The Modeling and Estimation of Statistically Self-Similar Processes in a Multiresolution Framework

Michael M. Daniel, *Student Member, IEEE*, and Alan S. Willsky, *Fellow, IEEE*

***Abstract*—Statistically self-similar (SSS) processes can be used to describe a variety of physical phenomena, yet modeling these phenomena has proved challenging. Most of the proposed models for SSS and approximately SSS processes have power spectra that behave as $1/f^{\gamma}$, such as fractional Brownian motion (fBm), fractionally differenced noise, and wavelet-based syntheses. The most flexible framework is perhaps that based on wavelets, which provides a powerful tool for the synthesis and estimation of $1/f$ processes, but assumes a particular distribution of the measurements. An alternative framework is the class of multiresolution processes proposed by Chou *et al.* [1994], which has already been shown to be useful for the identification of the parameters of fBm. These multiresolution processes are defined by an autoregression in scale that makes them naturally suited to the representation of SSS (and approximately SSS) phenomena, both stationary and nonstationary. Also, this multiresolution framework is accompanied by an efficient estimator, likelihood calculator, and conditional simulator that make no assumptions about the distribution of the measurements. In this paper, we show how to use the multiscale framework to represent SSS (or approximately SSS) processes such as fBm and fractionally differenced Gaussian noise. The multiscale models are realized by using canonical correlations (CC) and by exploiting the self-similarity and possible stationarity or stationary increments of the desired process. A number of examples are provided to demonstrate the utility of the multiscale framework in simulating and estimating SSS processes.**

***Index Terms*—Canonical correlations, fractional Brownian motion, multiscale, self-similarity.**

I. Introduction

A wide variety of physical phenomena are described by random processes that are statistically self-similar (SSS). A common characteristic of such processes is that the power spectral densities behave as $1/f^{\gamma}$, particularly in the range $0 < \gamma < 2$. Another common characteristic is long range dependence, in either the process or its increments. The long-range dependence is usually manifested by a covariance function $r(\tau)$ that decreases hyperbolically, i.e., $r(\tau) \sim \tau^{-\alpha}$ for some $\alpha > 0$, which is closely related to having a spectrum described by a power law. Examples of physical phenomena well described by $1/f$ processes are average temperature distributions [1], [2], annual flow rates in rivers [1], the noise in vacuum tubes and electrical components [2], traffic in communications networks [3], and biological and economic time series [4], [5]. Processes with $1/f$-like power spectra are also used to generate images that model real-world objects like clouds and mountain ranges [6], [7].

A number of models have been proposed that possess long-range dependence and $1/f$ spectra.[1] One is the class of fractional Brownian motions [4], which are zero-mean statistically self-similar Gaussian random processes with stationary increments. Another class of models is given by the fractional differencing and integration of discrete-time white noise [1], [8], [9], which yields discrete-time processes that are closely related to samples of fractional Brownian motion. A third class of models, motivated by the voltage response of a transmission line to a white-noise current source, was proposed by Keshner [2]. A more recently introduced class of $1/f$ models, which can approximate the statistics of the three aforementioned models, is based on the observation that the wavelet transform makes stationary and approximately whitens any $1/f$ process [5], [10], [11]. This framework was recently extended in [12] with the introduction of bi-orthogonal wavelets as a method for more accurately synthesizing fractional Brownian motion. The relative advantages of these models depend on the particular application and whether one is interested in synthesizing, estimating (smoothing and interpolating), or determining the parameters of $1/f$ processes. Wornell showed in [5] how the wavelet-based models can be used to efficiently synthesize, estimate, and determine model parameters for $1/f$ time series. The efficiency follows from efficient implementations of the discrete wavelet transform. However, the estimation algorithms based on this transform require that all of the measurements are equally spaced and that the measurement noise have constant variance. Also, for regular wavelets, one must account for boundary effects at the edges of the interval of interest.

An alternative framework for the modeling and processing of $1/f$ signals is the multiresolution stochastic processes proposed in [13]. These processes are indexed by the nodes of trees organized into scales, where the leaf-nodes of the tree represent the finest scale features of interest. Much like the wavelet models, these models are well-suited to the statistical self-similarity and possible nonstationarity of the $1/f$ processes. Also, the tree structure allows the multiresolution

Manuscript received December 10, 1997; revised October 17, 1998. This work was supported by the Office of Naval Research under Grant N00014-91-J-1004, by the Air Force Office of Sponsored Research under Grant F49620-95-1-0083, and by Boston University under subcontract GC123919NGD.

The authors are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Publisher Item Identifier S 0018-9448(99)02471-2.

[1] This paper considers only the first and second moments of a stochastic process, i.e., the process is either Gaussian or the higher order moments are considered to be insignificant.

models to compactly account for the long-range dependencies of $1/f$ processes. More importantly, these multiresolution processes are accompanied by very efficient estimation, simulation (conditional [14] and unconditional), and likelihood calculation [15] algorithms. These algorithms allow for measurements that are at different resolutions are irregularly spaced or are corrupted by measurement noise with nonconstant variance. This flexibility is necessary for many applications, like remote-sensing and geophysical inverse problems [16].

In this paper, we demonstrate that the multiresolution stochastic processes of [13] can accurately model a range of SSS and $1/f$ processes. In particular, we show that samples of fBm, discrete fractional Gaussian noise (DFGN), and fractionally differenced Gaussian noise (FDGN) can be represented to arbitrary accuracy by the multiresolution models. Multiresolution models representing fBm that are based on random midpoint displacement and the Haar wavelet basis were described in [17], [18], but these models are limited to low-order approximations with no clear method for refining the approximations. This paper develops a more general class of models that provides an optimal tradeoff between the model order and the accuracy of the representation. The realization algorithm for these models is based on canonical correlations (CC) [19], [20] and it exploits the statistical self-similarity and stationarity (or stationary increments property) of the $1/f$ processes proposed by Mandelbrot and Hosking. Representing these processes within the multiresolution framework not only allows for accurate statistics but also for efficient and flexible processing. In Section II, some relevant properties of fBm, discrete fractional Gaussian noise (DFGN) and fractionally differenced Gaussian noises (FDGN) are described. In Section III, the class of multiresolution models is defined. Next, in Sections IV and V, an algorithm based on CC is described that provides accurate multiresolution representations of fBm, DFGN, and FDGN. The algorithm is justified with examples in Section VI. Conclusions and outstanding problems are provided in Section VII.

II. Fractional Brownian Motion and Fractional Gaussian Noises

A fractional Brownian motion (fBm) is a Gaussian process with zero mean and covariance [4]

$$E[(t)x(s)] = \frac{\sigma^2}{2}\,[|t|^{2H} + |s|^{2H} - |t-s|^{2H}], \qquad 0 < H < 1. \tag{1}$$

Such a process is completely characterized by $E[x(1)^2] = \sigma^2$ and the Hurst exponent H. The covariance and variance functions are plotted in Fig. 1 for $\sigma^2 = 1$ and three values of H, where $H = 1/2$ corresponds to Brownian motion. Fractional Brownian motion is SSS in the sense that

$$x(at) \stackrel{\mathcal{P}}{=} a^H x(t), \qquad a > 0 \tag{2}$$

where $\stackrel{\mathcal{P}}{=}$ denotes equality in (finite-dimensional) distribution. While fBm is nonstationary, its power spectral density is well defined over any finite bandwidth observation window as [5]

$$S_x(f) = \frac{c}{f^{2H+1}}, \qquad f_1 < f \le f_2 \tag{3}$$

for some constant c and any two positive frequencies $0 < f_1 < f_2$. The nonstationarity of the process is evidenced near $f = 0$, where (3) implies infinite power in the low-frequency components of $x(t)$.

The nonstationary covariance function in (1) masks a more elegant definition of fBm in terms of its increments process. A process $x(t)$ is an fBm with Hurst exponent $0 < H < 1$ if and only if: 1) $x(t)$ is Gaussian and has zero mean; 2) $x(t)$ has stationary increments meaning that

$$E[(x(t+s) - x(t))^2] = g(s),$$

where $g(s)$ is called the *structure function*; and 3) $x(t)$ is self-similar, which implies $g(as) = a^{2H} g(s)$. The self-similarity of the process implies that the structure function $g(s)$ must have the form $g(s) = \sigma^2 |s|^{2H}$. The covariance function in (1) follows from this structure function and the stationary increments property.

The long-range dependence of the increments of fBm for $H > 1/2$ is manifested in the correlations among these increments. Consider the increments process

$$x_\Delta(t) \stackrel{\Delta}{=} x(t+\Delta t) - x(t).$$

For any $\Delta t > 0$, this process is stationary and its covariance function is [21]

$$r_{x_\Delta}(\tau) = \frac{\sigma^2(\Delta t)^{2H}}{2} \cdot \left(\left| \frac{\tau}{\Delta t} + 1 \right|^{2H} + \left| \frac{\tau}{\Delta t} - 1 \right|^{2H} - 2 \left| \frac{\tau}{\Delta t} \right|^{2H} \right) \tag{4a}$$

$$\approx \sigma^2 (\Delta t)^{2H}\, H(2H-1) |\tau|^{2H-2} \tag{4b}$$

where the approximation follows from a Taylor Series expansion for $|\tau| \gg \Delta t$. This approximation shows that the correlation between increments decays polynomially with distance and is positive for $1/2 < H < 1$ and negative for $0 < H < 1/2$. Also, the power spectral density of $x_\Delta(t)$ is proportional to $1/f^{2H-1}$ for $f \ll \Delta t^{-1}$ [21]. This spectrum should not be surprising, given (3) and that $x_\Delta(t)/\Delta t$ is an approximate derivative of fBm. The discrete-time process $x_\Delta[n] \stackrel{\Delta}{=} x_\Delta(n\Delta t)$ is commonly referred to as DFGN. The polynomial decay in the covariance of DFGN suggests statistical self-similarity. In fact, for $a_m[n] \stackrel{\Delta}{=} a_{m-1}[2n+1] + a_{m-1}[2n]$ with $a_0[n] = x_\Delta[n]$, the covariance of a_m is [22]

$$r_{a_m}[k] = 2^{2Hm} r_{x_\Delta}[k]. \tag{5}$$

This self-similarity will be invoked when applying the realization algorithm of Section IV.

Another class of discrete-time processes that possess long-range dependence is the FDGN's posed by Hosking [8] and by Granger and Joyeux [9]. A fractionally differenced noise is defined as the response of a linear time-invarient (LTI) system with system function $(1-z^{-1})^{-h}$ to a white noise input $w[n]$. For $-1/2 < h < 1/2$, the system function can be expressed as an infinite series that converges for $|z^{-1}| \le 1$, leading to [1], [8]

