

**January, 1995**

LIDS-P 2287

**Research Supported By:**

ONR N00014-91-J-1004

ARPA F49620-93-1-0604

AFOSR F49620-95-1-0083

**Fractal Estimation using Models on Multiscale Trees**

Fieguth, P.W.

Willsky, A.S.

# Fractal Estimation using Models on Multiscale Trees

Paul W. Fieguth

Alan S. Willsky

## Abstract

In this paper we estimate the fractal dimension of stochastic processes with  $1/f$ -like spectra by applying a recently-introduced multiresolution framework. This framework admits an efficient likelihood function evaluation, allowing us to compute the maximum likelihood estimate of this fractal parameter with relative ease. In addition to yielding results that compare well to other proposed methods, and in contrast to other approaches, our method is directly applicable, with at most very simple modification, in a variety of other contexts including fractal estimation given irregularly sampled data or nonstationary measurement noise and the estimation of fractal parameters for 2-D random fields.

## 1 Introduction

Many natural and human phenomena have been found to possess  $1/f$ -like spectral properties, which has led to considerable study of  $1/f$  processes. One class of such processes that is frequently used because of its analytical convenience and tractability is the class of fractional Brownian motion (fBm) processes, introduced by Mandelbrot and Van Ness[8]. For practical computation purposes we consider only sampled versions of continuous time fBm processes  $B(t)$ , i.e.,

$$B[k] = B(k\Delta t) \quad k \in Z \quad (1)$$

---

\*The authors are with the Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 35-427, 77 Massachusetts Ave., Cambridge MA, 02139. Email: pwfiegut@mit.edu Tel: (617) 263-3816 FAX: (617) 258-8553

Research support provided in part by the Office of Naval Research under Grant N00014-91-J-1004, by the Advanced Research Projects Agency under Grant F49620-93-1-0604, and by the Air Force Office of Scientific Research under Grant F49620-95-1-0083. P.W.F. was also supported by an NSERC-67 fellowship of the Natural Sciences and Engineering Research Council of Canada.

EDICS Number SP 3.5

for which the associated nonstationary covariance is

$$E[B[k], B[m]] = \frac{\sigma^2}{2} (\Delta t)^{2H} (|k|^{2H} + |m|^{2H} - |k - m|^{2H}) \quad (2)$$

where  $\sigma$  and  $H$  are scalar parameters which completely characterize the process.  $H$ , the quantity which we wish to estimate, determines the fractal dimension  $D = 2 - H$  of the process. Previous estimators have been developed addressing this problem, notably those of Wornell and Oppenheim[11], Kaplan and Kuo[4], Tewfik and Deriche[10], and Flandrin[3]. The exact maximum likelihood (ML) calculation for the fractal dimension of fBm is computationally difficult (see [10]); to address this difficulty, fractal estimators typically fall into one of the two following classes to achieve computational efficiency:

1. optimal algorithms, admitting an efficient solution, based on a  $1/f$ -model other than fBm;
2. approximate or suboptimal algorithms developed directly from the fBm model.

Our approach and that of [11] fall into the former category, while the methods in [3, 4, 9] fall into the latter. In particular the approach in [11] is based on a  $1/f$  process constructed using a wavelet basis in which the wavelet coefficients are independent, with variances that vary geometrically with scale with exponent equal to  $H$ . The method in [4] determines the exact statistics of the Haar wavelet coefficients of the discrete fractional Gaussian noise (DFGN) process  $F[k] = B[k+1] - B[k]$  and then develops an estimator by assuming, with some approximation, that the coefficients are uncorrelated.

The goal of our research, on the other hand, is the development of a fast estimator for  $H$  that functions under a broader variety of measurement circumstances, for example the presence of gaps in the measured sequence, measurement noise having a time-varying variance and higher dimensional processes (e.g., 2-D random fields). The basis for accomplishing this is the utilization of a recently-

introduced multiscale estimation framework.[1, 5] The next section gives a brief description of this framework, followed by the development of the estimator, and finally a description of estimation results.

