{K}_6 = \{1, 2, 3, 4, 5, 6\}$. For sample size larger than 45×45 almost all estimators are within (0,2]. The number of estimates outside (0,2] can be found in Zhu and Stein (2001).

6. The asymptotic results give less accurate approximations than in the first study for α near 2. This is expected because we estimated the weight matrix in the GLS estimator instead of using the true value of α. The difference between the asymptotic results and the simulation results for the smaller sample sizes can be substantial, which shows the limitations of asymptotic theory and the necessity of doing simulation studies for this problem.

Table 2 can be used to choose the best lag set for the GLS in practice if the sample size is similar to the one considered there. Otherwise, an iterative procedure can be used as follows. First use OLS to get an estimate $\hat{\alpha}_{(1)}$ based on $\mathcal{K}_2$. Then simulate the random field at $\hat{\alpha}_{(1)}$ and use all lag sets being considered to estimate α. The lag set that has the best performance for the simulated data is then used to get an estimate $\hat{\alpha}_{(2)}$. Repeat the above procedure until the best lag set for the data is found. Since the number of lag sets one needs to consider is usually small, one iteration would be enough to find the best lag set on most occasions. In our simulation study, it took 642 seconds to simulate 100 fractional Brownian surfaces on a 362×362 grid for $\alpha \leq 1.5$, and 2717 seconds for $\alpha > 1.5$ on a PIII 866. Computing the weight matrix for different lag sets took 165 to 574 seconds. The time it took to compute the estimator after the weight matrix was obtained was negligible. So the whole procedure can be implemented in a few hours for a data set with 131,044 observations on a 362×362 regular grid.

4.3. Comparison to Likelihood Method

An alternative approach to the estimation of α is to maximize the likelihood function of the contrasts, which is commonly known as restricted maximum likelihood (REML) estimation. See, for example, Stein (1999). In the settings considered here, the contrasts are all zeroth order differences of the observations. Theorem 1 in Section 6.7 of Stein (1999) suggests that the REML estimator of α has the same rate of convergence as the OLS/GLS estimators defined in this paper. The REML estimator should be more efficient than the OLS/GLS estimator when the sample size is large. We calculated the REML estimator of α for simulated data on a 45×45 grid, and the result is in Table 3. By comparing Table 2 with Table 3 we can see that the REML estimators outperform the GLS estimators by a large margin when $\alpha \geq 1.6$, but the difference is fairly small for smaller α. However, it is computationally difficult to compute the exact REML estimator when the sample size is moderately large, as the computation time is $O(N^3)$ and the memory required is $O(N^2)$, N being the sample size. On PIII 866, it took about 98 seconds to evaluate the likelihood once for observations on a 45×45 grid. It would take about 1.74 hours for observations on a 90×90 grid and 307 days for observations on a 362×362 grid. One could consider using various approximations to the restricted likelihood function such as a spectral method

(Stein 1995), but these are beyond the scope of this paper. On the other hand, the GLS can offer an estimator with the same convergence rate as the REML estimator in much shorter time, and thus can be recommended in practice for large data sets.

Table 3. Simulation results for REML estimators.

α	0.1	0.4	0.7	1.0	1.3	1.6	1.9
nBias	-0.09	-0.2	-0.3	-0.3	-0.3	0.1	-0.2
Var($n\hat{\alpha}$)	0.43	1.7	2.7	3.5	4.0	4.7	3.6
n^2MSE	0.44	1.8	2.8	3.6	4.1	4.7	3.7

Notes: The simulation results are based on 100 simulated fractional Brownian surfaces on a 45×45 grid.

4.4. Comparison to Chan and Wood (2000)

The simulation studies done in Chan and Wood (2000) showed a sharp increase of bias as α increases for all the estimators they considered except for the one based on square-increment, and they attributed this bias to the use of an approximate procedure when simulating smoother processes (i.e., processes with α close to 2). In our simulation studies we employed the exact simulation procedure by Stein (2002) and we see no evidence of this bias problem, confirming that the biases in Wood and Chan (2000) are due to the simulation method. In fact, in all our simulation studies the squared biases are much less than the variances, and in most cases are negligible.

The square-increment estimators in Chan and Wood (2000) correspond to the estimators based on Filter 3 in our paper, and the other four estimators they considered each corresponds to an estimator based on one of the four subcategories of Filter 1. For the purpose of comparison we list the asymptotic variance and simulation results for the Filter 1 estimators, the Square/Filter 3 estimators, and the Horizontal estimators in Table 4. Using vertical increments gives the same statistical properties as using horizontal increments, so we only list one of them.

Table 4 shows that if we only compare the Square estimator with the Horizontal estimator, the Square estimator is better except for $\alpha = 0.1$. These results are consistent with Chan and Wood (2000). However, the Horizontal estimators are worse than the Filter 1 estimators by a substantial margin. This is because the Horizontal estimators are based on single subcategory of Filter 1 without averaging the same filter over different directions. Let $\hat{\alpha}$, $\hat{\alpha}_H$ and $\hat{\alpha}_V$ be the estimators of α defined by (3.5) and based on Filter 1, horizontal increment and vertical increment respectively, then it can be shown that

Table 4. Comparing Filter 1 estimator with estimators defined in Chan and Wood (2000): Asymptotic variance and simulation results.

	Asymp. var. $\lim_{n\to\infty} \mathrm{Var}(n\hat{\alpha})$					Simulation results of $n^2\mathrm{Var}(\hat{\alpha})$				
	α					α				
Type of estimator	0.1	0.7	1.0	1.3	1.9	0.1	0.7	1.0	1.3	1.9
Filter 1	4.9	5.4	5.8	6.1	6.9	6.1	6.5	6.9	7.2	7.4
Square/Filter 3	10.2	9.3	9.0	8.7	8.4	(1)11.8	10.6	10.1	9.6	(1)8.6
Horizontal	9.5	9.6	10.1	10.6	12.2	(1)11.2	10.8	11.3	12.2	(1)12.8

Notes: All estimators are based on the lag set $\mathcal{K}_2 = \{1, 2\}$. The asymptotic variance is computed using (3.8). The simulation results are based on 500 simulated fractional Brownian surfaces on a 90×90 grid. If any estimates fall outside (0,2], the number of times this happens is given in brackets (.).

$$N^{1/2}\{\hat{\alpha} - \alpha\} = \frac{1}{2}N^{1/2}\{\hat{\alpha}_H - \alpha\} + \frac{1}{2}N^{1/2}\{\hat{\alpha}_V - \alpha\} + O_p(N^{-1/2}).$$

Since $N^{1/2}\{\hat{\alpha}_H - \alpha\}$ and $N^{1/2}\{\hat{\alpha}_V - \alpha\}$ have the same distribution, the asymptotic variance of Filter 1 estimators will be strictly smaller than that of horizontal/vertical estimators as long as the asymptotic correlations between horizontal and vertical estimators are less than 1. In our simulation study the variance of those estimators are 50% to 85% larger than that of estimators based on Filter 1, and the calculation of asymptotic variance confirms those results. Our study also found that estimators based on Filter 1 are better than estimates based on Filter 3 in both the asymptotic variance and the simulation study, so there is no evidence that Filter 3 should ever be preferred. Table 4 only includes the results of estimators based on lag set $\mathcal{K}_2$. Similar results hold for larger lag sets, which we omit. Although our theoretical results are stated just for fractional Brownian surfaces, they can be extended to the random fields Chan and Wood (2000) considered using essentially the same arguments. Chan and Wood (2000) considered a particular non-Gaussian case, the stationary χ_1^2 field, and proved that their estimator of α has the same rate of convergence as in the Gaussian case. Similar results also hold for the estimators we defined.

In this section we apply our method to the stationary Gaussian field with power exponential covariance function and the stationary χ_1^2 field and compare our estimators with Chan and Wood (2000). By adding a quadratic term to the covariance function as suggested in Stein (2002) we can simulate the increments of those random fields exactly, and simulation results are given in Table 5. From the table we can see that the Filter 1 estimator outperforms the other two for both Gaussian and χ_1^2 fields, and the estimator is generally less efficient for χ_1^2 fields than for Gaussian fields.

Table 5. Simulation results for stationary Gaussian process and stationary χ^2_1 processes.

	Filter 1			Square/Filter 3			Horizontal		
α	nBias	Var($n\hat{\alpha}$)	n^2MSE	nBias	Var($n\hat{\alpha}$)	n^2MSE	nBias	Var($n\hat{\alpha}$)	n^2MSE
Stationary Gaussian Random Field									
0.1	-2.5	4.9	(1)11.0	-2.5	9.6	(12)15.7	-2.6	9.2	(11)16.2
0.7	-0.8	5.4	6.0	-0.7	8.9	9.4	-0.9	8.8	9.6
1.0	0.0	5.7	5.7	-0.0	8.6	8.6	-0.1	9.2	9.3
1.3	0.2	6.1	6.1	0.1	8.3	8.3	0.1	9.9	9.9
1.9	0.6	7.2	7.6	0.4	8.8	(3)8.9	0.5	11.6	(2)11.9
Stationary χ^2_1 Random Field									
0.1	-6.4	7.0	(82)47.5	-6.4	13.2	(117)54.2	-6.5	13.4	(113)55.3
0.7	-2.7	15.1	22.4	-2.6	22.1	28.7	-2.9	24.1	32.6
1.0	-0.6	14.4	14.8	-0.7	20.0	20.5	-0.8	23.5	24.1
1.3	0.1	14.4	14.4	-0.2	18.9	18.9	-0.2	23.3	23.3
1.9	1.3	16.0	(16)17.7	0.8	18.2	(15)18.8	1.4	25.5	(36)27.5

Notes: All estimators are based on the lag set $\mathcal{K}_2 = \{1, 2\}$. The simulation results are based on 500 simulated Gaussian random field with power exponential covariance function on a 90×90 grid. If any estimates fall outside (0,2], the number of times this happens is given in brackets (.).

4.5. Estimating C

Figure 5 shows examples of the relationship between $\hat{\alpha}$ and $\hat{C}$ for Gaussian random fields. The simulations are on a 90×90 grid for $\alpha = 1.6$ or 1.9 and $C = 1$, and the estimators are based on Filter 1 and $\mathcal{K}_4$, using GLS. From Figure 5 we can see that $\hat{\alpha}$ and $\hat{C}$ have high correlation, and the variation of $\hat{C}$ is much larger than that of $\hat{\alpha}$, both consistent with the asymptotic theory.

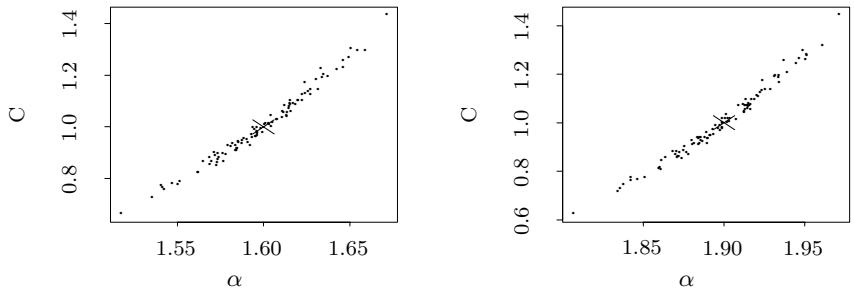

Figure 5. $\hat{\alpha}$ vs $\hat{C}$ based on 100 simulations on a 90×90 grid, with $(\alpha, C) = (1.6, 1)$ on the left and $(\alpha, C) = (1.9, 1)$ on the right (indicated by $\times$). The estimators are based on Filter 1 and lag set $\mathcal{K}_4$.

Acknowledgement

This research was supported in part by National Science Foundation grant 99-71127. We would like to thank the reviewers for helpful comments that led to a number of improvements in the substance and presentation of this work.

References

Chan, G. and Wood, A. T. A. (2000). Increment-based estimators of fractal dimension for two-dimensional surface data. *Statist. Sinica* **10**, 343-376.

Constantine, A. G. and Hall, P. (1994). Characterizing surface smoothness via estimation of effective fractal dimension. *J. Roy. Statist. Soc. Ser. B* **56**, 97-113.

Cressie, N. (1993). *Statistics for Spatial Data* (revised ed). John Wiley, New York.

Davies, S. and Hall, P. (1999). Fractal analysis of surface roughness by using spatial data. *J. Roy. Statist. Soc. Ser. B* **61**, 3-37.

Kent, J. T. and Wood A. T. A. (1997). Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J. Roy. Statist. Soc. Ser. B* **59**, 679-699.

Peitgen, H. O. and Saupe, D. (1988). *The Science of Fractal Images*. Springer-Verlag, New York.

Stein, M. L. (1995). Fixed-domain asymptotics for spatial periodograms. *J. Amer. Statist. Assoc.* **90**, 1277-1288.

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

Stein, M. L. (2002). Fast and exact simulation of fractional Brownian surfaces. *J. Comput. Graphical Statist.* In press.

Taylor, C. C. and Taylor, S. J. (1991). Estimating the dimension of a fractal. *J. Roy. Statist. Soc. Ser. B* **53**, 353-364.

Wilson, T. H. (2000). Some distinctions between self-similar and self-affine estimates of fractal dimension with case history. *Math. Geology* **32**, 319-335.

Zhu, Z. and Stein, M. L. (2001). Parameter estimation for fractional Brownian surfaces. University of Chicago, Department of Statistics technical report 501.

Department of Statistics, the University of Chicago, Chicago, IL 60637, U.S.A.

E-mail: zhu@galton.uchicago.edu

Department of Statistics, the University of Chicago, Chicago, IL 60637, U.S.A.

E-mail: stein@galton.uchicago.edu

(Received October 2000; accepted January 2002)