$$w_h[n] = \sum_{k=0}^{\infty} c[k] w[n-k] \tag{6a}$$

$$c[k] = \binom{-h}{k} (-1)^k. \tag{6b}$$

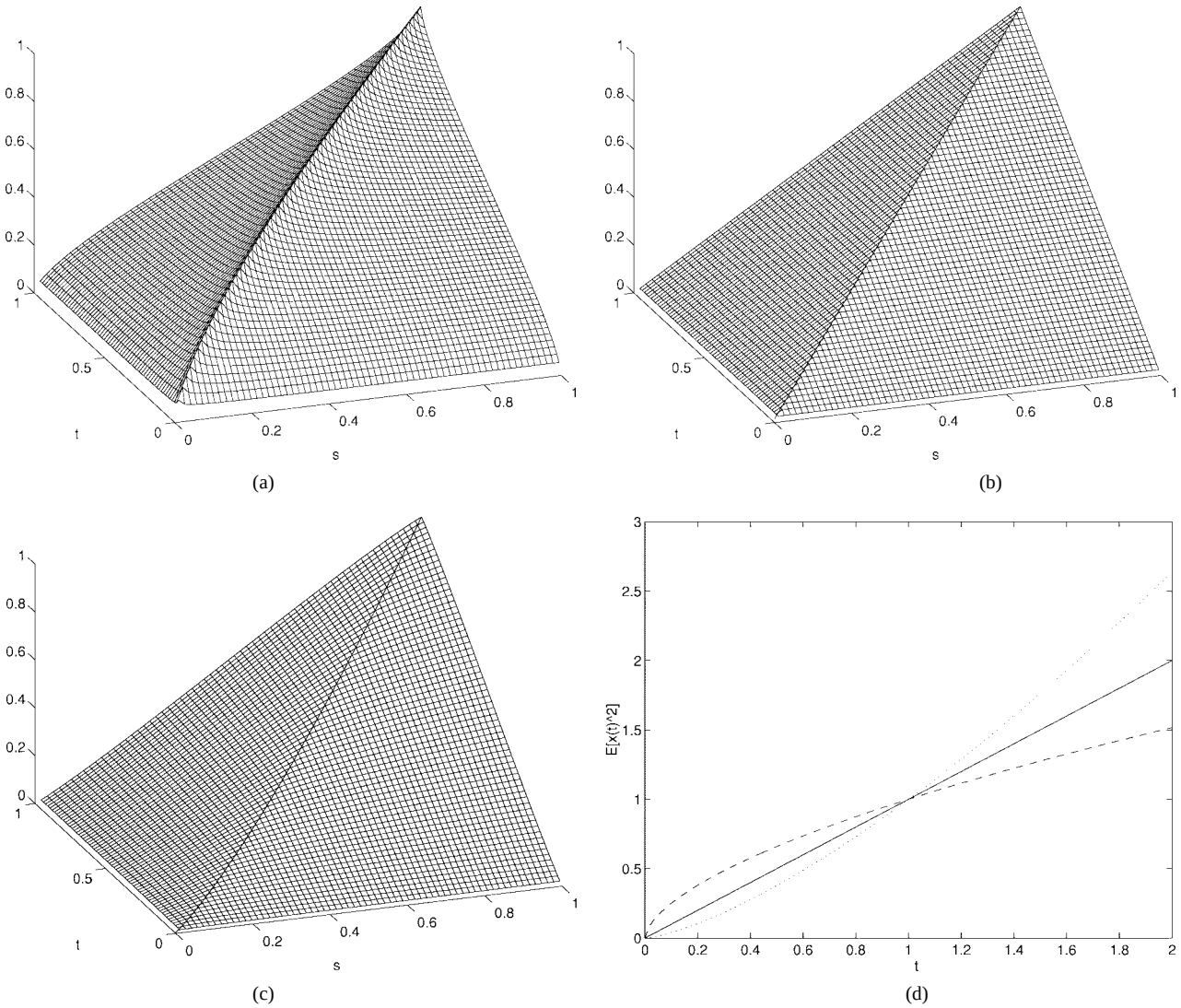


Fig. 1. The covariance functions of fBm for $\sigma^2 = 1$ and (a) $H = 0.3$, (b) $H = 0.5$, and (c) $H = 0.7$. (d) The variance function of fBm for $H = 0.3$ (dashed), $H = 0.5$ (solid), and $H = 0.7$ (dotted).

Since $c[n]$ is absolutely summable for $-1/2 < h < 1/2$, the power-spectral density of the fractional noise $w_h[n]$ follows as

$$S_h(\omega) = \frac{2^{-2h}\sigma_w^2}{\sin^{2h}(\omega/2)} \tag{7}$$

where σ_w^2 is the variance of $w[n]$ and ω is discrete-time frequency. The covariance function is defined by the recursion $r_h[m+1] = r_h[m](m+h)/(m+1-h)$ for $m \geq 0$, where $r_h[0]$ is controlled by σ_w^2. For small $\omega, S_h(\omega) \sim 1/\omega^{2h}$. Also, $r_h[m] \sim m^{2h-1}$ for large m [8], [23]. Thus, at least at low frequencies, $w_h[n]$ behaves identically to DFGN with $H = h + 1/2$. A similar analogy between samples of fBm and FDGN can be made if we allow $w_h[n]$ to be summed, i.e., passed through an LTI filter with system function $H(z) = (1 - z^{-1})^{-1}$.

III. Multiscale Stochastic Processes

The multiscale random processes introduced in [13] are indexed by the nodes of trees organized into scales. The coarsest scale is indexed by the root node, while the finest scale is indexed by the set of leaf nodes. For example, the multiscale process indexed by the binary tree illustrated in Fig. 2(a) consists of a random vector $z(s)$ for each node s on the tree. The *scale* of node s, which we denote by $m(s)$, is the distance between node s and the root node of the tree. Define $\overline{\gamma}$ to be the upward (in scale) shift operator so that $s\overline{\gamma}$ is the parent of node s, as illustrated in Fig. 2(b). The nodes $s\alpha_i, i = 1, \cdots, q$, are defined to be the children of s. For the rest of this paper, we will consider only binary trees, i.e., $q = 2$. The multiscale processes satisfy the following autoregression *in scale*[2]

$$z(s) = A_s z(s\overline{\gamma}) + w(s) \tag{8a}$$

$$w(s) \sim \mathcal{N}(0, Q_s). \tag{8b}$$

The autoregression is initialized at the root node $s = 0$ by

$$z(0) \sim \mathcal{N}(0, P_0). \tag{9}$$

[2]The notation $x \sim \mathcal{N}(m_x, P_x)$ denotes that x is a normal random vector with mean m_x and covariance P_x.

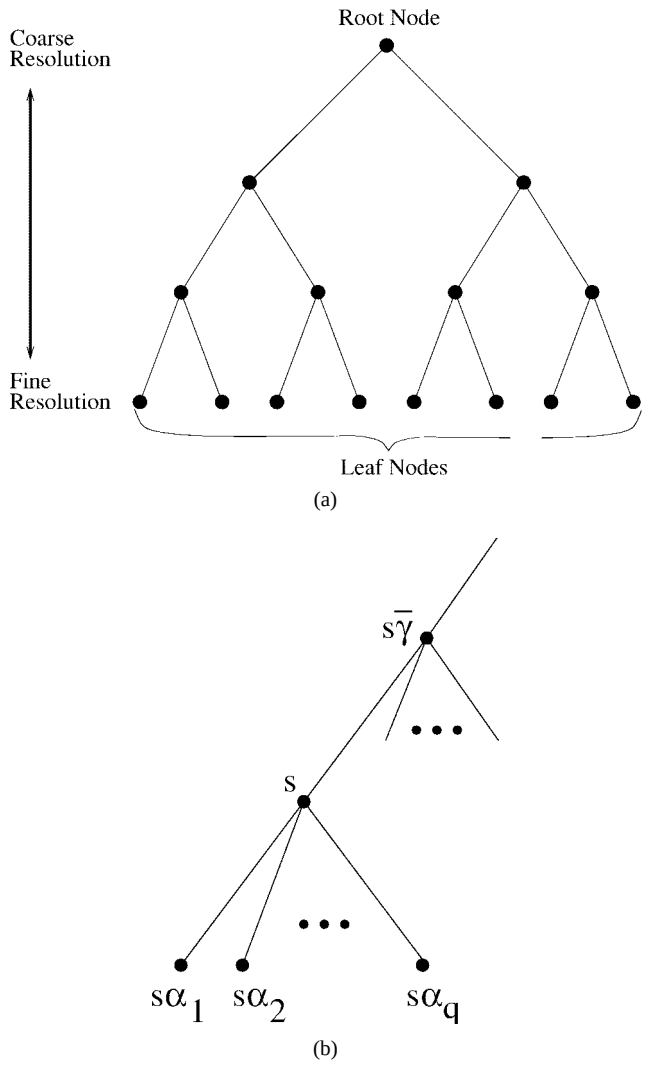


Fig. 2. (a) A binary tree used to index a random process at multiple resolutions. (b) The local labeling of the $q+1$ nodes connected to node s.

Since $z(0)$ and $w(s)$ are zero mean, every process value $z(s)$ will be a zero-mean random vector.

The process noise $w(s)$ is assumed to be a white-noise process uncorrelated across scale and space and also uncorrelated with the root-node variable, i.e., $E[w(s)\ z(0)^T] = 0$. The whiteness of the process noise implies that a multiscale tree model is characterized completely by P_0—the root-node covariance—and the autoregression parameters A_s and Q_s for all nodes $s \neq 0$. More importantly, the whiteness of the process noise implies a Markov property similar to the Markov property for one-dimensional (1-D) autoregressive processes driven by white noise [13], namely, every node $s \neq 0$ partitions the tree into $q+1$ sets of nodes (three sets for binary trees) while the root node partitions the tree into just q sets (two sets for binary trees). The Markov property is that conditioned on $z(s)$, the $q+1$ sets of random vectors partitioned by node s are mutually uncorrelated.

This Markov property leads to an efficient algorithm for the estimation of the tree process from measurements, each of which is a noise-corrupted observation of $z(\cdot)$ at some node of the tree

$$y(s) = C_s z(s) + v(s) \qquad v(s) \sim \mathcal{N}(0, R_s) \tag{10}$$

where $v(\cdot)$ is white and uncorrelated with $z(\cdot)$ at all nodes on the tree. The estimation algorithm, which is discussed in detail in [13], [24], is a generalization of the Rauch–Tung–Striebel (RTS) smoother [25]. The algorithm divides into two steps. The first step is a fine-to-coarse scale processing, where the optimal estimate of each state $z(s)$ is computed based on all measurements at node s and its descendents. Call this estimate $z(s|s)$. Analogous to the Kalman filter, $z(s|s)$ is a linear function of $y(s)$ and the estimates $z(s\alpha_i|s\alpha_i)$ at each of its descendent nodes. The Markov property essentially allows the estimates $z(s\alpha_i|s\alpha_i)$ to be computed independently and then incorporated into the computation of $z(s|s)$. The second step is a coarse-to-fine processing, analogous to the backward sweep of the RTS smoother. The output of the coarse-to-fine sweep is the optimal estimate $\hat{z}(s)$ based on all measurements and the estimation error covariance $E[(z(s)-\hat{z}(s))\,(z(s)-\hat{z}(s))^T]$ at every node on the tree. The estimation algorithm requires $\mathcal{O}(Nd^3)$ computations for a tree that has N nodes at the finest scale and constant variable dimension d.

A. Internal Multiscale Models

Recall that the finest-scale process of a multiscale model is the process indexed by the leaf nodes of the tree. An internal multiscale model is one for which the variable at each node s is a linear function of the finest-scale process that descends from s [20].[3] If f_s is defined to be the vector containing the finest-scale process descending from node s, then each variable of an internal multiscale model can be expressed as

$$z(s) = W_s f_s. \tag{11}$$

The matrices W_s are called *internal matrices* and $W_s f_s$ is the *internal variable* at node s.

Internal models are of interest because the tree parameters P_0, A_s, and Q_s can be derived directly from the internal matrices and $E[f_0 f_0^T]$. Specifically,

$$P_0 = W_0 E[f_0 f_0^T] W_0^T. \tag{12}$$

The parameters A_s and Q_s can be computed by noting that (8a) is just the optimal prediction of $z(s)$ based upon $z(s\overline{\gamma})$, plus the associated prediction error, i.e.,

$$z(s) = E[z(s)|z(s\overline{\gamma})] + w(s). \tag{13}$$

Using standard estimation equations, the model parameters follow as

$$A_s = W_s E[f_s f_{s\overline{\gamma}}^T] W_{s\overline{\gamma}}^T (W_{s\overline{\gamma}} E[f_{s\overline{\gamma}} f_{s\overline{\gamma}}^T] W_{s\overline{\gamma}}^T)^{-1} \tag{14a}$$

$$Q_s = W_s E[f_s f_s^T] W_s^T - A_s W_{s\overline{\gamma}} E[f_{s\overline{\gamma}} f_s^T] W_s^T. \tag{14b}$$

[3]The states of internal models actually can be functions of the entire finest-scale process, but this generality is not needed for the models of this paper.

B. Choosing the Internal Matrices

In this paper, the finest-scale process will represent a finite interval of fBm, DFGN, or FDGN. The only problem is how to determine the internal matrices. The internal variables must satisfy the Markov property, but an equally important consideration is minimizing the complexity of the resulting model, i.e., the state dimensions should be minimized so as to reduce the complexity of the multiscale estimation and likelihood calculation algorithms. In some cases, such internal matrices can be found without any computation, e.g., for Markov processes or Markov random fields [26]. In other cases, internal matrices that lead to approximate multiscale models are easily found. For example, the internal variables for the multiscale models in [17], [18] that approximately represent fBm are derived from the random midpoint displacement and wavelet synthesis algorithms. In cases where intelligent approximations are not easily found or when one desires more accurate representations that make optimal use of state dimension, a more systematic approach is required. One approach is to make use of CC [20], extending the results of 1-D realization theory [19] to multiscale trees. This approach can be computationally overwhelming, but, as shown in this paper, when adapted to make use of SSS and stationarity (or stationary increments), it is both efficient and accurate. In the remainder of this section, we briefly summarize the application of CC to multiscale realization.

If every variable satisfies (11), the Markov property is satisfied at each node s if and only if $W_s f_s$ conditionally decorrelates the three subsets of finest-scale variables partitioned by node s. In other words, if we define f_{s^c} to contain the elements of f_0 not in f_s, then $f_{s\alpha_1}, f_{s\alpha_2}$ and f_{s^c} must be uncorrelated after conditioning on $W_s f_s$. For approximate models, the problem is to minimize the residual correlation among these random variables while restricting the dimension of $z(s)$. For any integer $d > 0$, CC provides the d-dimensional linear function $W_s f_s$ that minimizes the correlation among the vectors $f_{s\alpha_1}, f_{s\alpha_2}$, and f_{s^c}. Before defining CC, however, we must first define the correlation among multiple random vectors.