## 2 Multiscale Framework

In the framework developed in [1, 5] stochastic models are constructed recursively in scale or in resolution on multilevel trees. Specifically, let  $s$  index the nodes of a tree  $\mathcal{T}$  (refer to Figure 1) which, for the purposes of this paper, may be considered a dyadic tree although the framework permits much greater flexibility. Let  $s_o \in \mathcal{T}$  designate the root node of  $\mathcal{T}$ ; also let  $s\bar{\gamma}$  denote the parent node of  $s \neq s_o$ . Each node  $s \in \mathcal{T}$  has associated with it a state vector  $x(s)$  and possibly an observation vector  $y(s)$ . Stochastic models on  $\mathcal{T}$  are written recursively in this form

$$x(s) = A(s)x(s\bar{\gamma}) + G(s)w(s) \quad \forall s \in \mathcal{T}, s \neq s_o \quad (3)$$

and where  $w(s)$  is a white Gaussian noise process with identity covariance. Similarly, noisy observations of the process are permitted on an arbitrary subset of the tree nodes:

$$y(s) = C(s)x(s) + v(s) \quad \forall s \in \mathcal{O} \subseteq \mathcal{T} \quad (4)$$

where  $v(s)$  is white and Gaussian with covariance  $R(s)$ . In general, tree variables at all scales may be physically meaningful and measurable, or they may be abstract – a by-product of achieving the desired statistics on the finest level of the tree. In this paper coarse scale nodes are abstract, with a  $1/f$ -like process residing on the finest scale.

For those multiscale stochastic models which can be written in the form (3),(4) the following two problems possess extremely efficient algorithmic realizations[1, 6]:

1. Given observations  $y()$ , determine the optimum least-square estimate for  $x()$

2. Determine the likelihood  $l \left[ A(), G(), C(), R(), y() \right]$  of a set of observations  $y()$

The latter algorithm permits the estimation of any parameter embedded in the multiscale model by maximizing the likelihood function; this is the very problem which we have set out to solve in this paper: by formulating an appropriate multiscale model  $A(s, H), G(s, H), C(s), R(s)$  an estimator for  $H$  may be written abstractly as

$$\hat{H} = \arg_H \max l \left[ A(s, H), G(s, H), C(s), R(s), B[k] \right] \quad (5)$$

As was the case with Wornell and Oppenheim[11], we do not construct an exact model of fBm, but rather choose an appropriate approximation – in our case within this multiscale framework. The selection of such a multiscale model is achieved in the next section.

### 3 Fractal Estimator

The statistical self-similarity of fBm makes the application of wavelets a logical choice. Kaplan and Kuo[4] apply the Haar wavelet to the incremental process  $F[k]$ , and Wornell and Oppenheim[11] apply higher order Daubechies wavelets to  $B[k]$ . We use the multiscale framework just described to develop a Haar wavelet multiscale stochastic model which applies directly to  $B[k]$ . This choice of wavelet is motivated by the particularly simple realization of the Haar wavelet in our multiscale framework by using a dyadic tree structure (see Figure 1):

$$\text{Coarse Scales : } \begin{cases} x(s) = \begin{bmatrix} 1 & +1 \\ 0 & 0 \end{bmatrix} x(s\bar{\gamma}) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} g(s, H)w(s) & \begin{array}{l} s \text{ is left descendant} \\ \text{of its parent} \end{array} \\ x(s) = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} x(s\bar{\gamma}) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} g(s, H)w(s) & \begin{array}{l} s \text{ is right descendant} \\ \text{of its parent} \end{array} \end{cases} \quad (6)$$

$$\text{Finest Scale : } \begin{cases} x(s) = \begin{bmatrix} 1 & +1 \end{bmatrix} x(s\bar{\gamma}) + 0 \cdot w(s) & \begin{array}{l} s \text{ is left descendant} \\ \text{of its parent} \end{array} \\ x(s) = \begin{bmatrix} 1 & -1 \end{bmatrix} x(s\bar{\gamma}) + 0 \cdot w(s) & \begin{array}{l} s \text{ is right descendant} \\ \text{of its parent} \end{array} \\ y(s) = x(s) + v(s) \end{cases} \quad (7)$$

That is, at coarse scales  $x(s)$  consists of two scalars: a coarse approximation to the  $1/f$  process, and a detail coefficient. The detail coefficient equals the difference in the coarse  $1/f$ -like representation between node  $s$  and its two children, where the sign of this difference depends on the parity of the child (i.e., left vs. right). At the finest scale  $x(s)$  is a single scalar, representing a sample of a  $1/f$ -like process, and measurements of the actual fBm sequence appear as observations  $y(s)$  at the finest scale. Note that the  $1/f$  process described by (6),(7) does not yield a finest scale process that is exactly an fBm process (and thus, as with the technique in [11], our model does not exactly match the statistics of the process to be estimated). However by an appropriate choice of the remaining model parameters we can produce a process with the same type of  $1/f$  behavior.

The elements which remain to be determined in the above multiscale model are the  $g(s, H)$ : the variance of the detail wavelet coefficients between node  $s$  and its children. Expressions for the statistics of the wavelet decomposition of fBm have been determined by others [3, 9], however the self-statistics for the special case of the Haar wavelet are easily computed as follows:

- Let  $B_o[k] = B[k]$  which is the fBm process of interest.
- Define  $B_m[k]$  as the process obtained by coarsening  $B[k]$   $m$  times,

$$B_m[k] = (B_{m-1}[2k] + B_{m-1}[2k + 1]) / 2 \quad (8)$$

a relation which follows from the multiscale model of (6).

- From (8) it follows that

$$B_{m-1}[k] = \sum_{i=0}^{2^{m-1}-1} \frac{B[2^{m-1}k + i]}{2^{m-1}} \quad (9)$$

$$B_{m-1}[k + 1] - B_{m-1}[k] = \sum_{i=0}^{2(2^{m-1}-1)} F[2^{m-1}k + i] \frac{2^{m-1} - |2^{m-1} - i - 1|}{2^{m-1}} \quad (10)$$

$$\equiv \sum_{i=0}^{2(2^{m-1}-1)} F[2^{m-1}k + i] c_i \quad (11)$$

- From the stationarity of the increments process  $F[k]$ , and from the symmetry

$$B_m[k] - B_{m-1}[2k] = -\{B_m[k] - B_{m-1}[2k+1]\} \quad (12)$$

(11) reduces to the desired variance expression

$$E[(B_m[k] - B_{m-1}[2k])^2] = \frac{1}{4} \sum_{i=-2(2^{m-1}-1)}^{2(2^{m-1}-1)} \lambda_F[i] \sum_{j=-\min(0,i)}^{2(2^{m-1}-1)+\min(0,-i)} c_i c_{i+j} = g^2(s, H) \quad (13)$$

where  $s$  is any node on scale  $m$  of the tree, and where  $\lambda_F$  is the covariance function of  $F[k]$ :

$$\lambda_F[i] = \frac{\sigma^2}{2} [|i+1|^{2H} + |i-1|^{2H} - 2|i|^{2H}] \quad (14)$$

By way of comparison, it has been shown[11] that  $1/f$  processes, of which fBm is a subset, may be approximated by wavelet synthesis in which the wavelet coefficients vary exponentially with scale:

$$g_m^2 = \beta 2^{-2mH} \quad \text{i.e., } \log_2 \frac{g_m}{g_{m-1}} = H \quad (15)$$

Table 1 shows the scale to scale variance ratios as predicted by (13). The deviation from the approximate scaling law of (15) is most pronounced at low  $H$ ; it is this deviation which leads to a bias for those estimators based on (15), as shown in Table 2.

The actual estimator for  $H$ , based on the multiscale model of (6),(7), takes precisely the form as outlined in (5), in which the likelihood maximization is performed using standard nonlinear techniques (e.g., the section search method of MATLAB).

## 4 Experimental Results

Sixty four fBm sample paths, each having a length of 2048 samples, were generated using the Cholesky decomposition method of [7], precisely the same approach as in Kaplan and Kuo[4] whose experimental results form the basis of comparison with ours.

The performance of three fBm estimators is compared in Table 3. The bias in the estimator of [11] for low  $H$ , as was argued earlier based on Table 1, is evident. Also recall that the multiscale model of (6),(7) assumed the wavelet detail coefficients to be uncorrelated; this assumption becomes progressively poorer as  $H$  increases[3], leading to an increase in the error variance for our estimator at large  $H$ . Nevertheless, our method still performs reasonably well over quite a wide range of values of  $H$ . Moreover, using the techniques developed in [5], we can construct higher-order multiscale models which account for most of the residual correlation in the wavelet coefficients. The same likelihood procedure applied to these higher-order models would then yield even better results, closely approaching the exact ML estimator based on the exact fBm statistics. However since fBm itself is an idealization, the benefit in practice of such higher-order models over that based on the low-order model (6),(7) depends upon the application.