The correlation between two random vectors $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$ can be defined as

$$\overline{\rho}(x_1, x_2) = \max_{\left\{ \begin{smallmatrix} g_1^T P_{x_1} g_1 = 1 \\ g_2^T P_{x_2} g_2 = 1 \end{smallmatrix} \right\}} g_1^T P_{x_1 x_2} g_2 \tag{15}$$

which is just the maximum correlation between any two linear functionals of x_1 and x_2. The correlation among m random vectors x_i is defined similarly as

$$\overline{\rho}(x_1, x_2, \cdots, x_m) = \max_{i \neq j} \overline{\rho}(x_i, x_j).$$

The conditional correlation $\overline{\rho}(x_1, x_2|y)$ is just the correlation after conditioning both x_1 and x_2 on the random vector y. To determine the matrix W with row dimension less than or equal to d that minimizes $\overline{\rho}(x_1, x_2|Wx_1)$, we make use of the following. There exist matrices $S_1 \in \mathbb{R}^{k_1 \times n_1}$ and $S_2 \in \mathbb{R}^{k_2 \times n_2}$ such that

$$\begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} P_{x_1} & P_{x_1 x_2} \\ P_{x_2 x_1} & P_{x_2} \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}^T = \begin{bmatrix} I_{k_1} & B \\ B^T & I_{k_2} \end{bmatrix} \tag{16}$$

where I_{k_i} is an identity matrix with k_i rows and B has nonzero entries only along its main diagonal. These nonzero entries $\beta_j = [B]_{jj}$ are the CC and they satisfy $1 \geq \beta_1 \geq \beta_2 \geq \cdots \beta_{\min(k_1,k_2)} \geq 0$. As shown in [20], the desired matrix W is given by the first d rows of S_1, i.e., $W = S_1(1:d,:)$, and

$$\overline{\rho}(x_1, x_2|Wx_1) = \beta_{d+1} \tag{17}$$

where $\beta_{d+1} \triangleq 0$ for $d \geq \min(k_1, k_2)$.

For multiscale modeling from CC, each internal variable $z(s) \in \mathbb{R}^{d(s)}$ must be computed in two steps. First, CC is used to determine the linear function $W_{s,1} f_{s\alpha_1}$ of dimension $d_1(s)$ that maximally decorrelates $f_{s\alpha_1}$ from $f_{s\alpha_1^c}$. Next, CC is used to determine the linear function $W_{s,2} f_{s\alpha_2}$ of dimension $d_2(s) = d(s) - d_1(s)$ that maximally decorrelates $f_{s\alpha_2}$ from $f_{s\alpha_2^c}$. The internal matrix is then given by $W_s = \text{diag}\{W_{s,1}, W_{s,2}\}$. Note that $z(s) = W_s f_s$ is not necessarily the $d(s)$-dimensional linear function of f_s that minimizes $\overline{\rho}(f_{s\alpha_1}, f_{s\alpha_2}, f_{s^c}|W_s f_s)$, but it is generally a very good approximation [20]. For the remainder of this paper, we will ignore the possible suboptimality and refer to the internal variable produced by CC as the optimal linear function.

Note that the multiscale realization algorithm based on CC computes the internal variables independently at each node [20]. Because computing the CC between two random vectors x_1 and x_2 requires the eigenvalue or Cholesky decomposition of both P_{x_1} and P_{x_2} as well as the SVD of $P_{x_1 x_2}$, the total algorithm will be computationally overwhelming when the number of finest-scale elements, equal to the dimension of f_0, is large. Secondly, computing the internal variables independently means that the approximations made at each node may be done inconsistently. However, if the multiscale model is to represent processes that are SSS with stationary increments or stationary processes presenting the long-range dependence phenomenon, these drawbacks can be overcome, as detailed in the following sections.

IV. The Multiscale Modeling of Statistical Self-Similarity

In this section, we show how to represent SSS processes at the finest scale of multiscale tree models. These models are determined using CC and thus provide an optimal tradeoff between model order (state dimensions) and statistical fidelity. The realization algorithm is made efficient by exploiting the SSS and the stationarity of the increments of the process to be represented at the finest scale. The basis for simplifying the realization algorithm for self-similar processes is provided in Section IV-A, where we show that the linear functionals that minimize the conditional correlation between two intervals of a SSS process can be derived from the linear functionals that minimize the conditional correlation between the process on a common contraction or dilation of these intervals. Because the internal variables of a multiscale model must minimize

the conditional correlation between intervals of the finest-scale process, the result of Section IV-A leads to a rule for determining an internal variable at one scale from an internal variable already computed at a coarser scale. This rule, however, requires that the result of Section IV-A be extended to discrete-time processes on a finite interval, since the finest scale of a multiscale process can only represent samples on a finite interval. These extensions are provided in Section IV-B.

The basis for using stationarity to simplify the multiscale realization algorithm is provided in Section IV-C. We show that the internal matrices produced by CC are relatively constant across a given scale of the multiscale process whenever the process to be represented at the finest scale is stationary or has stationary increments. The variation in the internal matrices is greatest at the coarsest scales and is due to the finite interval of representation. These results imply that once a single internal variable is known at scale m then so are all the other internal variables at scale m, assuming that m is not one of the coarsest scales.

The simplifications under SSS and stationarity lead to an efficient realization algorithm, namely, CC can be used to compute the internal variables at the first few scales of the tree. Next, self-similarity can be invoked to determine an internal variable at each finer scale. Finally, all of the other variables can be determined using stationarity or stationary increments. The end result is that the number of CC required by the algorithm is small and independent of the number of samples represented at the finest scale of the tree. The complete realization algorithm is summarized in Section V.

A. Statistical Self-Similarity and Canonical Correlations

Consider modeling N evenly spaced samples of a SSS process $x(t)$ at the finest scale of a binary multiscale tree. Let $x[n] \triangleq x(n\Delta t)$ for $1 \le n \le N$ be the samples. Assume for notational simplicity that that $N = 2^{M+2}$ for some positive integer M. Four consecutive samples of $x[n]$ can be mapped to each of the 2^M nodes at the finest scale of a binary tree with $M+1$ scales. The mapping of the first thirty-two samples is illustrated in Fig. 3 for $\Delta t = 1/16$. To see how SSS can be used in the modeling process, consider node s of the tree in Fig. 3. The finest-scale descendents of node s represent samples of $x(t)$ on the interval $(0,1]$, while the finest scale descendents of its parent $s\overline{\gamma}$ represent $x(t)$ on the interval $(0,2]$. Yet $x(t)$ on the interval $(0,1]$ is SSS to $x(t)$ on the interval $(0,2]$. Because of this similarity, we should expect that the internal variable at node s should be closely related, perhaps by some transformation, to the internal variable at node $s\overline{\gamma}$. Now consider node u in Fig. 3. Because $x(t)$ on the finest-scale interval descending from node $u, (1,2]$, is SSS to $x(t)$ on the interval $(1/2,1]$, which is the finest-scale interval descending from $s\alpha_2$, the internal variables at nodes u and $s\alpha_2$ should also be closely related.

To relate the internal variables at different scales, we first consider the decorrelation of two *intervals* of $x(t)$ and later apply this result to the decorrelation of vectors of samples of $x(t)$. We define the correlation between two intervals of $x(t)$ to be analogous to the correlation between two vectors

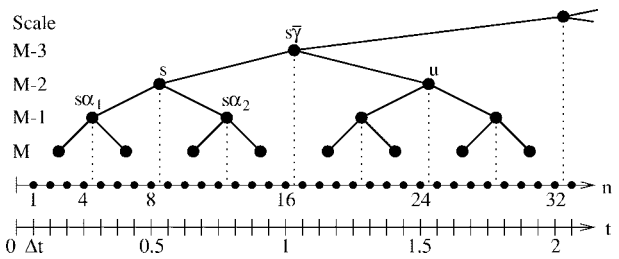


Fig. 3. The mapping of $x(n\Delta t)$ for $\Delta t = 1/16$ to the finest scale of a binary tree with $M+1$ scales. Four consecutive samples of $x(t)$ are represented by each variable at scale M.

in (15). Recall that the correlation between two scalar random variables u and v after conditioning on y is

$$\rho(u,v|y) = \frac{E[(u-m_u)(v-m_v)|y]}{\sigma_{u|y}\sigma_{v|y}}$$

where $\sigma_{u|y}$ is the variance of u after conditioning on y. Define $\mathcal{L}[t_1,t_2]$ to be the set of bounded linear functionals of $x(t)$ on the interval $[t_1,t_2]$. The correlation between $x(t)$ for $t \in [t_1,t_2]$ and $x(t)$ for $t \in [t_3,t_4]$, allowing for conditioning on a random variable y is defined as [20]

$$\overline{\rho}(x,[t_1,t_2],[t_3,t_4]|y) \triangleq \max_{\left\{\begin{smallmatrix}\ell_1 \in \mathcal{L}[t_1,t_2] \\ \ell_2 \in \mathcal{L}[t_3,t_4]\end{smallmatrix}\right\}} \rho(\ell_1(x),\ell_2(x)|y). \tag{18}$$

Due to the homogeneity of linear operators and because conditioning on γy is equivalent to conditioning on y, we have

$$\overline{\rho}(\alpha x,[t_1,t_2],[t_3,t_4]|\gamma y) = \overline{\rho}(x,[t_1,t_2],[t_3,t_4]|y) \tag{19}$$

for any two scalars α and γ.

Define $x^a(t) \triangleq x(at)$ for any $a > 0$. We will use the following theorem (proved in the Appendix) to relate internal variables at different scales when the finest-scale process is SSS.

Theorem 1: Assume that the bounded linear functional $\ell \in \mathcal{L}[t_1,t_2]$ satisfies

$$\ell = \arg \min_{\ell_0 \in \mathcal{L}[t_1,t_2]} \overline{\rho}(x,[t_1,t_2],[t_3,t_4]|\ell_0(x)). \tag{20}$$

For any SSS process $x(t)$ and any scalar $a > 0$, the linear functional $\hat{\ell}(x) \in \mathcal{L}[at_1,at_2]$ that satisfies $\hat{\ell}(x) = \ell(x^a)$ must also satisfy

$$\hat{\ell} = \arg \min_{\hat{\ell}_0 \in \mathcal{L}[at_1,at_2]} \overline{\rho}(x,[at_1,at_2],[at_3,at_4]|\hat{\ell}_0(x)). \tag{21}$$

Theorem 1 can be extended to open intervals or unions of intervals, e.g., $[t_1,t_2)$ or $(t_1,t_2) \cup (0,t_1/2)$ in lieu of $[t_1,t_2]$. Theorem 1 also applies if $\ell(x)$ is a vector of linear functionals.

Theorem 1 basically states that the linear functionals that maximally decorrelate two intervals of a SSS process can also be used to determine the linear functionals that maximally decorrelate any common dilations or contractions of these intervals. For realizing multiscale models when the finest-scale process is SSS, the utility of Theorem 1 is that internal variables at one scale can be used to determine internal variables at other scales, thereby reducing the number of

CC computations that are required. For example, consider deriving $z(s)$ from $z(s\overline{\gamma})$ in the tree illustrated in Fig. 3. The internal variable $z(s\overline{\gamma})$ must conditionally decorrelate $x(t)$ on the three intervals $(0,1], (1,2]$, and $(0,2]^c$, where $(0,2]^c$ is the complement of the interval $(0,2]$ in the entire interval represented by the finest scale of the tree. (Ignore for the moment that the finest scale of the tree represents samples of $x(t)$ not entire intervals.) Similarly, $z(s)$ must conditionally decorrelate $x(t)$ on $(0,1/2], (1/2,1]$, and $(0,1]^c$. Recall that the internal variables are computed in two steps when using CC. For $z(s\overline{\gamma})$, the first step is to compute the linear functionals of $x(t)$ on $(0,1]$ that maximally decorrelate $x(t)$ on $(0,1]$ from $x(t)$ on $(0,1]^c$. Similarly, for $z(s)$ the first step is to compute the linear functionals of $x(t)$ that maximally decorrelate $x(t)$ on $(0,1/2]$ from $x(t)$ on $(0,1/2]^c$. Because the interval $(0,1]$ is related to $(0,1/2]$ by a simple contraction, the first components of $z(s\overline{\gamma})$ and $z(s)$ should be related through Theorem 1. Define the linear functional that maximally decorrelates $x(t)$ on $(0,1]$ from $x(t)$ on $(0,1]^c$ by

$$\ell_1(x) = \int_0^1 g(t)x(t)\,dt \tag{22}$$

for some function $g(t)$. From Theorem 1, the linear functional $\hat{\ell}_1(x) = \ell_1(x^a)$ for $a = 1/2$ is the linear functional that maximally decorrelates $x(t)$ on $(0,1/2]$ from $x(t)$ on $(0,1/2]^c$. That is

$$\ell_1(x^a) = \int_0^1 g(t)x(t/2)\,dt \tag{23a}$$

$$= 2\int_0^{1/2} g(2t)x(t)\,dt \tag{23b}$$

$$= \hat{\ell}_1(x). \tag{23c}$$

From this analysis, once the internal variables at some scale of the tree have been computed, all of the other internal variables can be derived with the aid of Theorem 1. However, there are two issues with this line of reasoning. First, the finest-scale of the multiscale tree represents samples, not intervals, of $x(t)$ so that (23) cannot be directly applied. Secondly, in using Theorem 1 to derive $z(s)$ from $z(s\overline{\gamma})$, we implicitly assumed that $(0,1]^c$ is related to $(0,1/2]^c$ by a contraction of the time axis. This will be true only if $(0,1]^c = (1,\infty)$ and $(0,1/2]^c = (1/2,\infty)$, which is not the case if we are using a tree to model a finite interval of $x(t)$. The following subsection discusses how to overcome these problems and how the results lead to an efficient algorithm for representing SSS processes at the finest scale of multiscale trees.

B. Extending Theorem 1 to Sampled and Discrete-Time Processes

To illustrate how Theorem 1 can be adapted for multiscale modeling when the finest scale of the tree represents samples of a SSS process, return to the tree illustrated in Fig. 3. Assume that $z(s\overline{\gamma})$ has been computed as in Section III-B and we would like to determine the functional form of $z(s)$ directly from that of $z(s\overline{\gamma})$. Consider the first element of $z(s\overline{\gamma}), w_{s\overline{\gamma}}^T f_s$, which is the linear functional of f_s that maximally decorrelates f_s from f_{s^c}. The first element of $z(s), w_s^T f_{s\alpha_1}$, should maximally decorrelate $f_{s\alpha_1}$ from $f_{s\alpha_1^c}$. Since the intervals represented by f_s and $f_{s\alpha_1}$ are $(0,1]$ and $(0,1/2]$, respectively, we should expect that w_s is closely related to $w_{s\overline{\gamma}}$. However, the dimension of w_s is half that of $w_{s\overline{\gamma}}$, so that a contraction of the time axis, like that used in (23) for $g(t)$, cannot be directly applied to the derivation of w_s from $w_{s\overline{\gamma}}^T$. Instead, note that

$$w_{s\overline{\gamma}}^T f_s = \int_0^1 g(t)x(t)\,dt \tag{24a}$$

for

$$g(t) \triangleq \sum_{k=1}^{16} w_{s\overline{\gamma}}[k]\,\delta(t - k/16) \tag{24b}$$

where $w_{s\overline{\gamma}}[k]$ is the kth element of the vector $w_{s\overline{\gamma}}$. Because $\langle x, g\rangle$ is the linear functional of x that minimizes the conditional correlation between samples on $(0,1]$ and its complement, it will also approximately minimize $\overline{\rho}(x,(0,1],(0,1]^c|\ell(x))$ over all linear functionals $\ell \in \mathcal{L}(0,1]$. The accuracy of this approximation depends on the aliasing in the sampling of $x(t)$, i.e., the variation in $x(t)$ on $(0,1]$ not determined from the samples in f_s.

Assuming that conditioning on the linear functional $\langle x, g\rangle$ minimizes the correlation between $x(t)$ on $(0,1]$ and $x(t)$ on $(0,1]^c$, Theorem 1 shows that conditioning on $\langle x, \hat{g}\rangle$, where $\hat{g} = g(2t)$ minimizes the correlation between $x(t)$ on $(0,1/2]$ from $x(t)$ on $(0,1/2]^c$. (We assume for the moment that the interval represented at the finest scale has infinite length, i.e., $(0,1]^c = (1,\infty)$.) This linear functional will also minimize the correlation between $f_{s\alpha_1}$ and $f_{s\alpha_1^c}$, which vectors represent samples of $x(t)$ on the intervals $(0,1/2]$ and $(0,1/2]^c$, respectively. However, $\langle x, \hat{g}\rangle$ cannot be included in $z(s)$ since

$$\langle x, \hat{g}\rangle = \tfrac{1}{2}\sum_{k=1}^{16} w_{s\overline{\gamma}}[k]\,x(k/32) \tag{25}$$

is a function of samples of $x(t)$ that are not to be represented at the finest scale of the tree. A solution is to approximate this inner product with a linear function of the samples that are represented at the finest scale, e.g.,

$$w_s^T f_{s\alpha_1} = \tfrac{1}{2}\sum_{k=1}^{8} \underbrace{(w_{s\overline{\gamma}}[2k-1] + w_{s\overline{\gamma}}[2k])}_{w_s[k]}\,x(k/16) \tag{26}$$

which is obtained by replacing $x(k/32)$ in (25) for k odd with $x((k+1)/32)$. Another alternative is to replace $x(k/32)$ for k odd with $\rho\,x((k+1)/32)$, where ρ is the correlation coefficient between the two samples.

Assuming that the multiscale tree is binary and the SSS process to be represented at the finest scale is mapped to the finest-scale nodes as in Fig. 3, (26) leads to a more general method for determining internal variables from those already computed at coarser scales, namely, if the finest-scale interval descending from node τ represents samples of $x(t)$ on $(t_1,t_2]$,

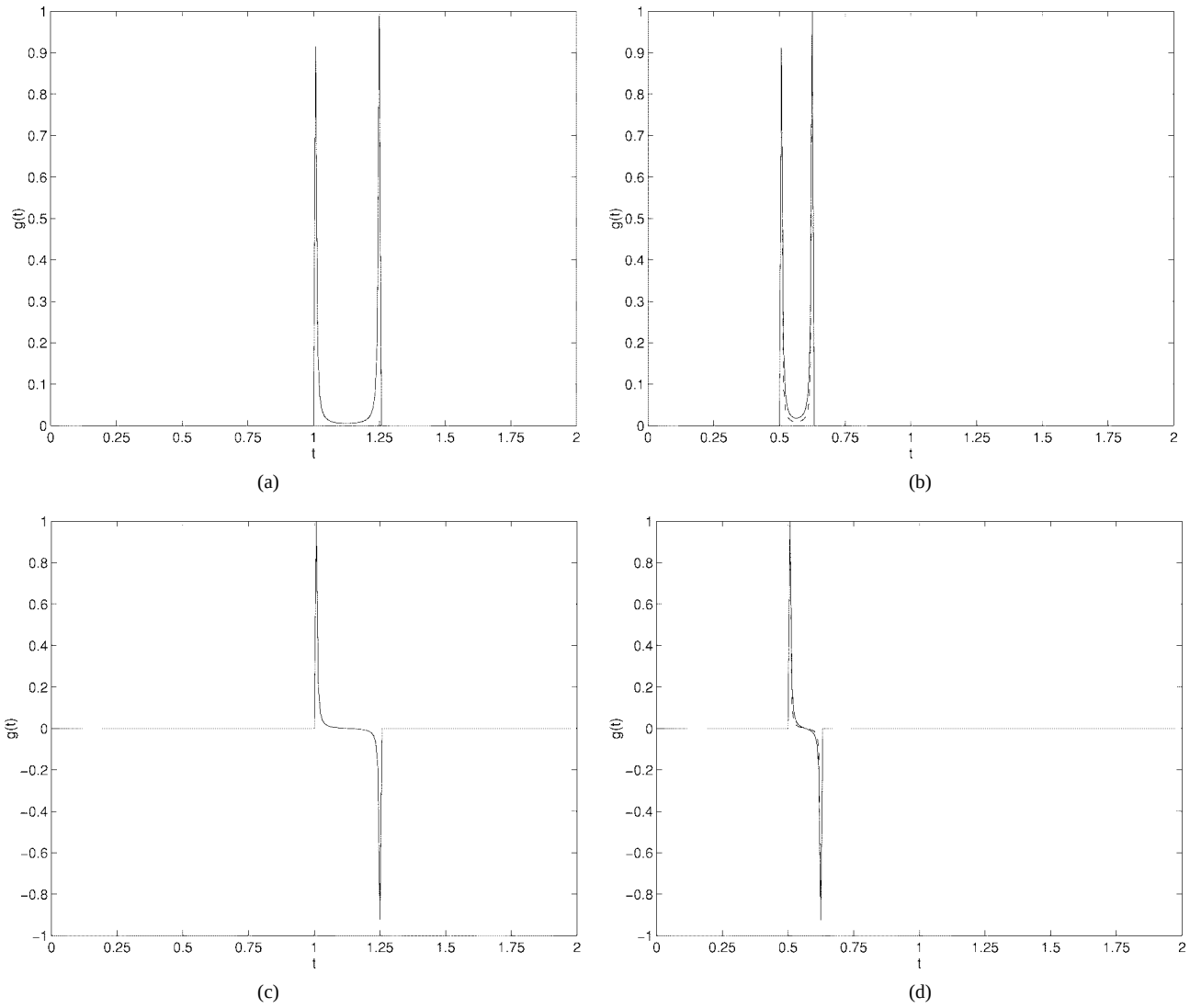


Fig. 4. For fBm with $H = 0.3$, the linear functional that minimizes the conditional correlation between $x(t)$ on the intervals (a) $(1, 5/4]$ and $(0, 1] \cup (5/4, 2]$ and (b) $(1/2, 5/8]$ and $(0, 1/2] \cup (5/8, 2]$. The dashed line in (b) is the linear functional derived from (27). (c) and (d) Identical plots for the next additional linear functionals that minimize the conditional correlation.

then the internal variable W_s can be derived from W_τ using (in MATLAB pseudocode)

$$W_s = \tfrac{1}{2}\left(W_\tau(:, 1:2:n-1) + W_\tau(:, 2:2:n)\right) \qquad (27)$$

whenever $m(s) = m(\tau) + 1$ and the finest-scale interval descending from node s is $(t_1/2, t_2/2]$. Note that W_s is determined by averaging neighboring columns of W_τ so that W_s has one-half the number of columns of W_τ. Assuming that $W_\tau f_\tau$ minimizes $\overline{\rho}(f_{\tau\alpha_1}, f_{\tau\alpha_2}, f_\tau^c | z(\tau))$ over all d-dimensional linear functions of $f_\tau, W_s f_s$ will approximately minimize $\overline{\rho}(f_{s\alpha_1}, f_{s\alpha_2}, f_s^c | z(s))$ over d-dimensional linear functions of f_s. This process can be continued recursively, using W_s to determine an internal variable at scale $m(s)+1$, assuming that $m(s)+1$ is not the finest scale.

To justify the use of (27) in the multiscale modeling of fBm and other SSS processes, we provide a few examples. Consider modeling fBm for $t \in (0, 2]$ with $\Delta t = 1/128$ and $H = 0.3$. Fig. 4(a) illustrates the linear functional[4] of (samples of) $x(t)$ on $(1, 5/4]$ that minimizes the conditional correlation between samples on $(1, 5/4]$ and $(1, 5/4]^c = (0, 1] \cup (5/4, 2]$. The solid line in Fig. 4(b) illustrates the linear functional of (samples of) $x(t)$ on $(1/2, 5/8]$ that minimizes the conditional correlation between samples on $(1/2, 5/8]$ and $(1/2, 5/8]^c = (0, 1/2] \cup (5/8, 2]$. Both linear functionals were computed using CC. The dashed line in Fig. 4(b) illustrates the linear functional derived from the linear functional in Fig. 4(a) using (27). The linear functionals derived from CC and from (27) are nearly identical. Fig. 4(c) and (d) show the linear functionals that when combined with the linear functionals in Fig. 4(a) and (b) minimize the conditional correlation between the respective intervals. Again, the linear functionals derived from CC and from (27) are nearly identical.

The difference between the linear functionals derived from (27) and the optimal linear functionals derived from CC is due

[4] By "illustrating the linear functional" we mean that the vector g is plotted where $g^T f$ is the linear functional and f is a vector of all of the samples on $(0, 2]$.

to both sampling and a finite interval of representation. Recall that Theorem 1 applies to continuous-time processes and that both of the intervals must be related by a common contraction. While $(1/2, 5/8]$ is the interval corresponding to $(1, 5/4]$ after the time axis is contracted by two, the complement $(1/2, 5/8]^c$ does not correspond exactly to $(1, 5/4]^c$ after a contraction by two. As a general rule, when the interval of interest increases in size, i.e., $(0, T]$ for T greater than two, the difference between the linear functionals computed by CC and those computed from (27) decreases.

The accuracy of (27) in determining linear functionals from those already computed applies to fBm for all $0 < H < 1$. Also, remember that (27) was derived assuming only that the finest scale of the multiscale process is to represent a SSS process, not just fBm. Consider FDGN, which is not SSS but has a geometrically decaying covariance function for large lags [see Fig. 4(b)]. Consider FDGN for $h = 0.2$ on the interval $n \in [1, N]$ for $N = 256$. (Recall that FDGN is a discrete-time process.) The linear functional that minimizes the correlation between $w_h[n]$ on $(N/2, 5N/8]$ and $(N/2, 5N/8]^c$ is illustrated in Fig. 5(a). The linear functional that minimizes the correlation between $w_h[n]$ on $(N/4, 5N/16]$ and $(N/4, 5N/16]^c$ is illustrated in Fig. 5(b). Both of these linear functionals are computed using CC. The linear functional derived from the linear functional in Fig. 5(a) using (27) is illustrated by the dashed line in Fig. 5(b). Again note the close correspondence between the two linear functionals plotted in Fig. 5(b). The additional linear functionals that decorrelate $(N/4, 5N/16]$ and $(N/4, 5N/16]^c$ can also be derived quite accurately from (27). This demonstrates that (27) or similar relationships can be used in the multiscale modeling of SSS processes such as fBm or long-range dependence processes such as DFGN and FDGN.

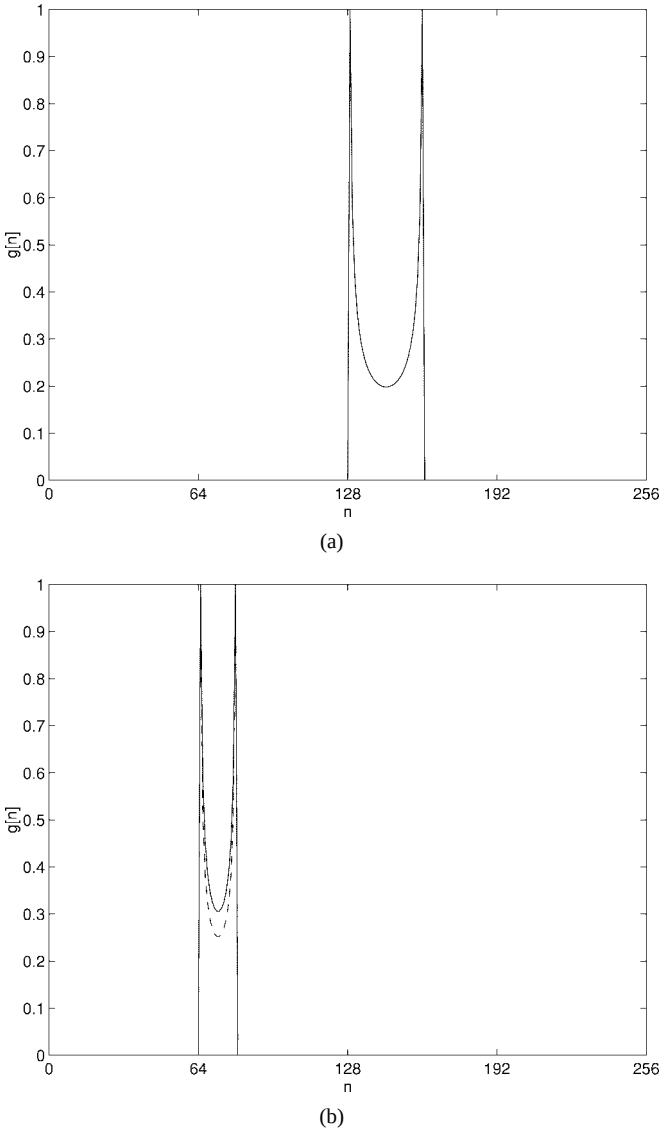


Fig. 5. For FDGN with $h = 0.2$ and $n \in [1, 256]$, the linear functional that minimizes the conditional correlation between $w_h[n]$ on the intervals (a) $(128, 160]$ and $(128, 160]^c$ and (b) $(64, 80]$ and $(64, 80]^c$. The dashed line in (b) is the linear functional derived from (27).

C. Using Stationarity and Stationary Increments

While (27) reduces the computation required for multiscale modeling, we are still left with a considerable number of CC's to compute. The problem is that the number of nodes increases by two with each increase in scale, while (27) can be used to compute only *one* internal variable at each scale from any internal variable at a coarser scale. Thus, even if all of the internal variables are known at some scale, only $(1/2)^k$ of the variables at the level k scales finer can be computed from (27). The solution is to take advantage of stationarity for FDGN and DFGN and stationary increments for fBm. The end result is that the internal matrices W_s remain approximately constant for all nodes at a given scale, meaning that no additional CC need to be computed once *one* internal variable has been computed at each scale.

To see why the internal matrices are relatively constant across a given scale when the finest-scale process is stationary, consider the multiscale modeling of samples of DFGN on the interval $(0, T]$. Assume a finest-scale mapping of the form illustrated in Fig. 3. At scale $m = 2$, the four internal variables have finest-scale descendents on the intervals $(0, T/4], (T/4, T/2], (T/2, 3T/4]$, and $(3T/4, T]$. The internal variable whose finest scale descendents represent DFGN on the interval $(T/4, T/2]$ must conditionally decorrelate samples of DFGN on the interval $(T/4, T/2]$ from samples on $(T/4, T/2]^c$. These two intervals are illustrated by the darkly and lightly shaded strips at the top of Fig. 6. Call this internal variable $z_1 = W_1 f_1$, where f_1 is the finest scale process on $(T/4, T/2]$. The internal variable whose finest scale descendents represent DFGN on the interval $(T/2, 3T/4]$ must conditionally decorrelate samples of DFGN on the interval $(T/2, 3T/4]$ from samples on $(T/2, 3T/4]^c$. These two intervals are illustrated by the darkly and lightly shaded strips in the middle of Fig. 6. Call this internal variable $z_2 = W_2 f_2$, where f_2 is the finest scale process on $(T/2, 3T/4]$. The internal matrices W_1 and W_2 can be computed independently using CC. Another solution, however, is to compute only W_1 and define $z_2 = W_1 f_2$. This internal variable is approximate in the sense that the intervals illustrated at the bottom of Fig. 6, rather than those in the middle, will be decorrelated by

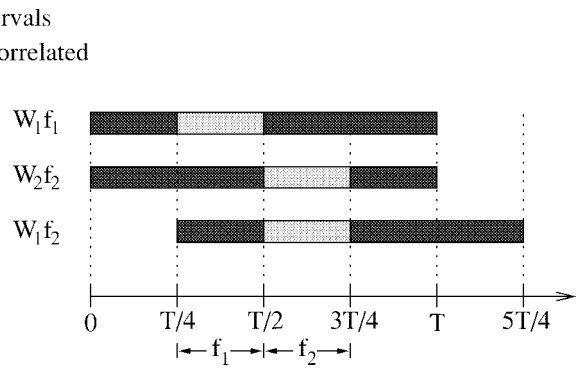


Fig. 6. The vector $W_1 f_1$ decorrelates the finest-scale process on the interval $(T/4, T/2]$ from $(0, T/4] \cup (T/2, T]$, while $W_2 f_2$ decorrelates $(T/2, 3T/4]$ from $(0, T/2] \cup (3T/4, T]$. The linear function $W_1 f_2$ can be used as an approximation of $W_2 f_2$, only it does not exactly decorrelate $(T/2, 3T/4]$ from $(0, T/4]$.

$z_2 = W_1 f_2$. The accuracy of the approximation, i.e., how well $W_1 f_2$ accomplishes the decorrelation of $W_2 f_2$, will depend on the size of the lightly shaded interval relative to the size of the entire interval of representation $(0, T]$ and how close the lightly shaded interval is to the edges of $(0, T]$. In other words, the approximation is best at the finer-scale nodes and at nodes away from the boundary of the tree.

For fBm, the reason why the internal matrices are approximately "shift-invariant" across any given scale is a little more subtle, since fBm is not stationary but has stationary increments. To show that the stationary increments of fBm leads to internal matrices that are effectively shift-invariant, we examine the linear function that minimizes the conditional correlation between samples of fBm on neighboring intervals. Denote the sampled process by $x[n] \triangleq x(n\Delta t)$ and consider the two p-sample intervals $n \in [n_0 + 1, n_0 + p]$ and $n \in [n_0 + p + 1, n_0 + 2p]$. For

$$x_{n_0} \triangleq \begin{bmatrix} x[n_0+1] \\ x[n_0+2] \\ \vdots \\ x[n_0+p] \end{bmatrix} \quad \text{and} \quad X_{n_0} \triangleq \begin{bmatrix} x[n_0+p+1] \\ x[n_0+p+2] \\ \vdots \\ x[n_0+2p] \end{bmatrix} \tag{28}$$

define T_{n_0} as the d-by-p matrix that minimizes $\overline{\rho}(x_{n_0}, X_{n_0} | T_{n_0} x_{n_0})$. The matrix T_{n_0} can be found using CC. To show that the internal matrices are approximately shift-invariant, we will show that T_{n_0} is asymptotically independent of n_0 for $n_0 \gg p$.

From the definition of $\overline{\rho}$, it follows that for any invertible m-by-m transformations Q_1 and Q_2, the d-row matrix T_{n_0} that minimizes $\overline{\rho}(x_{n_0}, X_{n_0} | T_{n_0} x_{n_0})$ is related to the d-row matrix W_{n_0} that minimizes $\overline{\rho}(Q_1 x_{n_0}, Q_2 X_{n_0} | W_{n_0} Q_1 x_{n_0})$ by $T_{n_0} = W_{n_0} Q_1$. Therefore, determining the d-dimensional linear function of x_{n_0} that minimizes the conditional correlation between x_{n_0} and X_{n_0} is equivalent to determining the d-dimensional linear function of $Q_1 x_{n_0}$ that minimizes the conditional correlation between $Q_1 x_{n_0}$ and $Q_2 X_{n_0}$.

Because fBm has stationary increments, the process $w[n] = x[n] - x[n-1]$ is stationary and its covariance is given by sampling $r_{x_\Delta}(\tau)$ in (4a). Define Q_1 and Q_2 to be the invertible transformations that satisfy

$$y_{n_0} = Q_1 x_{n_0} = \begin{bmatrix} x[n_0+1]/\sigma_{n_0+1} \\ w[n_0+2] \\ \vdots \\ w[n_0+p] \end{bmatrix}$$
$$Y_{n_0} = Q_2 X_{n_0} = \begin{bmatrix} x[n_0+p+1]/\sigma_{n_0+p+1} \\ w[n_0+p+2] \\ \vdots \\ w[n_0+2p] \end{bmatrix} \tag{29}$$

where σ_n is the standard deviation of $x[n]$. The CC analysis of y_{n_0} and Y_{n_0} depends on the covariance matrices $E[y_{n_0} y_{n_0}^T], E[Y_{n_0} Y_{n_0}^T]$, and $E[y_{n_0} Y_{n_0}^T]$. Each of these matrices has the form

$$\left[\begin{array}{c|c} \alpha & a^T \\ \hline b & P_w \end{array}\right] \tag{30}$$

where

- $P_w \in \mathbb{R}^{(p-1)\times(p-1)}$ has entries of the form

$$E[w[n_0] w[n_0 + r]]$$

  for $r = 1, \cdots, 2p-1$, and is independent of n_0 due to the stationarity of the increments process $w[n]$;
- $a \in \mathbb{R}^{(p-1)\times 1}$ and $b \in \mathbb{R}^{(p-1)\times 1}$ have entries of the form

$$E[x[n_0+1] w[n_0+r]]/\sigma_{n_0+1}$$

and

$$E[x[n_0+p+1] w[n_0+r]]/\sigma_{n_0+p+1}$$

  for $r = 2, \cdots, 2p$;
- α is a scalar.

To show that the matrix W_{n_0} returned by CC is slowly varying and asymptotically independent of n_0, we only need to show the same for α, a and b for each of the three covariance matrices.[5]

For $E[y_{n_0} y_{n_0}^T]$ and $E[Y_{n_0} Y_{n_0}^T], \alpha = 1$ for all values of n_0. For $E[y_{n_0} Y_{n_0}^T]$

$$\begin{aligned} \alpha &= E[x[n_0+1]\, x[n_0+p+1]]/\sigma_{n_0+1}\sigma_{n_0+p+1} \\ &= \frac{\frac{1}{2}\left(|n_0+1|^{2H} + |n_0+p+1|^{2H} - |p|^{2H}\right)}{|n_0+1|^H |n_0+p+1|^H} \\ &\approx 1, \qquad \text{for } n_0 \gg p. \end{aligned}$$

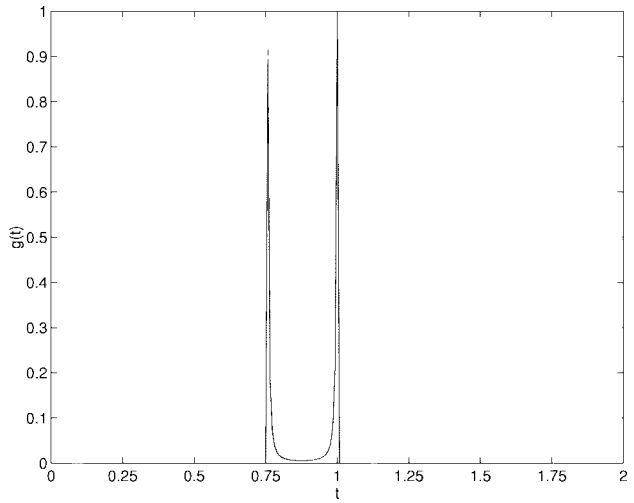
[5] Note that even if α, a, and b are slowly varying; W_{n_0} might not be slowly varying when the corresponding covariance matrices are ill-conditioned. However, this does not a detract from our case. Because of the stability of the SVD (singular-value-decomposition)-based CC decomposition, the subspace spanned by $W_{n_0} y_{n_0}$ should be relatively insensitive to perturbations in the corresponding covariance matrices, meaning that the conditioning information remains unchanged. In this case, there should still exist a series of vectors $\bar{W}_{n_0} y_{n_0}$ that is nearly statistically identical to $W_{n_0} y_{n_0}$ and for which $\bar{W}_{n_0}$ is slowly varying.

Now examine the entries of a and b that have the form

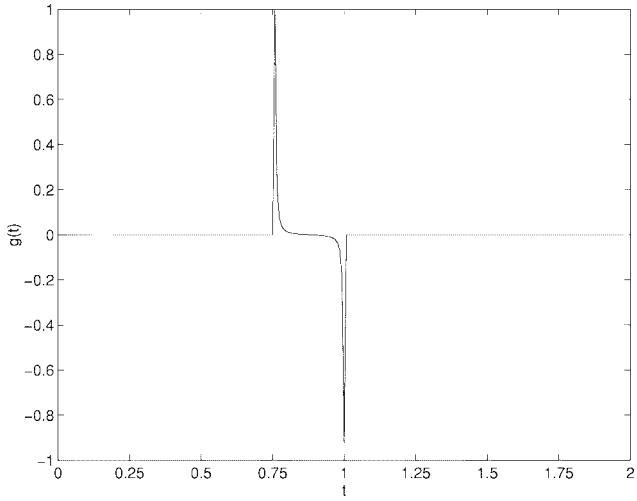
$$\begin{aligned}\frac{E[x[n_0]w[n_0+r]]}{\sigma_{n_0}} &= \frac{\sigma(\Delta t)^H}{2|n_0|^H}\left[|n_0+r|^{2H}-|n_0+r-1|^{2H}\right.\\ &\quad \left.+|r-1|^{2H}+|r|^{2H}\right]\\ &\approx \frac{\sigma(\Delta t)^H}{2}\left[2H|n_0|^{H-1}\right.\\ &\quad \left.+(|r-1|^{2H}-|r|^{2H})|n_0|^{-H}\right], \qquad \text{for } n_0 \gg r\end{aligned}$$

where the last approximation is made using a Taylor series expansion. This shows that a and b for all three covariance matrices are slowly varying and all of the elements decay to zero at the rate of $n_0^{\max(-H,H-1)}$. Thus, W_{n_0} is slowly varying and so must be the internal variables of the corresponding multiscale model (assuming, of course, that the edge effects due to the finite interval of representation are again negligible).

To demonstrate the shift-invariance of the internal variables, consider modeling fBm for $t \in (0,2]$ with $\Delta t = 1/128$ and $H = 0.3$. The linear functional that minimizes the conditional correlation between samples of $x(t)$ on $(3/4,1]$ and those on $(3/4,1]^c$ is illustrated by the solid line in Fig. 7(a). The solid line in Fig. 7(b) is the linear functional of $x(t)$ on $(3/4,1]$ that minimizes the remaining conditional correlation. The dashed lines in Fig. 7(a) and (b) are given by appropriately shifting in t the linear functionals in Fig. 4(a) and (c), respectively. The correspondence between the linear functionals computed by CC and those determined by invoking shift-invariance is so close that the dashed lines are hardly visible.



(a)



(b)

Fig. 7. For fBm with $H = 0.3$, a comparison of linear functionals computed from CC (solid line) and linear functionals determined from Fig. 4 by invoking shift-invariance (dashed line). Note that the dashed and solid lines are, as expected, indistinguishable.

V. An Algorithm for the Multiscale Modeling fBm, DFGN, and FDGN

This section outlines a complete algorithm for the multiscale modeling of fBm, DFGN, and FDGN that is based on the preceding analysis. Consider modeling N samples of a SSS process with stationary increments or of a stationary long-range dependent process. Assume for simplicity that $N = 2^{M+2}$ for some positive integer N. The N samples can be mapped to the finest scale of a binary tree with $M+1$ scales, where each finest-scale variable represents four consecutive samples of the process.

The first step of the multiscale modeling is to compute the internal matrices W_s at each node not at the finest scale, i.e., $m(s) \neq M$. (The internal matrices at $m = M$ are four-by-four identity matrices.) Remember that we are interested in approximate representations of the finest-scale process, e.g., fBm, since an exact model would generally require internal variables $W_s f_s$ with the same dimension as f_s. Thus, the following algorithm requires the specification at each node of either the maximum state dimension $d(s)$ or the maximum conditional correlation ϵ_s, i.e., $\overline{\rho}(f_{s\alpha_1}, f_{s\alpha_2}, f_{s^c}|W_s f_s) \leq \epsilon_s$.

To determine the internal matrices using a minimum number of CC decompositions, we will, of course, make use of the self-similarity and shift-invariance of the internal matrices, as detailed in the previous two subsections. Recall that the self-similarity and shift-invariance of the internal variables are most descriptive (in terms of minimizing residual correlation) when the finest-scale descendents of the internal variable (from which all other internal variables are computed) are on an interval that is both narrow relative to the length of the entire finest-scale interval and far from the edges of the entire interval. Therefore, one must decide for how many scales $(m = 0, \cdots, m_0)$ to compute the internal variables using CC where the remaining internal variables are computed using (27) and shift-invariance. The tradeoff is between accuracy and computations. The general algorithm has the following form.

1) Use CC to compute the internal matrices at scales $m = 0, \cdots, m_0 - 1$. The dimensions of the internal variables are controlled by $d(s)$ or ϵ_s.
2) Use CC to compute W_β, where β is the 2^{m_0}th (middle) node at scale m_0.
3) Determine all the remaining internal matrices at scales $m = m_0, \cdots, M-1$ using (27) and shift-invariance.

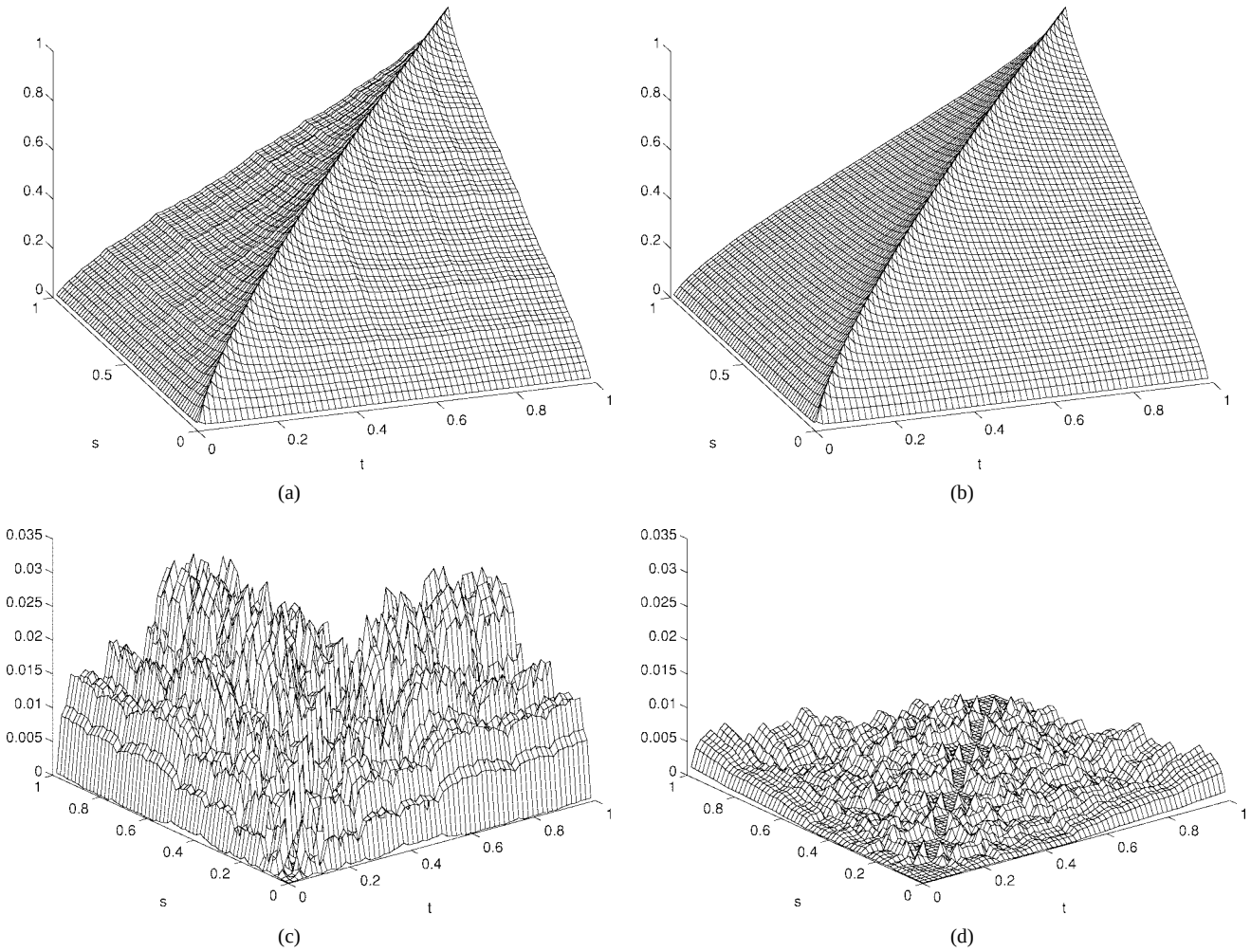


Fig. 8. The finest-scale covariance of multiscale approximations of fBm for $\sigma^2 = 1$ and $H = 0.3$ when the state dimensions are (a) four and (b) six. The absolute value of the difference between (1) and the finest-scale covariances in (a) and (b) are illustrated in (c) and (d), respectively.

4) Determine the multiscale model parameters using (12) and (14).

All of the examples in the following section use $m_0 = 2$. In this case, the entire algorithm requires seven[6] CC decompositions of the finest-scale covariance matrix $E[f_0 f_0^T]$ independent of the number of samples to be represented at the finest scale. The effects of the finite interval of representation can be further minimized by increasing m_0. However, the following examples show that multiscale representations of fBm and FDGN are very accurate for $m_0 = 2$.

VI. Example Multiscale Models

This section provides a number of examples that demonstrate both the effectiveness of the algorithm detailed in Section V for the multiscale modeling of SSS processes and the utility of the multiscale framework for estimating and simulating these random processes.

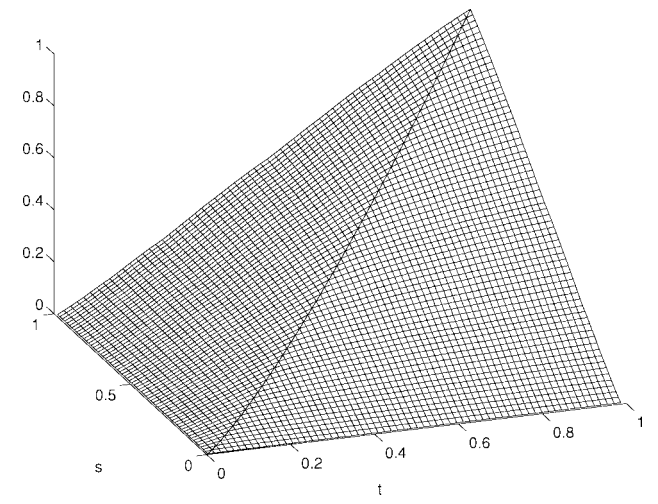
A. fBm

Consider modeling samples of fBm for $\sigma^2 = 1$ and $H = 0.3$ on $(0, 1]$ with $\Delta t = 1/256$. These 256 samples can be represented at the finest scale of a seven scale $(M = 6)$ binary tree. Using the algorithm of Section V, the finest scale of the multiscale model has the covariance illustrated in Fig. 8(a) when the dimension of all of the internal variables is four. The absolute value of the difference between (1) and the finest-scale covariance is plotted in Fig. 8(c). Note that the variance of fBm is modeled exactly by the multiscale process, which is a by-product of (14).

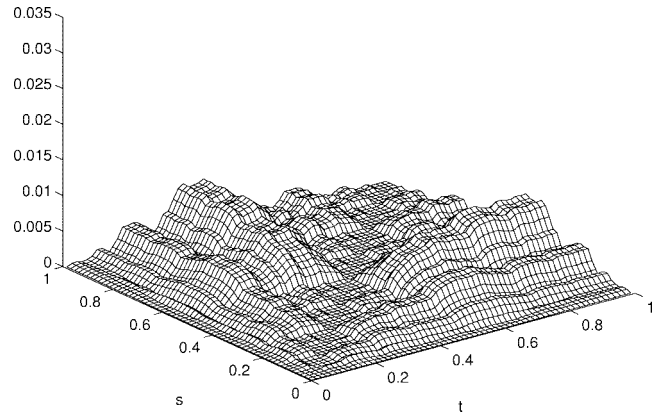
If the state dimensions are increased, the decorrelation produced by each internal variable will increase, leading to more accurate models. The finest-scale covariance and modeling error when the state dimensions are fixed at six—except at the finest scale, where the state dimension is fixed at four—are illustrated in Fig. 8(b) and (d), respectively. Note that the error in modeling the covariance function decreases by roughly an order of magnitude when the state dimensions increase from four to six.

As a second example, consider the multiscale modeling of fBm for $H = 0.7$. The finest-scale covariance and modeling error for a state dimension of four are illustrated in Fig. 9. As for $H = 0.3$, the finest scale of the multiscale model provides a very accurate approximation of fBm, even when the state dimensions are small. More generally, the multiscale models realized using the algorithm of Section V provide very

[6] Recall that W_0 requires only one CC, while the other internal variables require two.



(a)



(b)

Fig. 9. (a) The finest-scale covariance of multiscale approximations of fBm for $\sigma^2 = 1, H = 0.7$. All of the states have a dimension of four. (b) The absolute value of the difference between the finest-scale covariance and (1).

accurate representations of fBm for all values of $0 < H < 1$, even when the state dimensions are small. Also, the finest-scale covariances of these models are nearly identical to the finest-scale covariances of multiscale models for which CC is used to compute the internal matrix at every node on the tree. For example, the matrices formed by the errors in Figs. 8(c) and 9(b) have Frobenius norms of 3.6 and 0.69, respectively, while the errors for the corresponding four-dimensional multiscale models computed using CC at every node have Frobenius norms of 2.4 and 0.40.

Another important property of the multiscale models of fBm is that the modeling errors (when the state dimension is fixed) do not change significantly when the resolution $(1/\Delta t)$ or size of the interval represented at the finest scale increases. For instance, when $\sigma^2 = 1, H = 0.3$ and the state dimension is four, the ratio of the Frobenius norm of the error in modeling the covariance to the Frobenius norm of the covariance matrix of the samples represented at the finest scale is 0.023 for $\Delta t = 1/128$, 0.032 for $\Delta t = 1/256$, and 0.039 for $\Delta t = 1/512$. (We normalize by the Frobenius norm of the covariance, since increasing the number of samples will increase the Frobenius norm of the error, even when the magnitude of the errors remains constant.) This means that low-order multiscale models provide accurate representations of fBm now matter how fine the resolution or how large the interval to represented.

Multiscale models are of interest for the modeling of fBm and other SSS processes not only because they provide accurate and efficient representations, but also because the efficient estimation and simulation algorithms of the multiscale framework can be used. While the multiscale estimator requires $\mathcal{O}(Nd^3)$ computations when the state dimensions are d and the number of samples represented at the finest scale is N, we just demonstrated that the state dimensions are independent of N for a desired level of fidelity in the model. Thus, for a given error tolerance, the complexity of the estimator actually grows as $\mathcal{O}(N)$. This growth in computations is independent of the number of measurements incorporated, whereas a standard implementation of the normal equations would require $\mathcal{O}(N^3)$ computations when all or nearly all of the finest-scale samples are measured. As an example, consider the estimation of fBm for $\sigma^2 = 1, H = 0.2$, and $t \in (0, 1]$ from sparse, noisy, and irregularly sampled measurements. Assume a multiscale model for which the dimension of all the states is four. A sample path of fBm is illustrated in Fig. 10(a). The measurement noise is assumed to have a standard deviation of 0.05 and the measurements are indicated by the $\circ$'s in Fig. 10(a). The estimate of fBm produced by the multiscale model is illustrated by the solid line in Fig. 10(b). The dotted line indicates the estimate produced using an exact model of fBm. Without any additional computations, the multiscale estimator also produces the variance of the estimation error. The one standard deviation error lines are indicated by the dashed lines in Fig. 10(b). Note that the difference between the exact and approximate (multiscale) estimate is well within the one standard deviation error.

Another useful feature of the multiscale estimator is that it also produces the coefficients for a multiscale model that represents the estimation error. Because multiscale models can be efficiently simulated by simply evaluating the autoregression in (8), the multiscale error models can provide conditional simulations that are necessary for any applications requiring Monte Carlo analysis. Two conditional sample paths corresponding to the example of Fig. 10 are provided in Fig. 11. Note that the conditional sample paths nearly pass through each of the measurements. The advantage of the multiscale framework is that once the parameters of the multiscale error model are returned by the estimation algorithm, each conditional simulation can be computed in (approximately) Nd computations where N is the number of finest-scale samples and d is the dimension of the states in the model. By contrast, a Cholesky factorization of the estimation error covariance, which is nonstationary, will require $\mathcal{O}(N^3)$ computations.

Other features of the multiscale framework are that arbitrary nonlocal measurements of fBm can also be incorporated by the estimator [16], and the likelihood calculator can be used to estimate H and σ from noisy measurements of sample paths of fBm, although more crude models of fBm frequently suffice for the latter application [18].

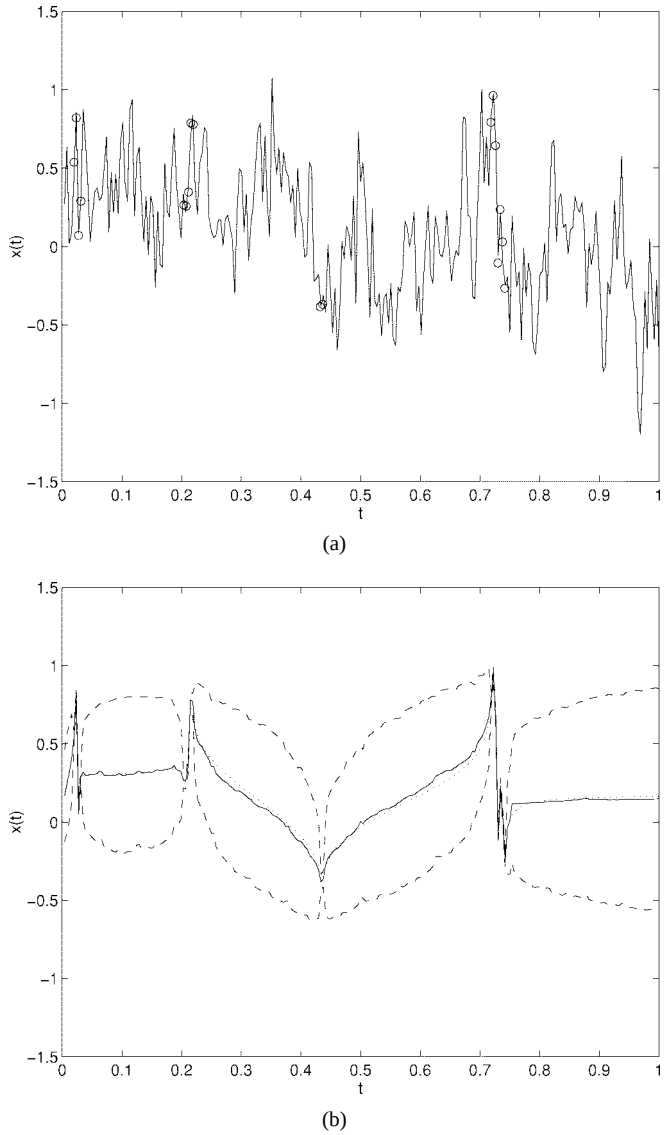


Fig. 10. (a) A sample path of fBm for $\sigma^2 = 1$ and $H = 0.2$. Noisy measurements are illustrated by o's. (b) The estimate produced at the finest scale of the multiscale model with state dimension four (solid line), the estimate based on the exact fBm covariance (dotted line), and the one standard deviation estimation errors (dashed line).

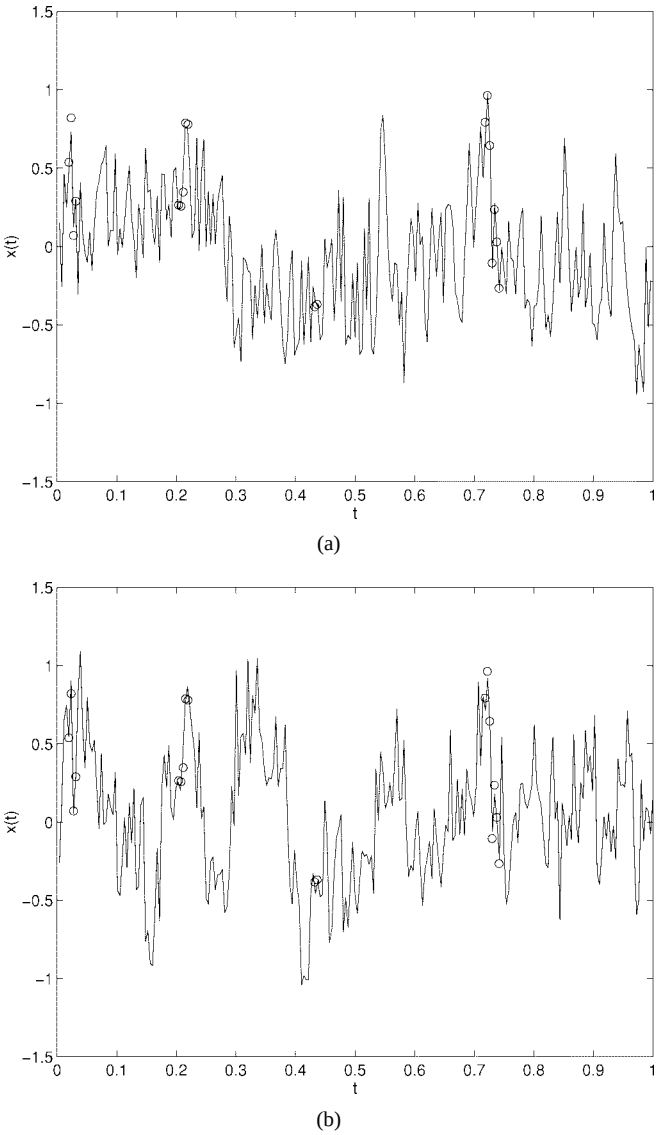


Fig. 11. Two conditional simulations of fBm conditioned on the measurements illustrated in Fig. 10.

B. FDGN

Now consider modeling 256 samples of FDGN using the multiscale realization algorithm of Section V. For $h = 0.2$, the finest-scale covariance of the multiscale model with state dimensions set to four is given by the solid line in Fig. 12(a). The dotted line in Fig. 12(a) is $r_h[n-128]$ where $r_h[n]$ is the covariance function of FDGN. While FDGN is stationary, the process at the finest scale of the multiscale model for FDGN is not exactly stationary, and will vary slightly with the location of the finest-scale sample. Thus, the solid line Fig. 12(a) really represents only a single column of the finest-scale covariance matrix, but the approximation errors given in this plot are typical of all the columns.

The finest-scale covariance of the multiscale model with state dimensions set to six is plotted in Fig. 12(b). As would be expected, increasing the state dimension from four to six leads to a noticeable reduction in modeling errors. For $h = -0.2$, the finest-scale covariance function is plotted in Fig. 12(c) and (d) for the multiscale models with state dimensions of four and six. Again, the multiscale model is quite accurate even when the state dimensions are limited to four and the representation improves as the state dimension increases from four to six.

Results similar to these were obtained for the multiscale modeling of DFGN, which should not be surprising given the close correspondence between the covariances of FDGN and DFGN.

VII. Conclusion

This paper developed an efficient realization algorithm for the modeling of SSS processes using the multiscale processes introduced in [13]. The models were then applied to fBm, DFGN, and FDGN, demonstrating the tradeoff between accuracy and state dimension. The multiscale models are quite accurate even when the state dimensions are very small, and the accuracy of the representations remains relatively constant

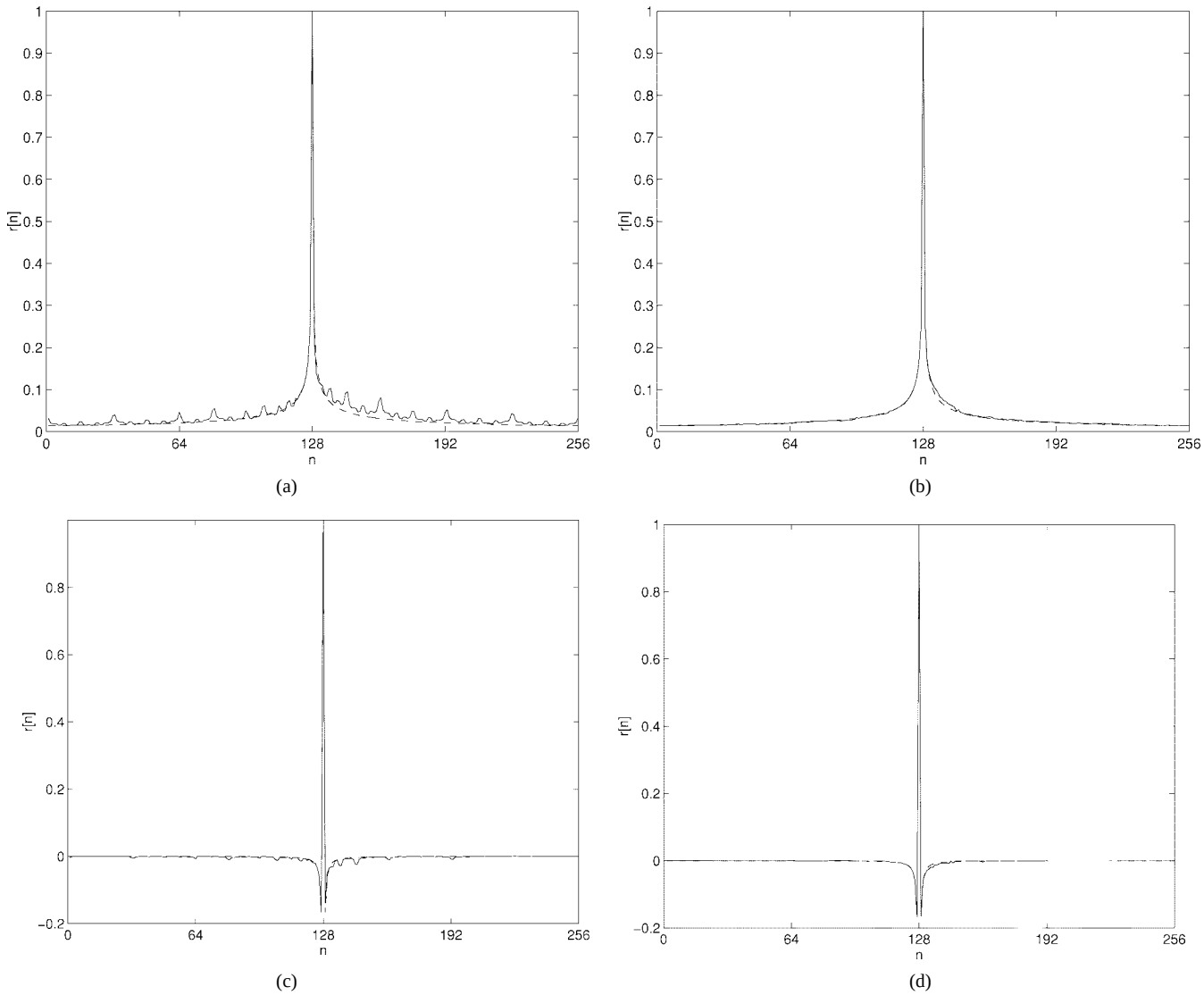


Fig. 12. The finest-scale covariance of multiscale approximations of FDGN for $h = 0.2$ and state dimensions of (a) four and (b) six and for $h = -0.2$ and state dimensions of (c) four and (d) six. The solid line is the covariance of the multiscale model, and the dashed line is that of FDGN.

when the state dimension is fixed and the number of samples represented at the finest scale is increased. This means that the processing algorithms of the multiscale framework have $\mathcal{O}(N)$ growth in computations for a desired level of accuracy. The flexibility and processing power of the multiscale framework were demonstrated by estimating fBm from sparse irregularly sampled measurements and then generating conditional simulations.

The most significant issue not addressed in this paper is the number of computations required by the CC in the algorithm of Section V. While the number of CC decompositions does not increase with N, the number of computations required to compute each CC is certainly a function of N and, in fact, grows cubicly with N. One possible solution to this growth in complexity is suggested by the fact that independent of the size of N, only a very small number of linear functions of the finest-scale process are generally desired from the CC since low-dimensional models are quite accurate even for large N. Also, these linear functions are rather smooth functions of the finest-scale process, with all of the detail concentrated at the boundaries of the finest-scale intervals to be decorrelated. Therefore, there is really no need to perform a CC analysis of the entire finest-scale covariance matrix. Instead, especially at the coarser-scale nodes, we are only interested in the linear functions that decorrelate a reduced-order subspace of the finest-scale process. The form of the linear functionals gives us some insight for choosing this subspace, but a complete analysis is beyond the scope of this paper.

A more long-term objective of this research is to model statistically self-similar processes using multiscale models without having to perform the intermediate step of computing the internal matrices. Namely, we would like to determine the form of the autoregressive moving-average (ARMA) parameters A_s and Q_s that leads to finest-scale processes that have polynomial decay in the correlation function without having to operate on the covariance matrix for the finest scale process. Another objective is to discover whether or not it is possible to represent two-dimensional SSS random processes with low-order multiscale models.

APPENDIX
PROOF OF THEOREM 1

For any two linear functionals $\ell_1 \in \mathcal{L}[t_1, t_2]$ and $\ell_2 \in \mathcal{L}[t_3, t_4]$, statistical self-similarity implies that

$$\rho(\ell_1(x), \ell_2(x)) = \rho(\ell_1(x^a), \ell_2(x^a)).$$

Riesz's Lemma [27] shows that there exists a function $g(t)$ such that

$$\ell(x) = \int g(t)x(t)\,dt$$

for any bounded linear functional $\ell(x)$. Applying a change of variables to such integrals, there exist linear functionals $\hat{\ell}_1 \in \mathcal{L}[at_1, at_2]$ and $\hat{\ell}_2 \in \mathcal{L}[at_3, at_4]$ such that $\hat{\ell}_1(x) = \ell_1(x^a)$ and $\hat{\ell}_2(x) = \ell_2(x^a)$. The existence of these linear functionals leads to

$$\begin{aligned}
&\max_{\left\{\substack{\ell_1 \in \mathcal{L}[t_1,t_2] \\ \ell_2 \in \mathcal{L}[t_3,t_4]}\right\}} \rho(\ell_1(x), \ell_2(x) | \ell(x)) \\
&\quad = \max_{\left\{\substack{\ell_1 \in \mathcal{L}[t_1,t_2] \\ \ell_2 \in \mathcal{L}[t_3,t_4]}\right\}} \rho(\ell_1(a^H x), \ell_2(a^H x) | \ell(a^H x)) \\
&\quad = \max_{\left\{\substack{\ell_1 \in \mathcal{L}[t_1,t_2] \\ \ell_2 \in \mathcal{L}[t_3,t_4]}\right\}} \rho(\ell_1(x^a), \ell_2(x^a) | \ell(x^a)) \\
&\quad = \max_{\left\{\substack{\hat{\ell}_1 \in \mathcal{L}[at_1,at_2] \\ \hat{\ell}_2 \in \mathcal{L}[at_3,at_4]}\right\}} \rho(\hat{\ell}_1(x), \hat{\ell}_2(x) | \hat{\ell}(x))
\end{aligned}$$

where the last equality follows from $\hat{\ell}(x) = \ell(x^a)$. Therefore,

$$\overline{\rho}(x, [t_1, t_2], [t_3, t_4] | \ell(x)) = \overline{\rho}(x, [at_1, at_2], [at_3, at_4] | \hat{\ell}(x)).$$

The result of the theorem follows. Q.E.D.

REFERENCES

[1] J. R. M. Hosking, "Modeling persistence in hydrological time series using fractional differencing," *Water Resources Res.*, vol. 20, no. 12, pp. 1898–1908, Dec. 1984.

[2] M. S. Keshner, "$1/f$ noise," *Proc. IEEE*, vol. 70, pp. 212–218, Mar. 1982.

[3] W. Willinger, M. Taqqu, W. E. Leland, and D. V. Wilson, "Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements," *Statistical Sci.*, vol. 10, pp. 67–85, 1995.

[4] B. Mandelbrot and J. van Ness, "Fractional Brownian motions, fractional noises, and applications," *SIAM Rev.*, vol. 10, pp. 422–437, 1968.

[5] G. W. Wornell, "Wavelet-based representations for the $1/f$ family of fractal processes," *Proc. IEEE*, vol. 81, pp. 1428–1450, Oct. 1993.

[6] M. F. Barnsley, R. L. Devaney, B. B. Mandelbrot, H.-O. Peitgen, D. Saupe, and R. F. Voss, *The Science of Fractal Images*. New York: Springer-Verlag, 1988.

[7] M. A. Stoksik, R. G. Lane, and D. T. Nguyen, "Practical synthesis of accurate fractal images," *Graphical Models Image Processing*, vol. 57, no. 3, pp. 206–219, May 1995.

[8] J. R. M. Hosking, "Fractional differncing," *Biometrika*, vol. 68, pp. 165–176, 1981.

[9] C. W. J. Granger and J. Joyeux, "An introduction to long memory time series models and fractional differencing," *J. Time Ser. Anal.*, vol. 1, pp. 15–29, 1980.

[10] P. Flandrin, "On the spectrum of fractional Brownian motions," *IEEE Trans. Inform. Theory*, vol. 35, pp. 197–199, Jan. 1989.

[11] ———, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–917, Mar. 1992.

[12] P. Abry and F. Sellan, "The wavelet-based synthesis for fractional brownian motion," *Appl. Computat. Harmonic Anal.*, vol. 3, pp. 377–383, 1996.

[13] K. C. Chou, A. S. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. Automatic Contr.*, vol. 39, pp. 464–478, Mar. 1994.

[14] M. R. Luettgen, W. C. Karl, and A. S. Willsky, "Multiscale smoothing error models," *IEEE Trans. Automatic Contr.*, vol. 40, pp. 173–175, Jan. 1995.

[15] M. R. Luettgen and A. S. Willsky, "Likelihood calculation for a class of multiscale stochastic mdoels with application to texture discrimination," *IEEE Trans. Image Processing*, vol. 4, pp. 194–207, Feb. 1995.

[16] M. M. Daniel and A. S. Willsky, "A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing," *Proc. IEEE*, vol. 85, pp. 164–180, Jan. 1997.

[17] ———, "Modeling and estimation of fractional Brownian motion using multiresolution stochastic processes," in *Proc. 3rd Conf. Fractals Eng.* (Arcachon, France, June 1997).

[18] P. W. Fieguth and A. S. Willsky, "Fractal estimation using models on multiscale trees," *IEEE Trans. Signal. Processing.*, vol. 44, pp. 1297–1300, May 1996.

[19] H. Akaike, "Markovian representations of stochastic processes by canonical variables," *SIAM J. Contr.*, vol. 13, Jan. 1975.

[20] W. W. Irving and A. S. Willsky, "A canonical correlations approach to multiscale stochastic realization," to be published.

[21] R. J. Barton and V. C. Poor, "Signal detection in fractional Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 34, pp. 943–959, Sept. 1988.

[22] L. M. Kaplan and C.-C. J. Kuo, "Fractal estimation from noisy data via discrete fractional Gaussian noise (DFGN) and the Haar basis," *IEEE Trans. Signal Processing*, vol. 41, p. 3554, Dec. 1993.

[23] M. Deriche and A. Tewfik, "Maximum likelihood estimation of the parameters of discrete fractionally differenced Gaussian noise process," *IEEE Trans. Inform. Theory*, vol. 41, pp. 2977–2989, Oct. 1993.

[24] P. W. Fieguth, W. C. Karl, A. S. Willsky, and C. Wunsch, "Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 280–292, Mar. 1995.

[25] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamics systems," *AIAA J.*, vol. 3, no. 8, Aug. 1965.

[26] M. R. Luettgen, W. C. Karl, A. S. Willsky, and R. R. Tenney, "Multiscale representations of Markov random fields," *IEEE Trans. Signal Processing*, vol. 41, p. 3377, Dec. 1993.

[27] M. Reed and B. Simon, *Functional Analysis*. New York: Academic, 1990.

[28] S. Cambanis and C. Houdre, "On the continuous wavelet transform of second-order random processes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 628–642, May, 1995.

[29] H. Krim and J.-C. Pesquet, "Multiresolution analysis of a class of nonstationary processes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1010–1020, July, 1995.