Finally, as we have said, our approach applies equally well in a variety of other settings. For example, Table 4 illustrates the performance of our estimator under non-uniform sampling (by randomly discarding 10% of the measurements), and under a varying measurement noise variance. Both of these special cases are accomplished with essentially no change in the algorithm. In addition, by using a quadtree rather than a dyadic tree we can also apply these techniques in 2-D. An example of such an application to the estimation of non-isotropic fractal parameters for a 2-D random field based on irregular, nonstationary data is given in [2].

## References

- [1] K. Chou, A. Willsky, A. Benveniste, "Multiscale Recursive Estimation, Data Fusion, and Regularization", *IEEE Trans. on Automatic Control* (39) #3, pp.464–478, 1994
- [2] P. Fieguth, W. Karl, A. Willsky, C. Wunsch, "Multiresolution Optimal Interpolation and



Statistical Analysis of TOPEX/POSEIDON Satellite Altimetry”, *IEEE Trans. Geoscience and Remote Sensing*, to appear

- [3] P. Flandrin, “Wavelet Analysis and Synthesis of Fractional Brownian Motion”, *IEEE Trans. Information Theory* (38) #2, pp.910–917, 1992
- [4] L. Kaplan, C.C. Kuo, “Fractal Estimation from Noisy Data via Discrete Fractional Gaussian Noise (DFGN) and the Haar Basis”, *IEEE Trans. S.P.* (41) #12, pp.3554–3562, 1993
- [5] M. Luetttgen, W. Karl, A. Willsky, R. Tenney, “Multiscale Representations of Markov Random Fields”, *IEEE Trans. Signal Processing* (41) #12, pp.3377–3396, 1993
- [6] M. Luetttgen and A. Willsky, “Likelihood Calculation for a Class of Multiscale Stochastic Models, with Application to Texture Discrimination.”, *IEEE Trans. I.P.* (4) #2, MIT, 1994
- [7] T. Lundahl, W. Ohley, S. Kay, R. Siffert, “Fractional Brownian motion: A maximum likelihood estimator and its application to image texture”, *IEEE Trans. Med. Im.* (5), pp.152–161, 1986
- [8] B. Mandelbrot, J. van Ness, “Fractional Brownian motions, fractional noises and applications”, *SIAM Review* (10), pp.422–437, 1968
- [9] A. Tewfik, M. Kim, “Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motion”, *IEEE Trans. Information Theory* (38) #2, pp.904–909, 1992
- [10] A. Tewfik, M. Deriche, “Maximum Likelihood Estimation of the Parameters of Discrete Fractionally Differenced Gaussian Noise Processes”, *IEEE Trans. Signal Processing* (41) #10, pp.2977–2989, 1993
- [11] G. Wornell, A. Oppenheim, “Estimation of fractal signals from noisy measurements using wavelets”, *IEEE Trans. Signal Processing* (40), pp.611–623, 1992

## List of Captions:

Table 1: Haar wavelet coefficient standard-deviation ratios as a function of  $H$  and scale:  $g_i^2$  represents the wavelet coefficient variance at scale  $i$ , where the finest scale is  $i = 0$ . The deviation of the variance progression from an exponential law is most pronounced at fine scales and for low values of  $H$ .

Table 2: The estimator in the top row is based on the premise that wavelet coefficient variances are exponentially distributed, leading to a biased estimate  $\hat{H}$ .

Table 3: Estimation results for three estimators, based on 64 fBm sample paths, each of length 2048 samples, with no additive noise. The experimental results for the first two estimators are from [4].

Table 4: Performance of the fBm estimator for two examples: irregular sampling (removal, at random, of 10% of the measurements), and nonstationary measurement noise variance (noise standard deviation  $\sigma[k] = \frac{1}{2} \exp\{ -[(k - 1024)/500]^2 \}$ ). In both cases the results are based on 64 fBm sample paths, each of length 2048 samples.

Fig. 1: Dyadic tree structure used for the estimator of this paper.

Variance Ratio:	$H = 0.25$	$H = 0.50$	$H = 0.75$	$H = 0.9$
$\log_2(g_8/g_7)$	0.250			
$\log_2(g_7/g_6)$	0.249	$\vdots$		
$\log_2(g_6/g_5)$	0.247	0.500	$\vdots$	
$\log_2(g_5/g_4)$	0.242	0.499	0.750	$\vdots$
$\log_2(g_4/g_3)$	0.228	0.496	0.749	0.900
$\log_2(g_3/g_2)$	0.188	0.484	0.745	0.898
$\log_2(g_2/g_1)$	0.091	0.437	0.727	0.892
$\log_2(g_1/g_0)$	-0.084	0.292	0.650	0.861

Table 1:

Variance Rule:		$H = 0.25$	$H = 0.50$	$H = 0.75$	$H = 0.90$
Variances assumed exponential with scale	$\hat{H}$ :	0.05	0.40	0.70	0.91
Variances based on exact result	$\hat{H}$ :	0.24	0.51	0.75	0.92

Table 2:

Estimator		$H = 0.25$	$H = 0.50$	$H = 0.75$	$H = 0.90$
W.O.	$\hat{H}$	0.082	0.398	0.683	0.846
	$\sigma_{\hat{H}}$	0.022	0.021	0.026	0.021
	$(H - \hat{H})_{\text{RMS}}$	0.169	0.109	0.072	0.058
K.K.	$\hat{H}$	0.252	0.499	0.748	0.899
	$\sigma_{\hat{H}}$	0.017	0.017	0.017	0.017
	$(H - \hat{H})_{\text{RMS}}$	0.017	0.017	0.017	0.017
Multiscale Haar	$\hat{H}$	0.249	0.503	0.768	0.919
	$\sigma_{\hat{H}}$	0.011	0.019	0.050	0.109
	$(H - \hat{H})_{\text{RMS}}$	0.011	0.019	0.054	0.110

Table 3:

Circumstance		$H = 0.25$	$H = 0.50$	$H = 0.75$	$H = 0.90$
Irregular Sampling	$\hat{H}$	0.246	0.507	0.781	0.937
	$\sigma_{\hat{H}}$	0.033	0.044	0.076	0.124
	$(H - \hat{H})_{\text{RMS}}$	0.033	0.045	0.082	0.128
Nonstationary Measurement Noise Variance	$\hat{H}$	0.268	0.511	0.769	0.918
	$\sigma_{\hat{H}}$	0.011	0.019	0.051	0.109
	$(H - \hat{H})_{\text{RMS}}$	0.021	0.022	0.054	0.110

Table 4:

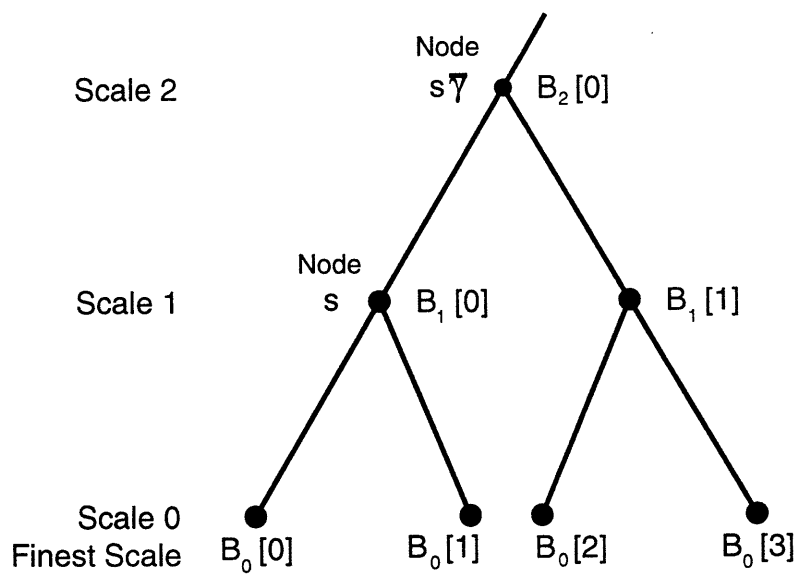


Figure 1: