Modeling Algorithms for a Class of Fractal Signals

by

Warren Michael Lam

B.S. University of California, Berkeley (1989)

Submitted to the Department of Electrical Engineering and
Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 1992

Author . . . .
Department of Electrical Engineering and Computer Science
June 5, 1992

Certified by. . . . .
Gregory W. Wornell
Assistant Professor
Thesis Supervisor

Accepted by . . . .
Campbell L. Searle
Chairman, Departmental Committee on Graduate Students

Modeling Algorithms for a Class of Fractal Signals

by

Warren Michael Lam

Submitted to the Department of Electrical Engineering and Computer Science
on June 5, 1992, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

In this work, we introduce a generalized class of fractal signals, a useful extension of the $1/f$ signal modeling framework. While the $1/f$ signal model plays an important role in modeling a wide variety of physical signals, we demonstrate that these generalized fractal signals allow the capturing of an even richer set of phenomena. Just as it provides a natural environment for the processing of $1/f$ signals, the wavelet transform is shown to be well-suited for the analysis of these generalized fractal signals. Exploiting this property of the transform, we formulate efficient wavelet-based algorithms for addressing a number of classification and estimation problems involving such signals. Results of simulations are presented to show the viability of these algorithms.

Thesis Supervisor: Gregory W. Wornell
Title: Assistant Professor

Acknowledgments

First, I wish to express my deepest gratitude to my thesis supervisor, Professor Gregory Wornell, for his guidance and support, without which this work would have been impossible.

I also wish to thank Helen Meng who painstakingly read and critiqued various drafts of this thesis.

Most of all, I thank my parents and my brother, Walter, for their love, patience, and encouragement.

Contents

List of Figures

Chapter 1

Introduction

The $1/f$ family of fractal signals has become increasingly important in modeling a variety of physical signals. Their characteristic *fractional*-power-law spectral behavior offers great versatility for modeling many types of structures and textures. In addition, the large amount of memory inherent in these $1/f$ signals is well-suited for capturing evolutionary and developmental time series associated with a host of natural and man-made phenomena. Texture variation in natural terrain, landscapes and cloud formations, average seasonal temperature, annual amount of rainfall, heartbeat arrival rate, economic data, and rate of traffic flow are among the numerous phenomena exhibiting $1/f$ behavior [6] [21]. Motivated by their ubiquity, much work has been devoted to the study of $1/f$ signals. Recently, computationally efficient wavelet-based algorithms have been developed for solving a number of signal processing problems involving such signals. While the $1/f$ signal model has received much attention, a number of useful generalizations of this model have not been as actively pursued. In this work, we focus on the development of a class of generalizations of the $1/f$ signal model.

Our interest in generalizations of the $1/f$ signal model is mainly motivated by the existence of physical phenomena which, though not well represented in the $1/f$ modeling framework, do possess $1/f$ type quality. As reflected by the model proposed by Goff and Jordan [4], the seafloor morphology is one such phenomenon. While observations of the seafloor structure over short length scales reveal $1/f$ behavior,

9

the application of a $1/f$ model for this modeling purpose is in conflict with the overall flatness of ocean bottom at long length scales. As another example, in their description of the structures of diluted gels and colloidal aggregates, Schaefer *et al* suggest a model which exhibit different types of $1/f$ behavior over different scales to account for the interplay of mass-fractals and surface-fractals [3]. The generalized fractal signal model we develop in this work makes possible the modeling of these signals.

1.1   Outline of the Thesis

This thesis is structured around the development of a class of generalized fractal signals. To show the desirability of this signal class for modeling, we shall mainly address two issues concerning such signals. We shall first demonstrate that this signal class offers much versatility for capturing many phenomena. We shall then show that efficient schemes exist for processing these generalized fractal signals. Before we delve into these issues, however, we shall first introduce our main tool in this work, the wavelet transform. As will be clear, the wavelet transform provides a naturally well-suited environment for the representation and processing of these generalized fractal signals. Because of the novelty of this transformation and its importance to this work, Chapter 2 is devoted to the discussion of this transform. The main purpose of this chapter is to convey to the reader a set of interpretations of the transform which will offer insight into seeing the relation between this transform and fractal signals. Also presented in this chapter is the architecture of the Discrete Wavelet Transform, a computationally efficient mechanism which we use throughout this work for actually computing the dyadic wavelet transform.

Chapter 3 introduces and describes the class of generalized fractal signals which we shall study in this work. In this chapter, we see how the notion of these generalized fractal signals arises from the $1/f$ signal model. The key result of the chapter is a synthesis relation of the generalized fractal signals using the dyadic wavelet transform. Through this, it becomes clear that the dyadic wavelet basis functions form

an approximate Karhunen-Loève basis for such generalized fractals. This result thus substantiates the importance of the wavelet transform in the processing of generalized fractals.

Chapter 4 gives a collection of algorithms for solving several signal processing problems involving generalized fractal signals. The problems which we focus on are mainly modeling problems including a parameter estimation and a signal separation problems. The algorithms given in this chapter exploit the role of the dyadic wavelet basis functions as an approximate Karhunen-Loève basis for fractal signals, and much of the processing takes place in the wavelet transform domain.

Chapter 2

Wavelet Transformations

A mathematical tool which we shall use extensively throughout this work is the wavelet transform. While our interest in this transform is primarily due to the naturally well-suited environment it provides for representing and analyzing fractal signals, the recent advent of the Discrete Wavelet Transform, an efficient algorithm for its computation, helps make the wavelet transform stand out as the ideal tool for this work.

Although the wavelet transform recently emerged as a novel idea involving representations of functions using *dilations* and *translations* of a prototype function, it is closely related to a number of well-established notions in signal processing. An objective of this chapter is to convey to the reader a set of signal processing interpretations of this transform which will provide insight for understanding the desirability of the wavelet transform for processing fractal signals. In this chapter, we also outline a derivation of the Discrete Wavelet Transform, a computationally efficient mechanism for computing the wavelet transform. Since we focus mainly on aspects of the wavelet theory of relevance to this work, the treatment in this chapter is in no way comprehensive. For more detailed coverage of the theory of wavelet transformations, sources [1], [9], [15], and [17] are highly recommended.

2.1 The Continuous Wavelet Transform

The *continuous wavelet transform* (CWT) is formally defined as

$$X_\nu^\mu = \mathcal{W}\{x(t)\} \triangleq \langle x(t), \psi_\nu^\mu(t) \rangle = \int_{-\infty}^{\infty} x(t)\, \psi_\nu^\mu(t)\, dt, \qquad \mu, \nu \in \mathbf{R},\ \mu \neq 0, \quad (2.1)$$

where the functions $\psi_\nu^\mu(t)$ are obtained by dilating and translating a *basic wavelet* $\psi(t)$:

$$\psi_\nu^\mu(t) \triangleq |\mu|^{-1/2} \psi\left(\frac{t - \nu}{\mu}\right).$$

The parameters μ and ν are known as the *scale* and *translation factors*, respectively. If the basic wavelet $\psi(t)$ is *admissible*, i.e., if its Fourier transform $\Psi(\Omega)$ satisfies

$$\int_{-\infty}^{\infty} |\Psi(\Omega)|^2\, |\Omega|^{-1}\, d\Omega = C_\psi < \infty, \tag{2.2}$$

then an *inverse continuous wavelet transform* (ICWT) can be defined and is given by

$$x(t) = \mathcal{W}^{-1}\{X_\nu^\mu\} \triangleq \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_\nu^\mu\, \psi_\nu^\mu(t)\, \mu^{-2}\, d\nu\, d\mu. \tag{2.3}$$

When defined, the ICWT can be viewed as a reconstruction of the original signal using scaled and translated versions of the basic wavelet, with the CWT X_ν^μ acting as the weighting factors. The CWT, therefore, gives the scale contents of a signal, just as the Fourier transform gives its frequency contents.

Since μ and ν are continuous variables, it is clear that the set $\{\psi_\nu^\mu(t);\ \mu, \nu \in \mathbf{R},\ \mu \neq 0\}$ lacks orthogonality regardless of the choice of the wavelet. Consequently, the CWT carries highly redundant information, and the reconstruction given in (2.3) is very inefficient. Fortunately, under certain general circumstances, we only need a subset of the CWT points to reconstruct the original signal. Indeed, it can be shown

13

that if for some positive constants A, B, a_0, and b_0,

$$A\langle x(t),\ x(t)\rangle < \sum_{m,n} |\langle x(t),\ \psi(a_0^{-m}t - nb_0)\rangle|^2 < B\langle x(t),\ x(t)\rangle$$

for all $x(t) \in L^2(\mathrm{R})$ (the set of real-valued finite-energy functions), then any $L^2(\mathrm{R})$ function $x(t)$ is completely specified by the CWT points $\{X_\nu^\mu;\ \mu = a_0^m, \nu = nb_0a_0^m\}$, where m and n range over the set of integers [1]. Such a sufficient set of CWT points is generally referred to as a *frame*. In this work, we focus mainly on the frame corresponding to $a_0 = 2$, $b_0 = 1$, and are particularly interested in the case where the associated transform is orthonormal. We shall refer to this transform as the *dyadic wavelet transform*.

2.2   The Dyadic Wavelet Transform

The *dyadic wavelet transform* can be interpreted as a representation of functions using a complete orthonormal basis made up of dilations and translations of a basic wavelet $\psi(t)$. Formally, the transform pair is defined as

$$x_n^m = \mathcal{W}_d\{x(t)\} \triangleq \int_{-\infty}^{\infty} x(t)\ \psi_n^m(t)\,dt, \qquad m,n \in \mathrm{Z}, \tag{2.4}$$

$$x(t) = \mathcal{W}_d^{-1}\{x_n^m\} \triangleq \sum_m \sum_n x_n^m \psi_n^m(t), \tag{2.5}$$

where the basis functions are defined as

$$\psi_n^m(t) \triangleq 2^{m/2}\psi(2^m t - n). \tag{2.6}$$

The parameters m and n are referred to as the *scale* and *translation indices*, respectively.

To relate this transform to more familiar ideas in signal processing, we first consider the Fourier transform of a typical wavelet. It can be shown that if $\psi(t)$ decays sufficiently fast at infinity, the admissibility condition given in (2.2) coincides with

14

the condition

$$\Psi(0) = \int_{-\infty}^{\infty} \psi(t) \, dt = 0. \tag{2.7}$$

Thus, a basic wavelet cannot have lowpass frequency behavior. Next, the orthonormality of the set $\{\psi_n^m(t); \; m, n \in Z\}$ requires that

$$|\Psi(\Omega)| \le 1, \qquad \text{for all } \Omega. \tag{2.8}$$

In addition to these two requirements, a *regularity* constraint is often imposed on the design of the basic wavelet to yield smooth basis functions. A function $\psi(t)$ is said to be *kth-order regular* if for $j = 0, 1, \ldots, k - 1$,

$$\int_{-\infty}^{\infty} t^j \, \psi(t) \, dt = 0. \tag{2.9}$$

Hence (2.7) states that an admissible wavelet is at least first-order regular. Although wavelets having very general frequency behavior do exist, in this work we focus only on those having bandpass frequency response. Figure 2-1 shows the frequency response of two typical wavelets, the Haar wavelet and the Daubechies wavelet. Note that they both have a passband of roughly $[\pi, 2\pi]$.

Rewriting (2.4) as

$$x_n^m = \left\{ x(t) * \psi_0^m(-t) \right\}\big|_{t=2^{-m} n},$$

we see that the dyadic wavelet transform is equivalent to the filter bank shown in Figure 2-2(a). The sequence x_n^m associated with each scale m can be viewed as the result of the filtering of $x(t)$ with $\psi_0^m(-t)$ followed by a sampling operation. Since $\psi(t)$ has a passband of approximately $[\pi, 2\pi]$, it follows from (2.6) that $\psi_0^m(-t)$ approximately passes the frequency range $[2^m \pi, 2^{m+1} \pi]$. Figure 2-2(b) depicts the approximate passband of $\psi_0^m(-t)$ for several values of m. This diagram shows that the dyadic wavelet transform is in fact a particular type of constant-Q or octave-band

15
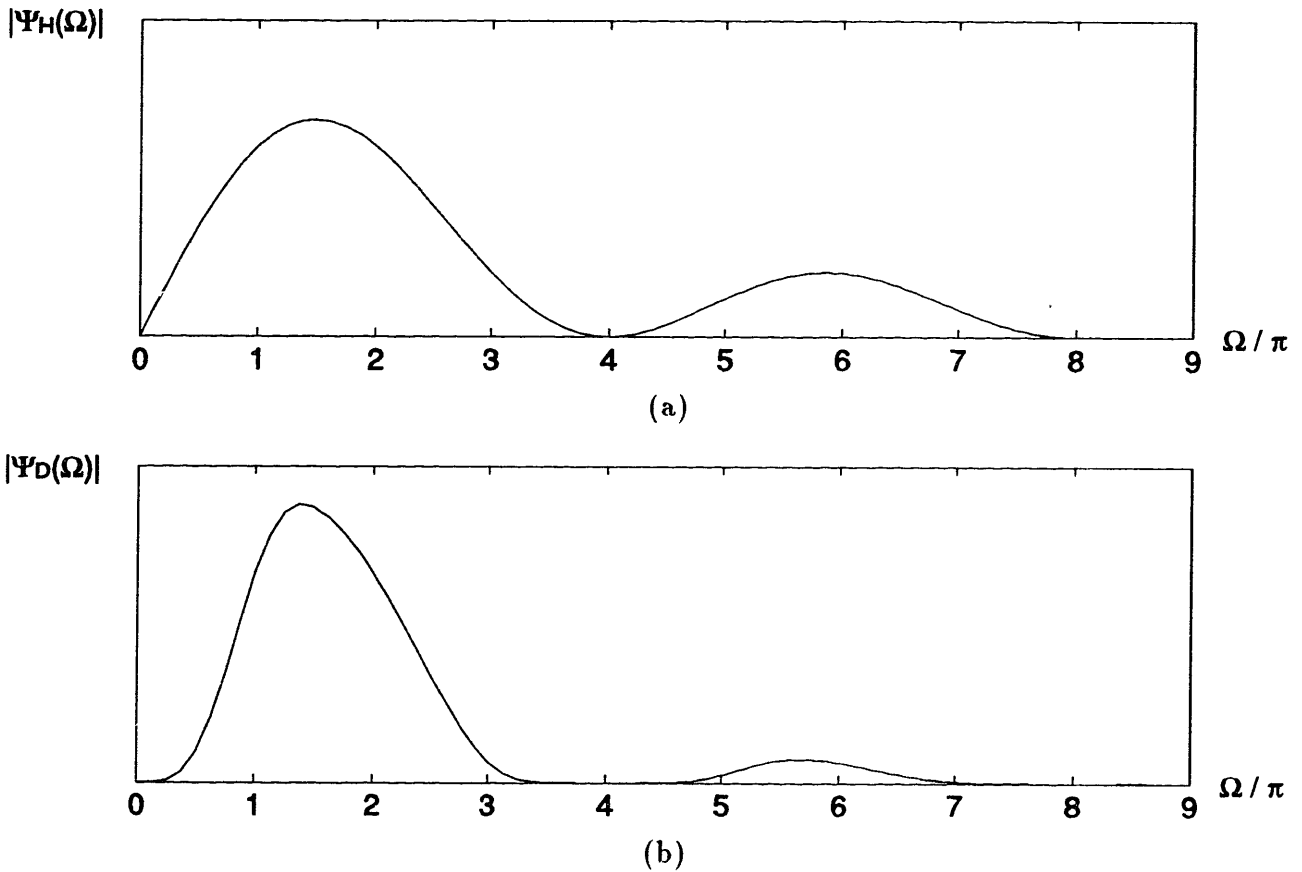
Figure 2-1: *Frequency response of two typical wavelets: (a) the Haar wavelet; (b) the 4th-order Daubechies wavelet*

filter bank, and can be viewed as a uniform partitioning of the frequency axis on a logarithmic scale.

Similarly, the inverse dyadic wavelet transform can be interpreted as the superposition of signals derived by modulating the functions $\psi_0^m(t)$ with the sequences x_n^m, and is hence equivalent to the system shown in Figure 2-3. Here, "D/I" denotes a system that converts a discrete-time sequence to a weighted impulse train, operating at a duty cycle of T.
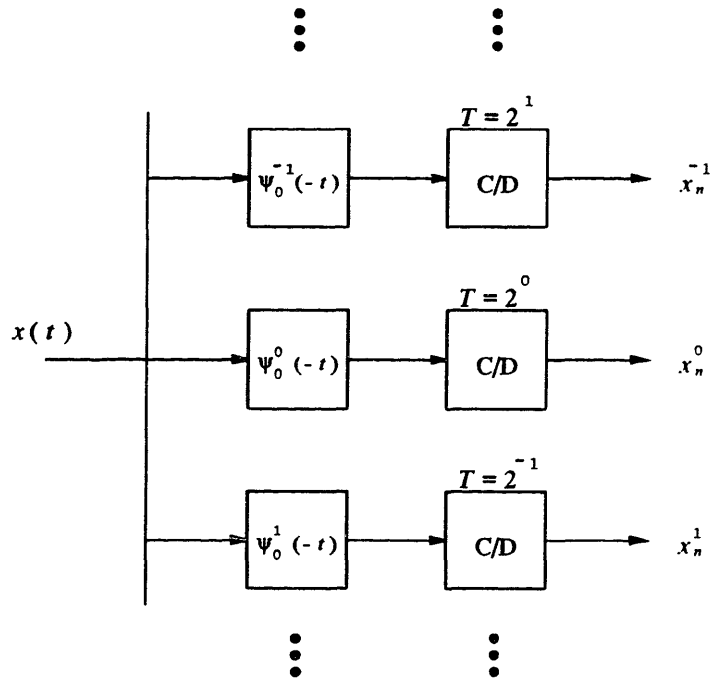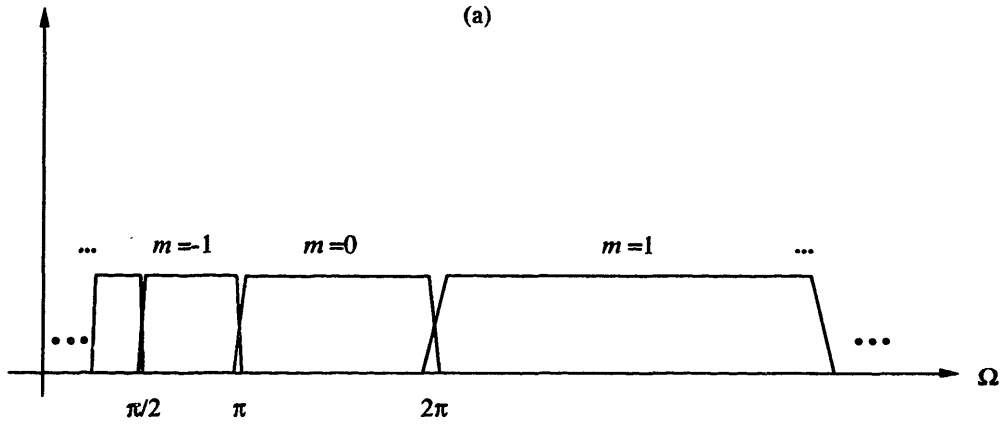
Figure 2-2: *Octave-band-filter-bank interpretation of the dyadic wavelet transform:* (a) *an octave-band filter bank;* (b) *approximate passband of each branch*
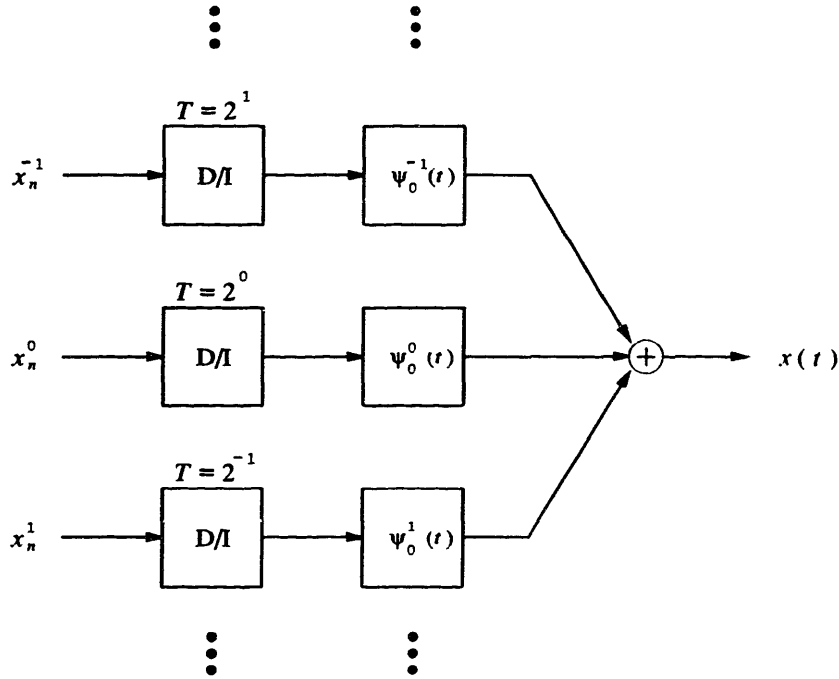
Figure 2-3: *Multirate-modulation interpretation of the inverse dyadic wavelet transform*

2.3 Multiresolution Analysis Interpretation of the Dyadic Wavelet Transform

The theory of multiresolution analysis offers an alternative perspective for viewing the dyadic wavelet transform. More importantly, as we shall see in this section, it leads naturally to an efficient implementation of the transform.

Multiresolution analysis basically involves the approximation of functions using basis functions of *approximation subspaces* at various resolutions. In this context, the wavelet transform is directly constructed from the difference between approximations at successive resolutions. The set of approximation subspaces is denoted by $\{V_m; \ m \in Z\}$, with increasing m corresponding to finer resolutions. To ensure that approximations at any resolution contain the information needed for generating

approximations at coarser resolutions, we require that

$$V_m \subset V_{m+1} \qquad \text{for all } m \in \mathbf{Z}.$$

This is referred to as the *causality property*. The orthogonal complement of V_m in V_{m+1} is denoted by O_m. More specifically, we have that

$$O_m \perp V_m, \qquad \text{and}$$

$$O_m \oplus V_m = V_{m+1}.$$

The *approximation* of $x(t)$ at resolution m is defined as the least-squares-error approximation of $x(t)$ in V_m, and is denoted by $A_m x(t)$. The difference between the approximations at resolutions $m+1$ and m, or the *detail* at resolution m, is denoted by $D_m x(t)$. Note that $A_m x(t) \in V_m$ and $D_m x(t) \in O_m$. Note also that

$$
\begin{aligned}
A_m x(t) &= A_{m-1} x(t) + D_{m-1} x(t) \\
&= A_{m-2} x(t) + D_{m-2} x(t) + D_{m-1} x(t) = \cdots = \sum_{i=-\infty}^{m-1} D_i x(t).
\end{aligned}
$$
(2.10)

To ensure that all $L^2(\mathbf{R})$ functions are well represented at fine resolutions, we require that the union of the approximation subspaces be dense in $L^2(\mathbf{R})$, i.e.,

$$\overline{\bigcup_{m=-\infty}^{\infty} V_m} = L^2(\mathbf{R}).$$

This leads to

$$\lim_{m \to \infty} A_m x(t) = x(t),$$
(2.11)

for all $x(t) \in L^2(\mathbf{R})$.

In addition to the causality property, [9] gives a list of conditions on $\{V_m; m \in \mathbf{Z}\}$ which ensure that V_m and O_m have orthonormal bases with very elegant properties.

If $\{V_m;\ m \in \mathbf{Z}\}$ are chosen accordingly, then for each m, V_m is spanned by the orthonormal basis $\{\phi_n^m(t);\ n \in \mathbf{Z}\}$, where

$$\phi_n^m(t) = 2^{m/2}\phi(2^m t - n),$$

and, for each m, O_m has the orthonormal basis $\{\psi_n^m(t);\ n \in \mathbf{Z}\}$, where

$$\psi_n^m(t) = 2^{m/2}\psi(2^m t - n).$$

The functions $\phi(t)$ and $\psi(t)$ are referred to as the *scaling function* and *wavelet*, respectively. While $\psi(t)$ has the properties given in (2.7)–(2.9) and is roughly bandpass, $\phi(t)$ is roughly lowpass in nature:

$$|\Phi(0)| = 1, \tag{2.12}$$

$$|\Phi(\Omega)| \leq 1, \qquad \text{for all } \Omega, \tag{2.13}$$

and

$$\Phi(\Omega) \sim \mathcal{O}(\Omega^{-1}), \qquad \text{as } \Omega \to \infty. \tag{2.14}$$

Since $\{\phi_n^m(t);\ n \in \mathbf{Z}\}$ is an orthonormal basis for the function space to which $A_m x(t)$ belongs, $A_m x(t)$ can be written as

$$A_m x(t) = \sum_n a_n^m \phi_n^m(t),$$

with a_n^m obtained via inner products:

$$a_n^m = \langle x(t),\ \phi_n^m(t) \rangle. \tag{2.15}$$

Also, using (2.10) and the fact that $D_m x(t) \in O_m$, $A_m x(t)$ can be expressed in terms

20

of $\psi_n^m(t)$ as

$$A_m x(t) = \sum_{i=-\infty}^{m-1} \sum_n x_n^i \psi_n^i(t) \tag{2.16}$$

where x_n^m are obtained via

$$x_n^m = \langle x(t),\ \psi_n^m(t) \rangle, \tag{2.17}$$

Passing (2.16) to the limit and using (2.11), we obtain

$$x(t) = \lim_{m \to \infty} A_m x(t) = \sum_{m=-\infty}^{\infty} \sum_n x_n^m \psi_n^m(t). \tag{2.18}$$

We recognize (2.17) and (2.18) as the dyadic wavelet transform pair.

Exploiting the causality property, we can express a_n^m and x_n^m in terms of a_n^{m+1}, and vice versa. Since for each m and n, $\phi_n^m(t)$ and $\psi_n^m(t)$ reside in subspaces of V_{m+1}, $\phi(t)$ and $\psi(t)$ satisfy the two-scale equations

$$\phi_n^m(t) = \sum_{k=-\infty}^{\infty} \langle \phi_n^m(t),\ \phi_k^{m+1}(t) \rangle \phi_k^{m+1}(t) = \sum_{k=-\infty}^{\infty} \langle \phi_n^{-1}(t),\ \phi_k^0(t) \rangle \phi_k^{m+1}(t) \tag{2.19}$$

$$\psi_n^m(t) = \sum_{k=-\infty}^{\infty} \langle \psi_n^m(t),\ \phi_k^{m+1}(t) \rangle \phi_k^{m+1}(t) = \sum_{k=-\infty}^{\infty} \langle \psi_n^{-1}(t),\ \phi_k^0(t) \rangle \phi_k^{m+1}(t). \tag{2.20}$$

The last equality in these equations follow from a change of variables. From (2.19) and (2.20), we derive the relations

$$a_n^m = \sum_{k=-\infty}^{\infty} h[2n - k] a_n^{m+1} \tag{2.21}$$

$$x_n^m = \sum_{k=-\infty}^{\infty} g[2n - k] a_n^{m+1} \tag{2.22}$$

(a)   (b)
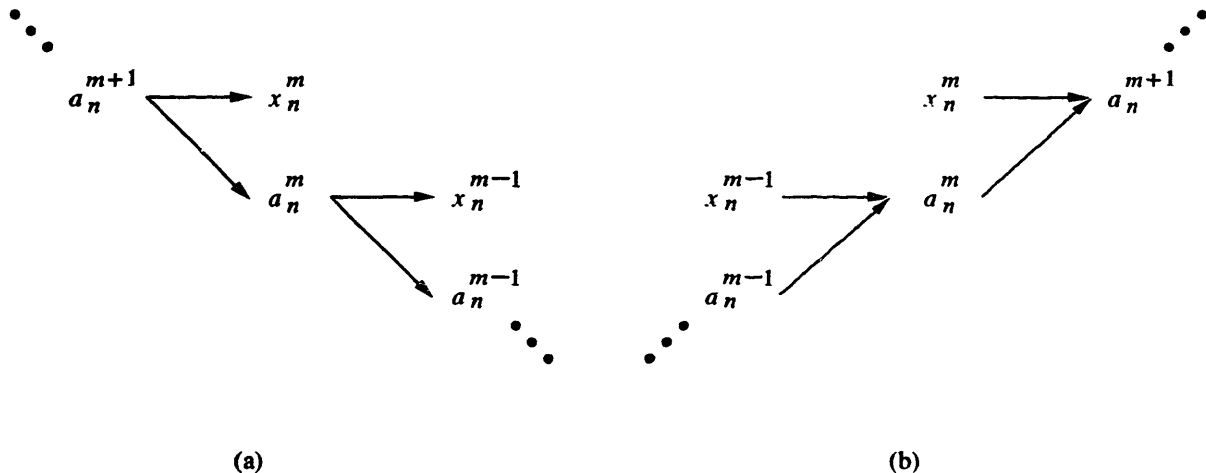
Figure 2-4: *Multiresolution scheme for computing the dyadic wavelet transform pair: (a) computation of the dyadic wavelet transform; (b) computation of the inverse dyadic wavelet transform*

and

$$a_n^{m+1} = \sum_{k=-\infty}^{\infty} h[2k - n]a_n^m + \sum_{k=-\infty}^{\infty} g[2k - n]x_n^m, \qquad (2.23)$$

where

$$h[n] = \langle \phi_n^{-1}(t), \phi_0^0(t) \rangle, \qquad (2.24)$$

and

$$g[n] = \langle \psi_n^{-1}(t), \phi_0^0(t) \rangle. \qquad (2.25)$$

Applying the operation on the right hand side of (2.21) and (2.22) to a_n^m, we can obtain a_n^{m-1} and x_n^{m-1}. Given an approximation a_n^m, therefore, we can compute the dyadic wavelet transform at scales coarser than m using the hierarchical scheme shown in Figure 2-4.

It can be shown that the sequences $h[n]$ and $g[n]$ given in (2.24) and (2.25) form a

conjuate quadrature discrete-time filter pair and that their Fourier transforms satisfy

$$G(\omega) = e^{-j\omega}H^*(\omega + \pi) \qquad (2.26)$$

and

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 = 1. \qquad (2.27)$$

More remarkably, although we have derived $h[n]$ and $g[n]$ from $\phi(t)$ and $\psi(t)$, given a conjugate quadrature discrete-time filter pair, we can conversely determine the scaling function and wavelet function of the associated multiresolution signal analysis. More precisely, in the frequency domain, we have the recurrent relations

$$\Phi(\Omega) = H(\Omega/2)\Phi(\Omega/2) \qquad (2.28)$$

and

$$\Psi(\Omega) = G(\Omega/2)\Phi(\Omega/2), \qquad (2.29)$$

which can be used for obtaining $\phi(t)$ and $\psi(t)$, respectively.

2.4 Implementation of the Dyadic Wavelet Transform

In this research, we employ the *discrete wavelet transform* (DWT) for the actual computation of the dyadic wavelet transform. Just as the Fast Fourier Transform (FFT) is an efficient algorithm for computing a discretized Fourier Transform of a discretized signal, the DWT is an efficient mechanism for computing the dyadic wavelet transform of a discretized signal. The idea of the DWT arises very naturally from the results of the theory of multiresolution analysis, and is, in fact, an efficient implementation of discrete-time computations (2.21)–(2.23). The architecture of the

DWT and the IDWT is shown in Figures 2-5 and 2-6, respectively. Note that their structures parallel the scheme shown in Figure 2-4. The filters used in this architecture are related to the filters given in the previous section via

$$h_0[n] = h[n]; \qquad h_1[n] = h[-n];$$
$$g_0[n] = g[n]; \qquad g_1[n] = g[-n].$$

Several issues concerning the computation of the dyadic wavelet transform of physical signals are noteworthy. First, because the DWT is an operator on discrete-time sequences, we cannot apply it directly to a continuous-time signal. Instead, we must first derive the approximation of the signal at the finest scale (using a C/D conversion, for instance), and then apply the DWT to the approximation.

Another issue has to do with the computation of the dyadic wavelet transform of physical signals in general. Because in practice, we can only process signals of finite duration, we are limited to a finite number of scales in the dyadic wavelet transform calculations. This can be easily seen from the architecture of the DWT. Indeed, a finite duration signal gives rise to a finite length sequence as the input of the DWT. Due to the downsampling operation at each stage of the DWT, the number of samples decreases by roughly a factor of 2 per scale. Thus, it follows that we can only obtain a finite set of scales. By the constant-Q filter bank interpretation of the dyadic wavelet transform, this scale limitation is equivalent to a limitation to a finite frequency range of the signal. As will become clear in Chapter 4, many of our wavelet-based algorithms for processing generalized fractals are affected by this limitation in the number of scales.
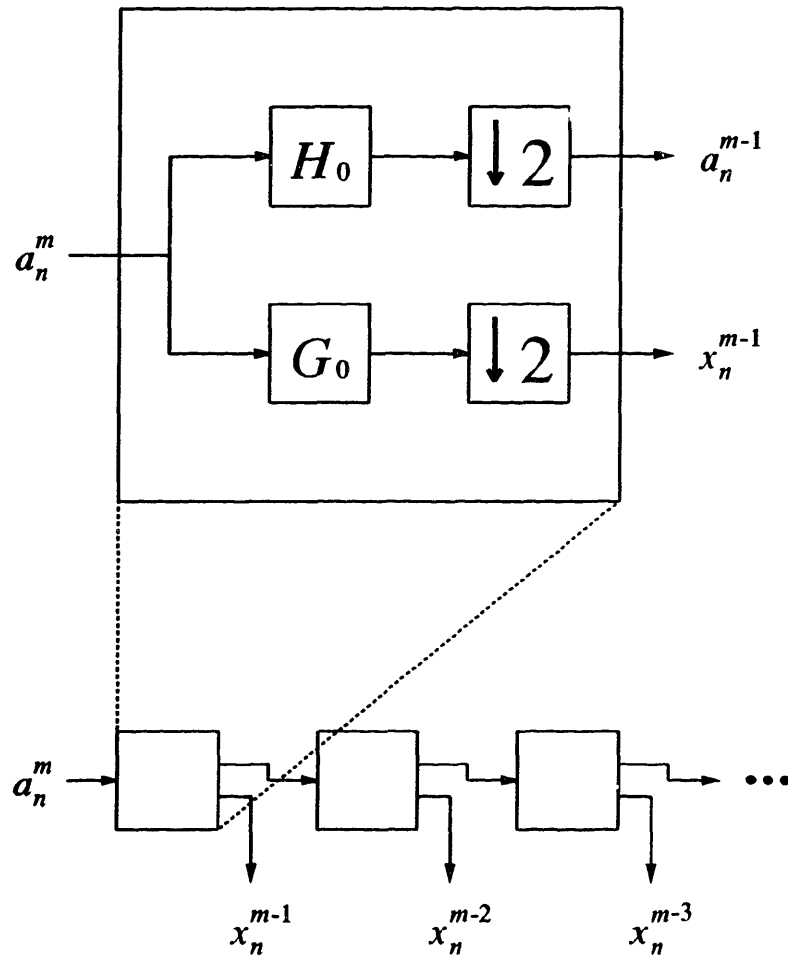
Figure 2-5: *Discrete wavelet transform implementation of the dyadic wavelet transform*
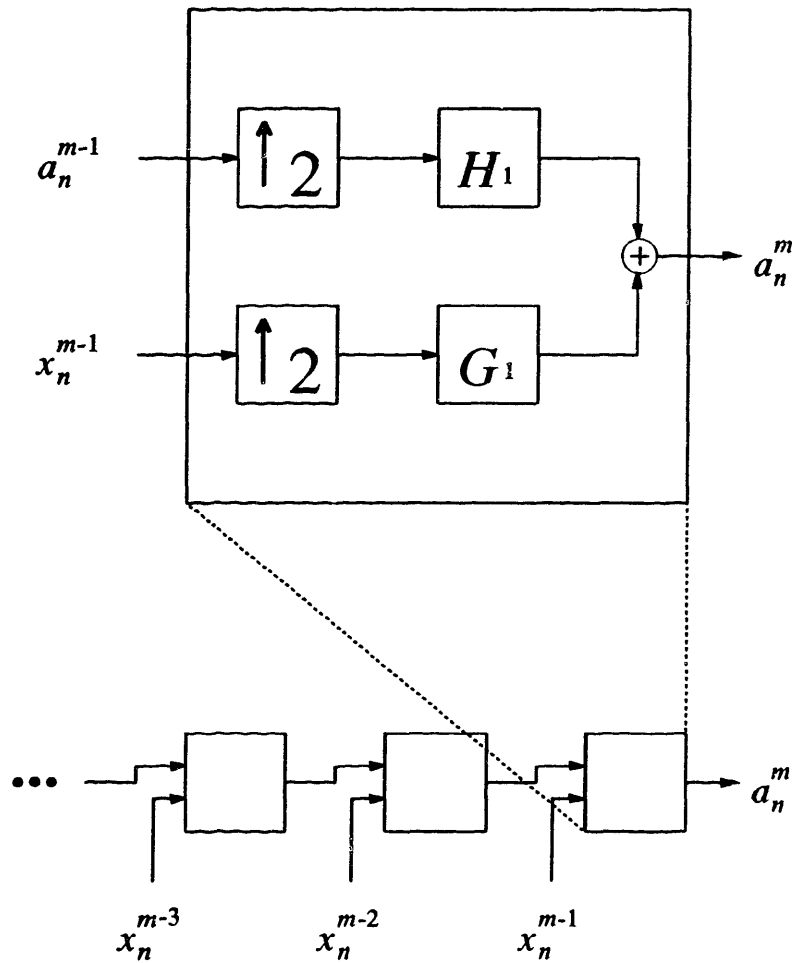
Figure 2-6: *Discrete wavelet transform implementation of the inverse dyadic wavelet transform*

Chapter 3

A Generalized Fractal Signal Model

In this chapter, we introduce a class of generalized fractal signals which are developed based on the notion of the $1/f$ signals. In the first half of this chapter, we briefly review the theory of $1/f$ signals. Through this discussion, we establish that although the fractional-power-law spectral behavior and the persistent correlation property of the $1/f$ signal class allow the modeling of an already rich class of phenomena, reasonable extensions of the $1/f$ signal model make possible the capturing of an even broader set of physical signals. Motivated by this, we focus on the development of a class of generalizations of the $1/f$ signals in the second half of the chapter. We begin this development by considering a prototype generalized fractal model which we refer to as the first-order generalized fractal. By showing that the dyadic wavelet basis functions constitute as an approximate Karhunen-Loève basis for these first-order generalized fractals, we motivate that just as it serves as the natural tool for processing $1/f$ signals, the wavelet transform plays a central role in the processing of more generalized fractal signals. Finally, using first-order generalized fractals as building blocks, we develop generalized fractals of higher complexity.

3.1   The $1/f$ Signal Model

A $1/f$ random process is characterized by its fractional-power-law power spectral density

$$S(\Omega) \sim \frac{\sigma^2}{|\Omega|^\gamma}. \tag{3.1}$$

The parameter γ is a real constant typically found between 0 and 2, although it does lie outside this range in some occasions. It is clear that when plotted on a logarithmic scale, the power spectrum given in (3.1) yields a straight line with slope $-\gamma$.

A well-known property of the $1/f$ signal family is their *self-similar* or *fractal* behavior, which roughly pertains to the scale-invariance of these signals. We present here an informal plausibility argument which allows us to relate this self-similar behavior with the frequency-domain specification given in (3.1). To understand the relation between scaled versions of a signal, we first note that as a random process is dilated or compressed, its autocorrelation is dilated or compressed in precisely the same way. Further, invoking the Fourier transform relation

$$f(t) \stackrel{\mathcal{F}}{\longleftrightarrow} F(\Omega) \qquad \Rightarrow \qquad f\left(\frac{t}{a}\right) \stackrel{\mathcal{F}}{\longleftrightarrow} |a| F(|a|\Omega),$$

we see that as a random process is dilated in the time domain (i.e., $a > 1$), the log-log frequency plot of its power spectrum is first shifted to the left by $\log a$, and then shifted up by $\log a$. Similarly, a compression of the signal results in a right shift followed by a downward shift of the log-log frequency plot of the power spectrum. Figure 3-1 shows such a transformation on a $1/f$ power spectrum for the case where the dilation factor $a > 1$. Because the log-log frequency plot of this power spectrum is a straight line, it is obvious that, on a logarithmic scale, the transformed spectrum is always a vertically displaced version of the original spectrum. Thus, the autocorrelation functions of a $1/f$ process $x(t)$ and a scaled version of itself, $x(t/a)$, are related by

$$E\left\{x\left(\frac{t+\tau}{a}\right)x\left(\frac{\tau}{a}\right)\right\} \propto E\{x(t+\tau)x(\tau))\}. \tag{3.2}$$

28
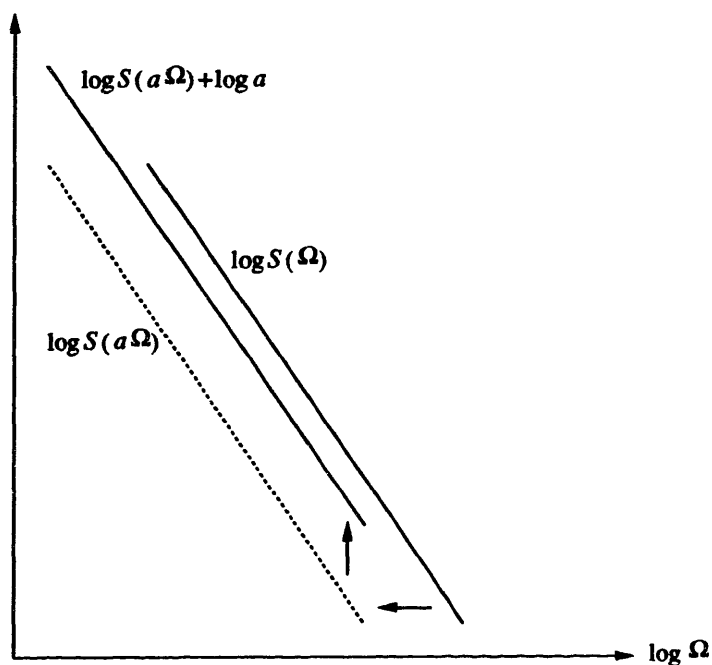
Figure 3-1: *Transformation of* $1/f$ *power spectrum corresponding to scaling in time domain*

This implies that, up to a multiplicative factor, the processes have the same second-order statistics.

The parameter γ governs the ratio between the two autocorrelations in (3.2), and plays the key role of characterizing the self-similar behavior. In fact, it can be shown that γ is related to the fractal dimension of the $1/f$ process via

$$D = E + \frac{3 - \gamma}{2}$$

[10]. Here, D and E are, respectively, the fractal dimension and the Euclidean dimension of the $1/f$ process. The Euclidean dimension pertains to the usual notion of dimension: a curve has a Euclidean dimension of 1, a surface has a Euclidean dimension of 2, and so on. Unlike the Euclidean dimension, the fractal dimension need not be an integer. A fractal curve, in general, can have fractal dimension between 1 and 2, with larger values correspond to rougher curves. Because of its relation with

the fractal dimension D, the parameter γ can be interpreted as a measure of the roughness of the signal. Figure 3-2 shows a few computer-generated sample functions of $1/f$ processes having different values of γ. As can be seen from these plots, the roughness of the functions is indeed dependent on γ. That γ is not restricted to being an integer, then, allows the modeling of a rich set of structures and textures as $1/f$ signals.

Another characteristic property of a $1/f$ process is its strong dependence on its distant history. Indeed, by applying the inverse Fourier transform to the power spectral density given in (3.1), it can be shown that, for suitable values of γ, the autocorrelation function of a $1/f$ process decays at a polynomial rate which is much slower than the exponential decay rate associated with the autocorrelation functions of the familiar ARMA processes. This persistent correlation, or equivalently, large memory property makes the $1/f$ signal class an ideal candidate for modeling many developmental and evolutionary systems.

While the $1/f$ signal family is characterized by its self-similar behavior, we realize that a much broader class of signals also exhibit such behavior. This is a consequence of the typical inability to observe the entire frequency contents of a physical signal. A low-frequency cutoff arises because of the finite data length, and a high-frequency cutoff exists due to limited sampling resolution. With this in mind, let us consider the piecewise linear log power spectrum shown in Figure 3-3. Applying the same type of plausibility argument as before, we can obtain the power spectrum of a dilation of the process by merely shifting the plot. However, in this case, it is clear that strictly speaking, the process is not self-similar; in fact, for any choice of a, the new power spectrum is *not* a vertically shifted version of the old one. Nevertheless, for an observer who is limited to frequency range I, the process appears self-similar over a range of scales. For the dilation corresponding to the transformation shown in Figure 3-3(a), for example, the transformed power spectrum differs from the original spectrum by the same amount over the observed frequencies. Thus, the observer is led to the conclusion that the process is self-similar. As the process is further dilated to give the power spectral density $S'(\Omega)$ shown in Figure 3-3(b), a second type of
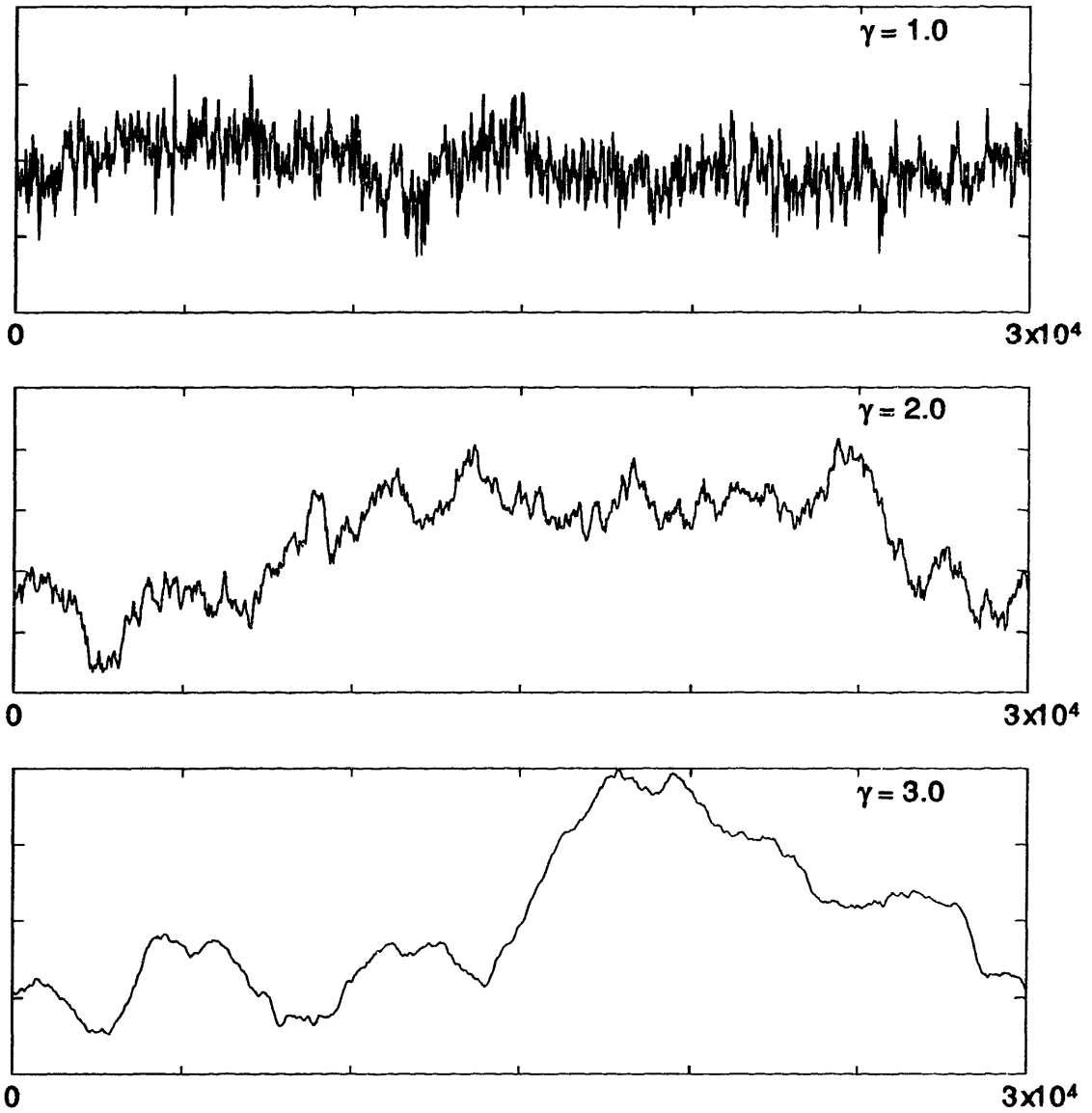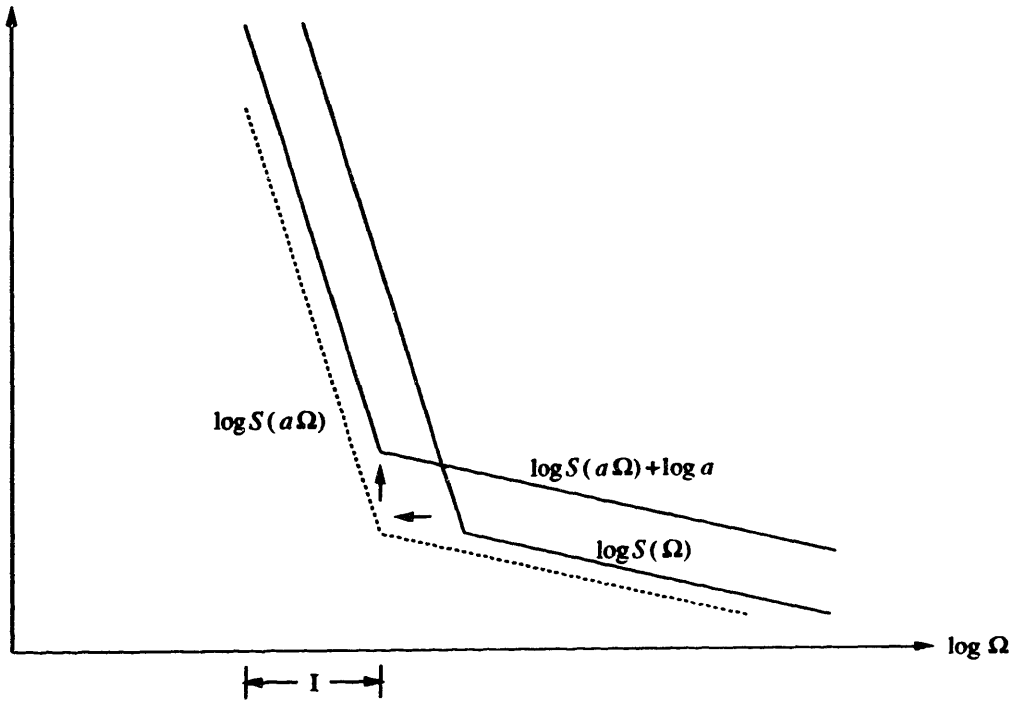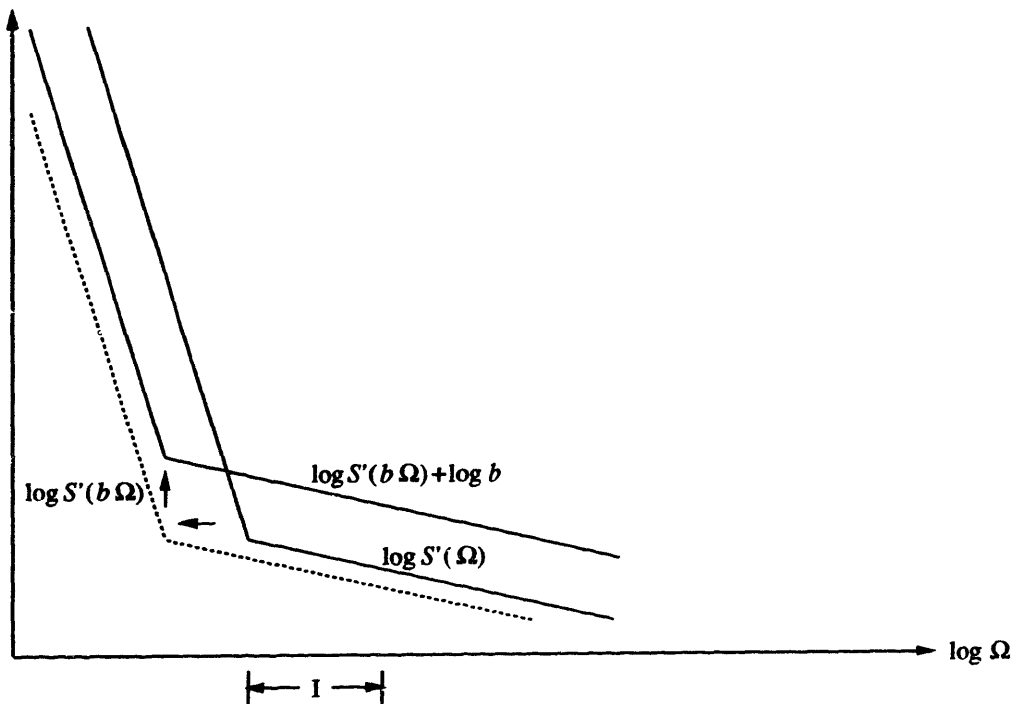
Figure 3-2: *Examples of* $1/f$ *signals*

Figure 3-3: *Behavior of power spectrum upon scaling in time domain*

self-similarity is observed due to a different slope in the log-log plot, or effectively a different γ, in the observed frequency regime. From this example, we see that because of the limitations in the frequencies observed, local self-similar or fractal behavior can arise if the log-log plot of the power spectral density is linear in certain frequency regimes. Moreover, the quality of local fractal behavior is governed by the slope of the plot in these regimes. A $1/f$ signal, then, can be thought of as a special case in which the local fractal behavior is the same over all scales.

That the $1/f$ model exhibits the same type of self-similarity at all scales greatly limits the kinds of signals it can describe. To see this modeling shortcoming, we consider the two computer-generated coastlines given in Figures 3-4 and 3-5. In each case, the coastline is viewed at four different scales. Starting from the second plot on, each plot is a dilated version of the middle tenth of the previous plot (i.e., the portion enclosed by the dashed lines) viewed on a 10 times finer scale. While the $1/f$ model can only generate coastlines with the same type of roughness at all scales, a model allowing variation of fractal behavior over frequencies can give rise to more general coastlines. This capability of modeling varying roughness over scales is crucial for the capturing of, for instance, the type of varying fractal behavior described by Schaefer et al observations of diluted gels and colloidal aggregates [3]. In such cases, the existence of mass-fractals and surface-fractals gives rise to two different types of local fractal behavior observed at two different scale regimes.

Recall that other than its uniform roughness at all scales, the $1/f$ signal model is also characterized by a persistent autocorrelation function, or equivalently, strong correlation over large scales. While this property is central in the modeling of many slow-varying time series, it is quite unrealistic for a number of phenomena. Indeed, in their modeling of seafloor morphology, Goff and Jordan propose a model which exhibits $1/f$ behavior over small scales but is white at large scales [4]. The power spectral density corresponding to their model, thus, has a log-log plot which is flat at low frequencies and rolls off linearly at high frequencies. This example suggests that a more realistic model should allow points very far apart to be uncorrelated, if appropriate.
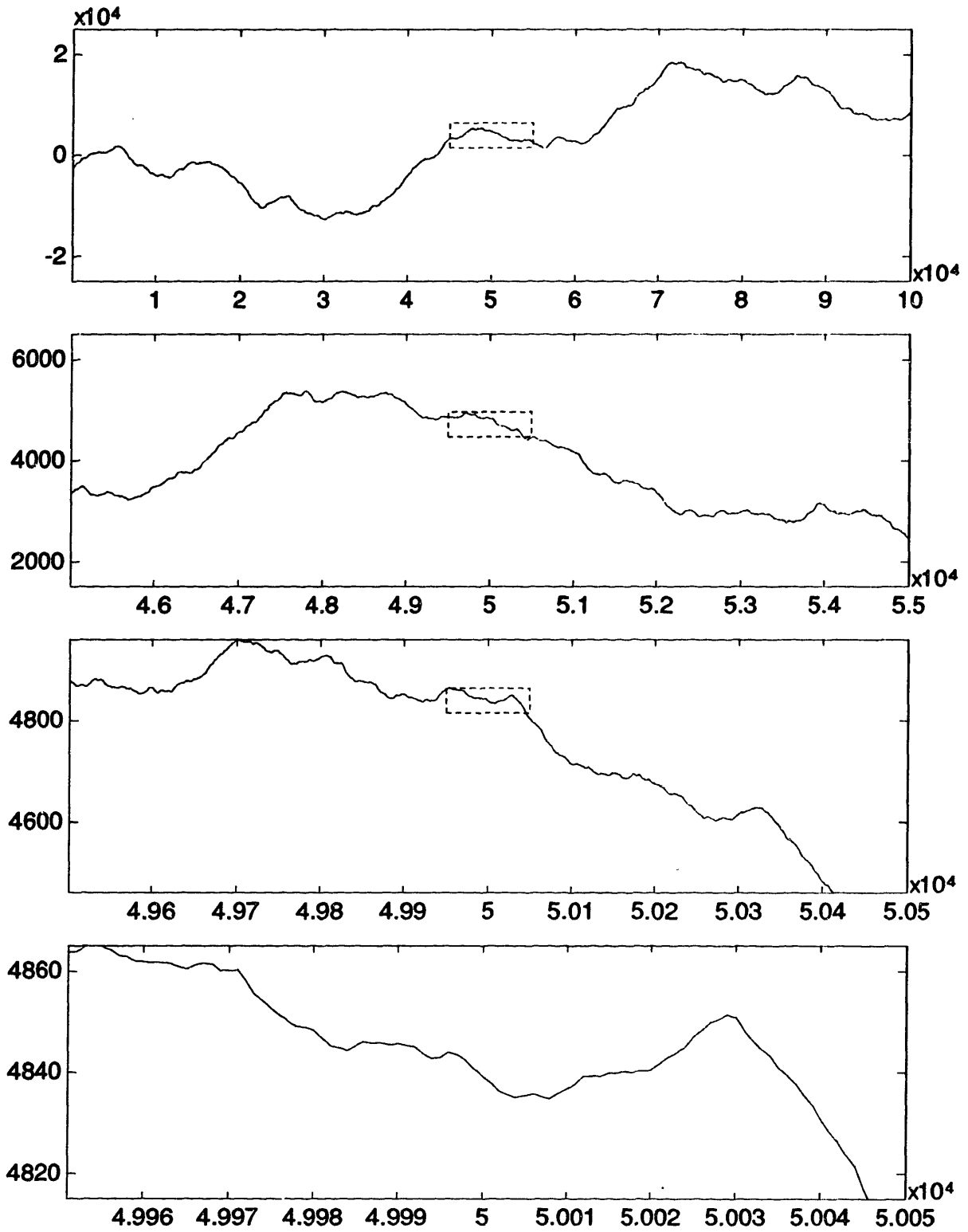
Figure 3-4: *A 1/f coastline viewed at various scales*
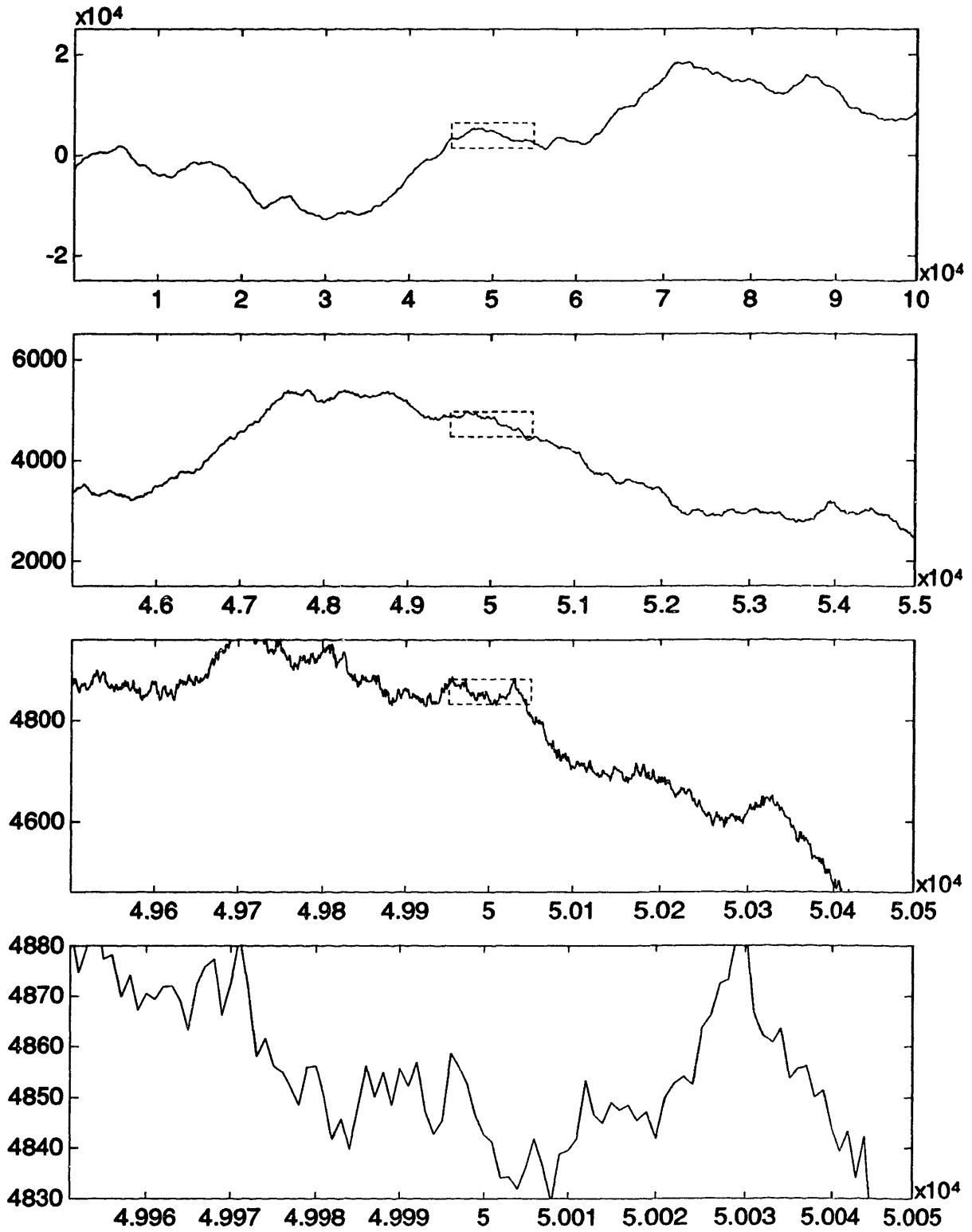
Figure 3-5: *A generalized fractal coastline viewed at various scales*

To summarize, although the $1/f$ signal model is desirable for modeling many signals in nature, it is inadequate for some modeling applications. In our development of a more generalized signal model, we shall keep in mind two types of behavior that our new model should be capable of modeling:

1. It should allow variations in self-similar behavior over scale regimes.

2. It should allow points far apart to be less correlated, if appropriate.

3.2   First-Order Generalized Fractal Signals

We begin our discussion of more generalized fractal signals by considering a family of random processes whose power spectral densities are of the form

$$S(\Omega) \sim \frac{\sigma^2}{|\Omega|^\gamma + p},\qquad (3.3)$$

where γ is a real constant as in (3.1), and p is a real, non-negative constant. This class of processes deserves much attention because it plays a prototype in the development of a much richer class of signals. Throughout, we shall refer to this signal model as the first-order generalized fractal signal model.

It should be clear that the definition of this model is motivated mainly by models like the previously mentioned seafloor morphology model proposed by Goff and Jordan. Indeed, the spectral behavior of a first-order generalized fractal corresponds very closely to the spectral description of the seafloor model. At high frequencies, or fine scales, the log-log plot of the power spectrum prescribed by (3.3) decays linearly with a slope of $-\gamma$ and the process behaves like a $1/f$ process. On the other hand, at low frequencies, or coarse scales, the log-log plot is roughly flat, indicating that points far apart are essentially uncorrelated. The two types of spectral behavior are separated by a corner frequency of approximately $\Omega = p^{1/\gamma}$. Note that by adding the variable p to the $1/f$ spectral specification, we have already achieved one of the two goals given at the closing of Section 3.1.

Several special cases of this first-order generalized fractal signal model have already

been studied, and should be quite familiar. When $p = 0$, for instance, (3.3) reduces to

$$S(\Omega) \sim \frac{\sigma^2}{|\Omega|^\gamma},$$

which corresponds to the power spectrum of a $1/f$ signal. Thus, the class of first-order generalized fractal signals is a superset of the $1/f$ signal family. Next, when $\gamma = 0$, (3.3) becomes

$$S(\Omega) \sim \frac{\sigma^2}{1 + p}.$$

This implies that stationary white processes are also among the collection of first-order generalized fractals. Finally, when $\gamma = 2$, (3.3) corresponds to the power spectrum of the first-order autoregressive (AR) model.

Because of its potential lack of integrability, the expression given in (3.3) may not be a valid power spectral density of a stationary process. In general, therefore, technical difficulties arise in the definition of first-order generalized fractal processes as stationary processes. In order to formally characterize such processes, we adopt the scheme used in [21] for characterizing $1/f$ processes. By this scheme, we consider only processes which give rise to stationary processes when filtered by ideal bandpass filters of the form

$$B_{l,h}(\Omega) = \begin{cases} 1 & \Omega_l < |\Omega| \le \Omega_h \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

where $0 < \Omega_l < \Omega_h < \infty$. This class of signals is made up of numerous signals, some of which are already familiar to us. For example, included is the set of stationary signals. Also, from [21], the family of $1/f$ signals is included in this set.

Based on this scheme, a first-order generalized fractal is characterized as follows: A continuous-time random process $x(t)$ is called a first-order generalized fractal if for some non-negative real constants σ^2, γ, and p, $x(t)$ gives rise to a stationary process

$x_{l,h}(t)$ having power spectral density of the form

$$S_{l,h}(\Omega) = \begin{cases} \dfrac{\sigma^2}{|\Omega|^\gamma + p} & \Omega_l < |\Omega| \le \Omega_h \\ 0 & \text{otherwise,} \end{cases} \qquad (3.5)$$

when it is filtered by an arbitrary ideal bandpass filter $B_{l,h}(\Omega)$ as specified in (3.4).

So, intuitively speaking, we say that a signal is a first-order generalized fractal if the power spectrum derived from any finite resolution measurement of the signal is in accord with a function of the form given in (3.3). Note that this characterization is generally sufficient since we are typically limited to a finite frequency range when processing physical signals. More importantly, this characterization allows us to exploit tools from the context of stationary signal processing such as the Fourier transform and various wavelet transforms for processing first-order generalized fractals.

3.3   Wavelet-Based Canonical Form of First-Order Generalized Fractals

The following theorem shows a close relation between first-order generalized fractals and orthogonal wavelets which will be heavily exploited in this work.

Theorem 3.1 *Suppose that for each m, x_n^m is a wide-sense stationary, zero-mean, white Gaussian sequence with variance*

$$\text{Var } x_n^m = \frac{\sigma^2}{2^{\gamma m} + p}, \qquad (3.6)$$

where $\gamma \ge 0$. Suppose also that for $m \ne m'$, x_n^m and $x_n^{m'}$ are uncorrelated. Define

$$x(t) = \sum_m \sum_n x_n^m \psi_n^m(t), \qquad (3.7)$$

where $\psi_n^m(t)$ is as defined in (2.6), and $\psi(t)$ is a wavelet associated with a multiresolution signal analysis whose scaling function $\phi(t)$ has a Fourier Transform $\Phi(\Omega)$ which

38

10^6

10^3

10^0

10^{-3}

10^{-6}

10^{-9}

10^{-12}

10^{-5} 10^{-2} 10^1 10^4 10^7 10^{10} 10^{13} 10^{16}
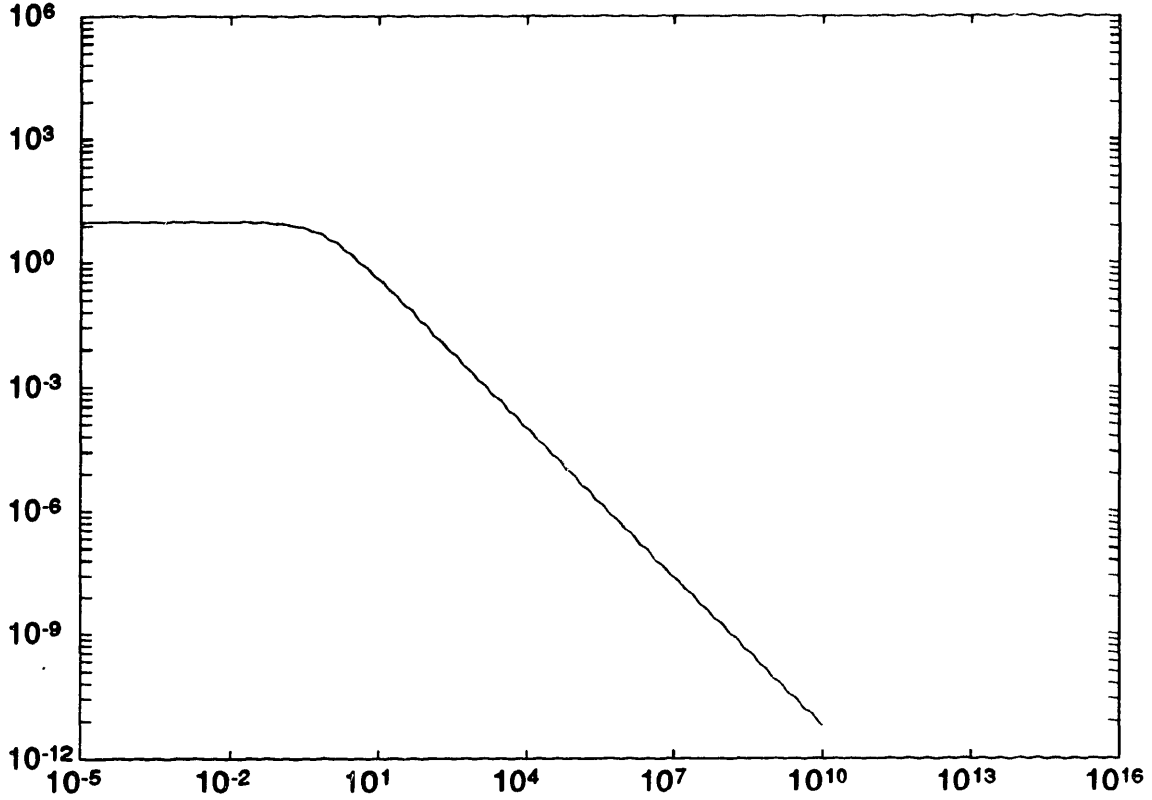
Figure 3-6: *Karhunen-Loève-like expansion of first-order generalized fractals using the Battle-Lemarié wavelet:* $\gamma = 1.2, p = .1$

satisfies (2.12), (2.13) and (2.14), and is continuous at $\Omega = 0$. Then,

$$0 < \frac{k_1}{|\Omega|^\gamma + p} \leq S_{xx}(\Omega) \leq \frac{k_2}{|\Omega|^\gamma + p} < \infty \tag{3.8}$$

for some k_1, k_2.

A proof of this theorem is given in Appendix A. Relation (3.8) suggests that $x(t)$ is a first-order generalized fractal. Indeed, as can be seen in Figure 3-6, which shows the synthesis of a first-order generalized fractal using the Battle-Lemarié wavelet, the shape of the synthesized spectrum has the quality of the spectrum of a first-order generalized fractal described earlier in section 3.2. Due to the orthonormality of the wavelet basis functions $\psi_n^m(t)$, and the lack of correlation of the set of random variables $\{x_n^m; m \in Z, n \in Z\}$, the synthesis scheme given in (3.7) is a Karhunen-Loève expansion. Theorem 3.1, therefore, suggests that the wavelet basis functions provide

an approximate Karhunen-Loève basis for first-order generalized fractals. An issue worth mentioning at this point is that to achieve tight bounds in (3.8), values different from that of the parameter p found in (3.6) are generally needed for the parameters p in the bounding values. This implies a modeling inaccuracy in the synthesis relation and that the value of the p used in the wavelet synthesis (i.e., found in (3.6)) is in general different from the parameter p found in the power spectral density of the resulting first-order generalized fractal.

Theorem 3.1 suggests a canonical form for conveniently viewing first-order generalized fractals. This is illustrated in Figure 3-7 and can be thought of as a wavelet-based synthesis filter for generating first-order generalized fractal signals. The input to the filter, $w(t)$, is white, with zero mean and unit variance. Because the wavelet basis functions form a complete orthonormal basis, the wavelet coefficients w_n^m extracted from the white input $w(t)$ are a set of uncorrelated zero-mean unit-variance random variables. Upon weighting each of these random variables by the factor

$$\sigma_m = \left[\frac{\sigma^2}{2^{\gamma m} + p} \right]^{1/2},$$

we obtain a set of uncorrelated random variables x_n^m with variances as given in (3.6). Theorem 3.1 then states that $x(t)$, the inverse dyadic wavelet transform of x_n^m, is approximately a first-order generalized fractal.

3.4 Higher-Order Generalized Fractal Signals

We now turn to the development of more complex generalized fractal signal models. In general, we shall use first-order generalized fractals discussed in the previous section as the basic building blocks for constructing such higher-order generalized fractals. Although there are several ways of constructing such higher-order generalized fractal signals, we shall see that they all lead to models capable of capturing variations in fractal spectral behavior over scales.

A way of constructing higher-order generalized fractal signals is illustrated in Fig-
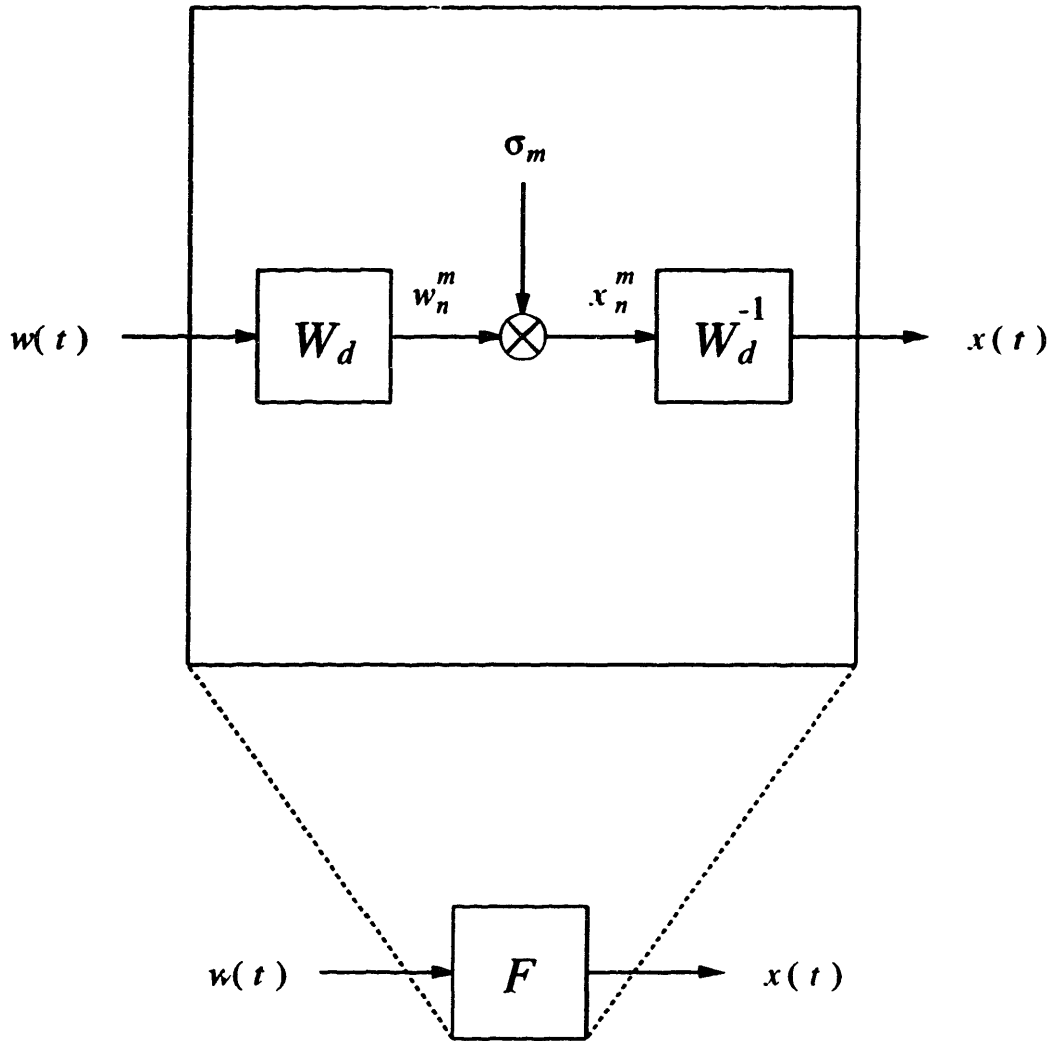
Figure 3-7: *Wavelet-based system for the generation of first-order generalized fractal processes*
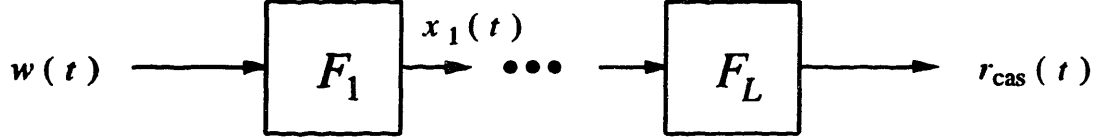
Figure 3-8: *Cascade arrangement of synthesis filters for first-order generalized fractals*

ure 3-8. By cascading synthesis filters for first-order generalized fractals, $F_1, \ldots, F_L$, we obtain a synthesis filter for a process whose power spectral density is the product of the power spectral densities associated with the filters $F_1, \ldots, F_L$. More specifically, the power spectral density of $r_{cas}(t)$ in Figure 3-8, $S_{cas}(\Omega)$, is given by

$$S_{cas}(\Omega) = \prod_{i=1}^{L} S_i(\Omega), \tag{3.9}$$

where $S_i(\Omega)$ is the power spectral density of the first-order generalized fractal process associated with the synthesis filter F_i. Using

$$S_i(\Omega) = \frac{\sigma_i^2}{|\Omega|^{\gamma_i} + p_i} \tag{3.10}$$

in (3.9), we obtain

$$
\begin{aligned}
S_{cas}(\Omega) &= \prod_{i=1}^{L} \frac{\sigma_i^2}{|\Omega|^{\gamma_i} + p_i} \\
&= \frac{K}{(|\Omega|^{\gamma_1} + p_1) \cdots (|\Omega|^{\gamma_L} + p_L)},
\end{aligned}
\tag{3.11}
$$

where

$$K = \prod_{i=1}^{L} \sigma_i^2.$$

It is clear from (3.11) that the corner frequency of each $S_i(\Omega)$ is also a corner frequency of the power spectrum of $S_{cas}(\Omega)$. Using this cascade form, therefore, we can build models which exhibit up to L types of local self-similar behavior over scales. Also, the use of first-order generalized fractals as building blocks also leads to possibility
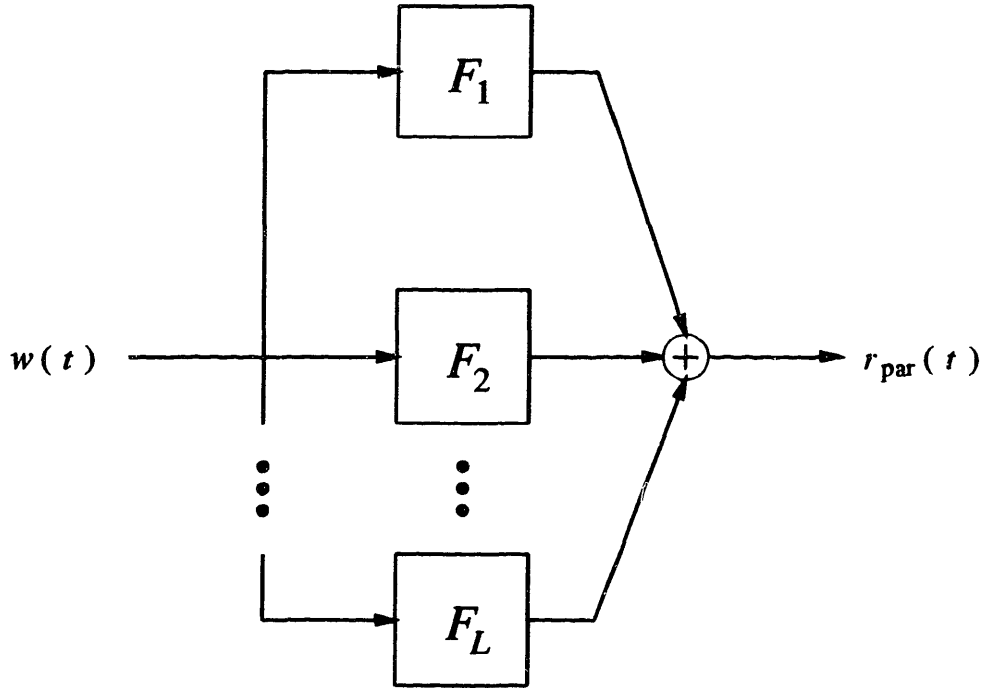
Figure 3-9: *Parallel arrangement of synthesis filters for first-order generalized fractals*

of a flat spectrum at low frequencies.

Similarly, we can consider a parallel arrangement of synthesis filters for first-order generalized fractals driven by a unit-variance white input. This scheme is illustrated in Figure 3-9. In the wavelet domain, we have the relation

$$\text{Var } (r_{\text{par}})_n^m = \left[\left(\frac{\sigma_1}{2^{\gamma_1 m} + p_1}\right)^{1/2} + \cdots + \left(\frac{\sigma_L}{2^{\gamma_L m} + p_L}\right)^{1/2}\right]^2.$$

A generalization of this case which eases analysis is shown in Figure 3-10. The systems $F_1, \ldots, F_L$ are now driven by independent white inputs with unit variance. Thus, $x_1(t), \ldots, x_L(t)$ are a set of independent first-order generalized fractals. Moreover, their power spectral densities are of the form

$$S_i(\Omega) = \frac{\sigma_i^2}{|\Omega|^{\gamma_i} + p_i}.$$
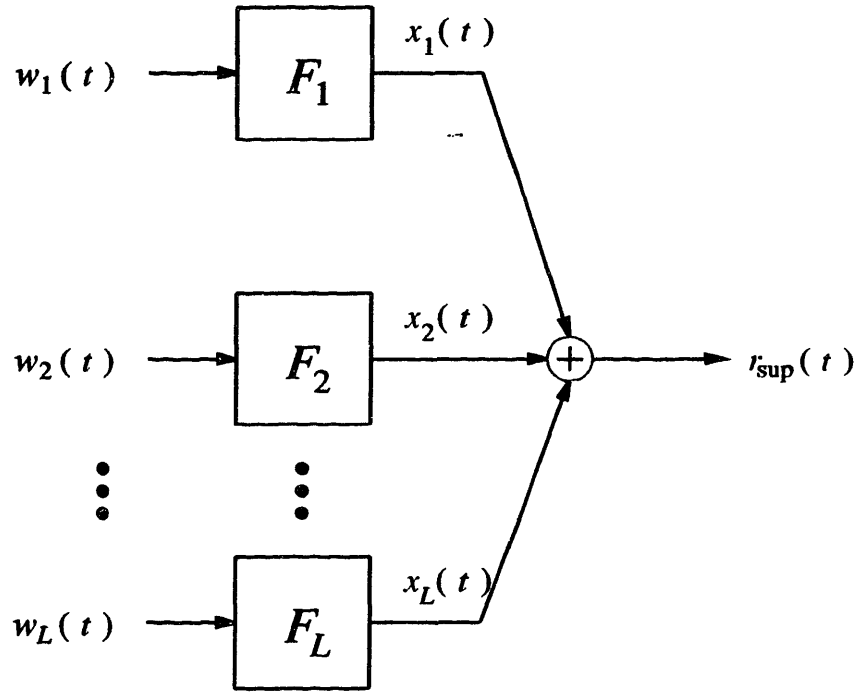
Figure 3-10: *Superposition of synthesis filters for first-order generalized fractals*

Due to the independence of $x_i(t)$, the power spectral density of $r_{sup}(t)$ is given by

$$S_{sup}(\Omega) = \sum_{i=1}^{L} \frac{\sigma_i^2}{|\Omega|^{\gamma_i} + p_i}. \tag{3.12}$$

Since the log-log plot of the sum of curves is roughly equal to the most dominant one, it is clear that the self-similar behavior observed at each scale is roughly due to only one of the first-order generalized fractals making up the higher-order generalized fractals. Thus, the local self-similar behavior of this higher-order generalized fractal is captured by one of the γ_i. This suggests that using this type of construction, we can also get higher-order generalized fractals exhibiting up to L types of local self-similar behavior.

This last scheme for constructing higher-order generalized fractals is of interest because in addition to models for processes having varying self-similar behavior over scales, this also gives models for certain scenarios of noise-corrupted signals. To see

this, we note that if both a signal and an additive noise are generalized fractals and are independent of each other, then the power spectral density of the signal observed in the noise is of the form given in (3.12). In the rest of this thesis, we shall focus on the type of generalized fractal processes associated with the scheme given in Figure 3-10. By an Lth-order generalized fractal, then, we shall be referring to the generalized fractal process associated with the type of construction given in Figure 3-10.

Chapter 4

Processing of Generalized Fractal Signals

Having introduced a class of generalized fractal signals, we now turn to a number of basic signal processing problems involving such signals. As suggested by Theorem 3.1, the dyadic wavelet transform provides an ideal environment for studying these generalized fractal signals. Our strategy for processing such signals, therefore, generally involves analysis in the wavelet domain.

Throughout this chapter, we shall focus on the case where the signal in question is zero-mean Gaussian. This is motivated by a number of reasons. First, this gives rise to more tractable problems. Indeed, because Gaussianity is preserved by linear transformations, the wavelet transform of a Gaussian process will be a set of jointly-Gaussian random variables. Other probability densities, on the other hand, are generally not as tractable under the wavelet transformation. Furthermore, we note that this assumption is also somewhat realistic since signals observed in nature are often the superposition of many elementary processes, and are thus approximately Gaussian by the Central Limit Theorem.

Although a number of different problems and algorithms will be considered in this chapter, several aspects are common to all our experimentations. First, the data used in the simulations are all synthesized using a frequency sampling scheme. By this scheme, we weight its Discrete Fourier Transform of a white Gaussian random

sequence according to the desired profile. While this synthesis scheme is easy to implement and allows much control over synthesis parameters, its fundamental difference from the wavelet-based analysis framework also allows us to test the robustness of our algorithms. The second aspect found in all our experimentations is that for actually computing the wavelet transform, we shall use a DWT scheme with the 4th-order Daubechies wavelet, which corresponds to an $h[n]$ with 8 nonzero coefficients. This filter length implies that the range of frequencies we can observe is roughly $\pi \cdot (8/N) \le \omega \le \pi$.

Before we present the algorithms, we first introduce some wavelet-related notation which will prove convenient in the sequel. While in theory, the dyadic wavelet transform of a function $r(t)$ can be thought of as the projection of the function onto the wavelet basis via

$$r_n^m = \int_{-\infty}^{\infty} r(t)\, \psi_n^m(t)\, dt,$$

in practice, we can only compute the wavelet transform over a limited range of scales due to the limitations in signal duration and resolution. In general, the set of scales is indexed by the set

$$\mathcal{M} = \{m_1, m_2, \ldots, m_M\},$$

and the set of points at each scale is indexed by the set

$$\mathcal{N}(m) = \{n_1(m), n_2(m), \ldots, n_{N(m)}(m)\}.$$

N is defined as the total number of wavelet transform points extracted from the signal $r(t)$, i.e.,

$$N = \sum_{m \in \mathcal{M}} N(m).$$

47

We also use the compact symbol **r**, given as

$$\mathbf{r} = \{r_n^m, m, n \in \mathcal{R}\} = \{r_n^m, m \in \mathcal{M}, n \in \mathcal{N}(m)\},$$

to denote the set of wavelet transform points extracted from $r(t)$. As we shall be using the computationally efficient DWT for obtaining such transform values, the set of indices can be more specifically labeled as

$$\mathcal{M} = \{m_c, m_c + 1, \ldots, m_f - 1, m_f\}$$

where m_c and m_f refer to the coarsest and finest scales, respectively, and

$$\mathcal{N}(m) = \{1, 2, \ldots, N_0 2^{m-1}\}.$$

Here, N_0 is a constant dependent on the length of the filter $h_0[n]$ used in the DWT.

4.1 Parameter Estimation

While second-order statistics of a random process are very useful in a wide variety of problems involving the process, this type of information is rarely known *a priori*. Referring back to (3.12), we see that the power spectral density of a generalized fractal process is completely captured in the set of parameters $(\gamma_1, \ldots, \gamma_L, p_1, \ldots, p_L, \sigma_1^2, \ldots, \sigma_L^2)$. The first problem we consider then is a parameter estimation problem involving the estimation of these parameters. In addition to its role in the solving of other problems, this problem is of significance in its own rights. As pointed out in Chapter 3, a number of physical properties of generalized fractal signals are reflected by these parameters. The exponents γ_i, for instance, sufficiently capture the local self-similar behavior, or roughness, of a generalized fractal over all scales. Estimators for γ_i are, therefore, directly applicable to problems such as texture classification and texture segmentation in image processing.

We shall approach this parameter estimation problem using the wavelet transform.

Recall that in the context of this thesis, an Lth-order generalized fractal process $r(t)$ is given by

$$r(t) = \sum_{i=1}^{L} x_i(t),$$

where $x_i(t)$ are independent first-order generalized fractal processes having power spectra

$$S_i(\Omega) = \frac{\sigma_i^2}{|\Omega|^{\gamma_i} + p_i}.$$

It is straightforward to see that since $x_i(t)$ are independent signals and since the dyadic wavelet basis functions are effectively eigenfunctions of generalized fractals, r is a set of essentially independent random variables with variance

$$\text{Var } r_n^m = \sigma_m^2 = \sum_{i=1}^{L} \frac{\sigma_i^2}{2^{\gamma_i m} + p_i}. \tag{4.1}$$

Further, since $r(t)$ is zero-mean Gaussian,

$$r_n^m \sim N(0, \sigma_m^2),$$

where $N(\mu, \nu)$ denotes a normal distribution with mean μ and variance ν. This information on the statistical behavior of r_n^m allows us to set up various estimators for the parameters. In the rest of this section, we shall give two such estimators, the maximum-likelihood (ML) estimator, and the maximum-spectral-flatness-measure (MSFM) estimator.

4.1.1 Maximum-Likelihood Estimator

Using the fact that r_n^m are independent Gaussian random variables with mean zero and variance given in (4.1), we can readily set up the likelihood function as

$$\mathcal{L}(\Theta) = p_{\mathbf{r}}(\mathbf{r}; \Theta) = \prod_{m,n \in \mathcal{R}} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{(r_n^m)^2}{2\sigma_m^2}\right], \qquad (4.2)$$

where Θ is the vector of parameters $(\gamma_1, \ldots, \gamma_L, p_1, \ldots, p_L, \sigma_1^2, \ldots, \sigma_L^2)$. The ML estimator, then, calls for the maximization of this function over the parameter space $\{-\infty < \gamma_i < \infty,\ 0 \leq p_i < \infty,\ 0 \leq \sigma_i^2 < \infty;\ i = 1, \ldots, L\}$. Due to the close coupling of the parameters in this estimator, simplification of (4.2) is difficult, and at first sight, it appears inevitable to numerically search the $3L$-dimensional parameter space to locate the maximum of the likelihood. Instead of attempting to solve this general problem directly, we shall first consider the first-order case. It will turn out that the solution for this case plays a key role in the solving of the general problem.

First-Order Generalized Fractals

If $r(t)$ is a first-order generalized fractal, then the variance given in (4.1) will be

$$\text{Var } r_n^m = \sigma_m^2 = \frac{\sigma_1^2}{2^{\gamma_1 m} + p_1}.$$

For notational convenience, we set $\beta = 2^{\gamma_1}$, $p = p_1$, and $\sigma^2 = \sigma_1^2$ to obtain, for σ_m^2, the expression

$$\sigma_m^2 = \frac{\sigma^2}{\beta^m + p}. \qquad (4.3)$$

Hence, it is the parameter set

$$\Theta = (\beta, p, \sigma^2)$$

we wish to estimate. Note that since the logarithm function is monotonic, the ML estimate of γ_1 can be derived from the ML estimate of β via

$$\hat{\gamma}_1 = \log_2 \hat{\beta}.$$

It is worth noting that if p is known *a priori* to be 0, then this problem reduces to the estimation of the parameters β and σ^2 of a $1/f$ signal. Detailed treatment of this case is given in [22].

Proceeding to the more general ML problem in which no *a priori* information of p is available, we first recall that the likelihood is given by

$$\mathcal{L}(\Theta) = p_{\mathbf{r}}(\mathbf{r}; \Theta) = \prod_{m,n \in \mathcal{R}} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{(r_n^m)^2}{2\sigma_m^2}\right]. \tag{4.4}$$

Taking the logarithm of (4.4) and grouping terms, we obtain

$$L(\Theta) = -\frac{1}{2} \sum_{m \in \mathcal{M}} N(m)\left\{ \frac{\hat{\sigma}_m^2}{\sigma_m^2} + \ln(2\pi\sigma_m^2) \right\}, \tag{4.5}$$

where

$$\hat{\sigma}_m^2 = \frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (r_n^m)^2,$$

or the sample variance of the data at scale m. Our interest then lies in the maximization of the log-likelihood function (4.5), with σ_m^2 as specified in (4.3). Taking the first-order partial derivatives of $L(\Theta)$ with respect to β, p, and σ^2, respectively, and setting them to zero, we obtain a set of necessary conditions for a stable point

$$\sum_{m \in \mathcal{M}} \frac{N(m)m\beta^{m-1}}{(\beta^m + p)^2}\left\{ \frac{\hat{\sigma}_m^2}{(\sigma_m^2)^2} - \frac{1}{\sigma_m^2} \right\} = 0, \tag{4.6}$$

$$\sum_{m \in \mathcal{M}} \frac{N(m)}{(\beta^m + p)^2}\left\{ \frac{\hat{\sigma}_m^2}{(\sigma_m^2)^2} - \frac{1}{\sigma_m^2} \right\} = 0, \tag{4.7}$$

$$\sum_{m \in \mathcal{M}} \frac{N(m)}{\beta^m + p}\left\{ \frac{\hat{\sigma}_m^2}{(\sigma_m^2)^2} - \frac{1}{\sigma_m^2} \right\} = 0. \tag{4.8}$$

51

Substituting (4.3) into (4.8), we obtain the equation

$$\sum_{m \in \mathcal{M}} N(m) \left\{ \frac{\hat{\sigma}_m^2}{(\sigma^2)^2}(\beta^m + p) - \frac{1}{\sigma^2} \right\} = 0. \tag{4.9}$$

Solving (4.9) for σ^2, then, we obtain

$$\sigma^2 = \frac{\displaystyle\sum_{m \in \mathcal{M}} N(m)\hat{\sigma}_m^2(\beta^m + p)}{\displaystyle\sum_{m \in \mathcal{M}} N(m)}. \tag{4.10}$$

So, the estimate of σ^2 can be expressed in terms of the estimates of β and p. Substituting (4.10) into (4.3), and substituting the result into (4.5), we can express the likelihood as a function of only β and p. Moreover, the resulting function is empirically shown to be very easily optimized. Simple optimization algorithms such as the Nelder-Mead simplex search algorithm [8] appear to be sufficient for locating the correct estimates. Once the estimates of β and p are obtained, the estimate of σ^2 can be readily derived using (4.10).

Preliminary simulations have been done to test the performance of this ML estimator. The estimation was performed on signals synthesized with various combinations of parameter values, and for each parameter combination, the RMS errors of the parameter estimates were computed over 32 trials. These RMS errors are plotted in Figures 4-1 to 4-3. Much can be seen from these plots. First, these plots suggest that all three estimators are in some sense consistent, as the errors generally decrease with increasing data length. This is reasonable because as N increases, we can observe the behavior of the generalized fractal over a broader set of scales (and thus frequencies), and therefore have more data on which to base our estimates. Although this argument suggests a monotonic decrease in the RMS errors as data length increases, a lack of monotonicity is observed in a number of the plots. This is mainly because the RMS errors were computed over only 32 trials which was not enough for averaging out the effects of random errors, and for the RMS errors to converge to their expected values.
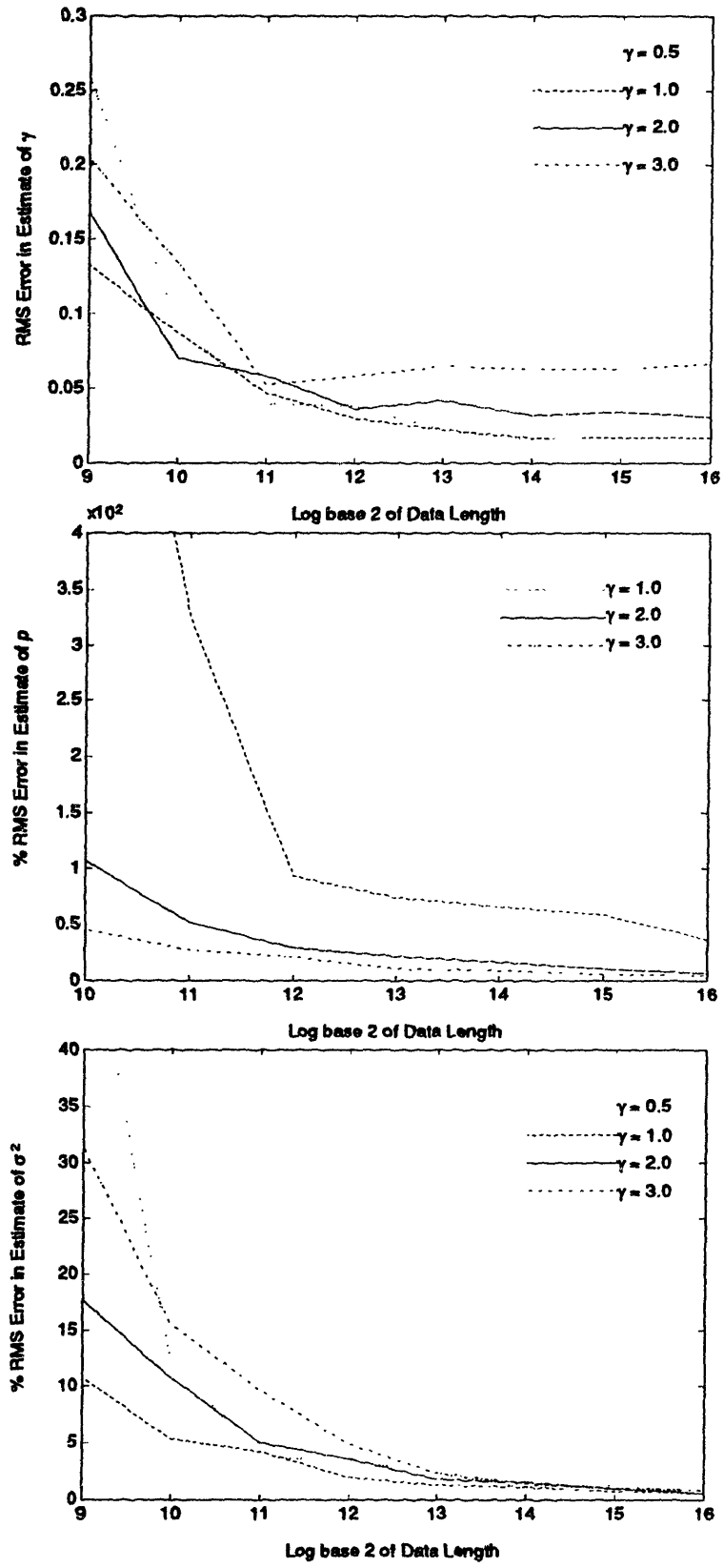
Figure 4-1: *Performance of the ML parameter estimation algorithm for a first-order generalized fractal process (p = .001)*

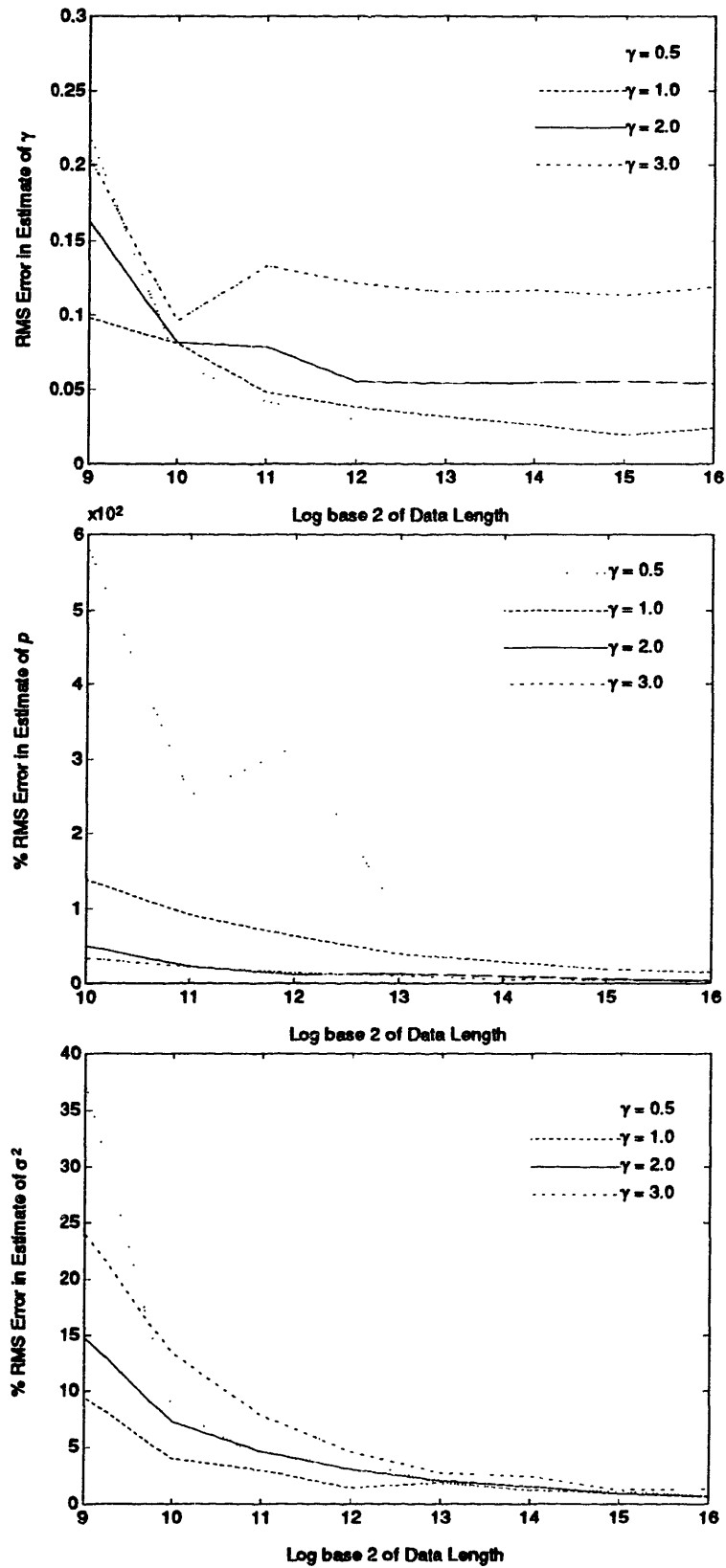**Figure 4-2:** *Performance of the ML parameter estimation algorithm for a first-order generalized fractal process (p = .01)*
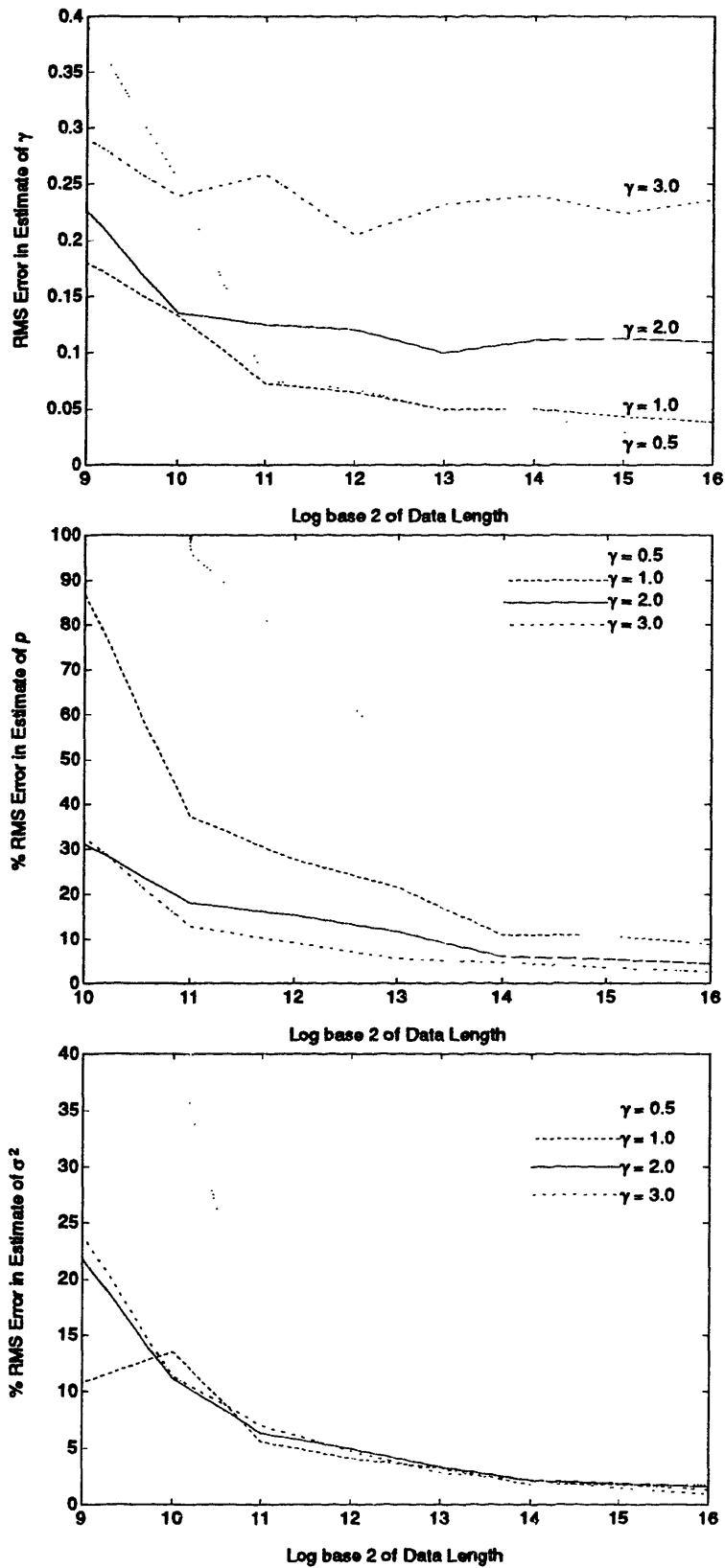
54

Figure 4-3: *Performance of the ML parameter estimation algorithm for a first-order generalized fractal process (p = .1)*

As we focus on the performance of the individual estimators over different sets of synthesis parameters, we see that while the performance of the estimator for σ^2 is quite independent of the true values of the parameters, the performance of the estimators for γ and p both degrade as p increases. Recall that on a logarithmic scale, the shape of the power spectral density of a first-order generalized fractal is approximated as being flat at low frequencies and decaying linearly with a fractional slope at higher frequencies. As the fractional pole is pushed higher, the fractional linear decay portion of the spectrum cannot be seen over as many scales. Consequently, it becomes more difficult to correctly fit a generalized fractal model onto the data and the estimators give poor results. We note that for very small p, a similar problem arises. In this case, while the flat part of the power spectrum is absent in the visible frequencies, leading to difficulty in correctly fitting a first-order generalized fractal model onto the data, the effects of γ can be seen over virtually all scales. Consequently, the estimator for γ is very accurate while the estimator for p suggests that $p = 0$. It is therefore worth noting that for the estimators to work correctly, it is important that the first-order generalized fractal behavior be sufficiently captured in the scales that are available to us.

The effects of varying γ in the synthesis are slightly different. Referring to (3.3), we see that the spectral breakpoint occurs at roughly $\omega_c = p^{1/\gamma}$. To keep p within the range of frequencies observable in the DWT, we have chosen p to be smaller than 1 in all of the simulations. This implies that as γ increases, the transition frequency ω_c moves up. This has the same effects on the estimator for γ as that due to increasing p, and causes an increase in its error. The performance of the estimator for p, on the other hand, improves as γ increases. This is because when γ is large, the transition of the spectral behavior is very sharp, making it easier for the estimator to locate the point of transition.

Higher-Order Generalized Fractals

Turning now to the higher-order case, we first note that in the derivation of (4.2), we have not made full use of the structure of an Lth-order generalized fractal as a superposition of L independent first-order generalized fractals. In this section, we shall see that this fact allows us to exploit the results of the first-order case in solving the Lth-order generalized fractal parameter estimation problem. The main tool that enables us to achieve this is the Estimate-Maximize, or the EM algorithm.

The EM algorithm is an iterative algorithm for solving ML problems [7]. Starting with an initial guess, the EM algorithm adjusts the parameter estimates to yield a higher value of the likelihood at every iteration. It is, therefore, certain to locate at least a local maximum point of the likelihood function. To achieve this monotonic increase in the likelihood, the EM algorithm requires, at each iteration, the solving of a maximization problem typically less complex than the original ML problem. Thus, the EM algorithm can be viewed as a decomposition of a large maximization problem into a series of small maximization problems. While a detailed derivation of the EM algorithm is presented in Appendix B, we will present here the key results of the derivation.

Each iteration of the EM algorithm consists of two steps, the E step and the M step. In the E step, the variance of each constituent process $x_i(t)$ is estimated based on the current parameter estimates $\Theta^{[k]}$ via

$$S_i^m(\Theta^{[k]}) = A_i^m(\Theta^{[k]}) + B_i^m(\Theta^{[k]})\hat{\sigma}_m^2,$$

where

$$A_i^m(\Theta) = \lambda_i^m \left(1 - \frac{\lambda_i^m}{\lambda_1^m + \lambda_2^m + \cdots + \lambda_L^m} \right)$$

$$B_i^m(\Theta) = \left(\frac{\lambda_i^m}{\lambda_1^m + \lambda_2^m + \cdots + \lambda_L^m} \right)^2,$$

and

$$\lambda_i^m \triangleq \frac{\sigma_i^2}{\beta_i^m + p_i}.$$

In the M step, these estimated variances are used for obtaining the next set of parameter estimates via

$$\Theta_i^{[k+1]} = \arg\max_{\Theta_i}\left(-\frac{1}{2}\sum_{m\in M} N(m)\left\{\frac{1}{\lambda_i^m}S_i^m(\Theta^{[k]}) + \ln 2\pi\lambda_i^m\right\}\right), \qquad i = 1,\ldots,L, \tag{4.11}$$

where Θ_i denotes the vector of parameters $(\beta_i, p_i, \sigma_i^2)$. Because (4.11) is of the same algebraic form as (4.5) techniques used for the first-order case can be applied in the M step.

We have tested this estimator on second-order generalized fractals synthesized with one set of parameters, and have obtained the following results.

Parameter	True Value	RMS Error
γ_1	3.0	0.0709
p_1	1.0×10^{-15}	2.31×10^{-13}
σ_1^2	1.0	0.2035
γ_2	0.3	0.1392
p_2	0.1	0.1
σ_2^2	10	1.29

As in simulations discussed earlier, the RMS errors were computed over 32 trials. In choosing the parameters, we were concerned that the effects of both constituent first-order generalized fractals be observable over the scales to which we had access. If we had allowed either constituent to be dominant over all the scales we could see, we would end up trying to fit a second-order generalized fractal model onto an essentially first-order generalized fractal, and this would lead to erroneous estimates of a set of parameters, Θ_i. To make the effects due to the two constituent generalized fractal discernible as well as observable, however, we were forced into a set of parameters which does not favor the estimation of p_2. Recall from our previous discussion in the

58

first-order case that a small γ together with a large p degrades the performance of the estimator for p. Other than the error in p_2, however, the estimates in this case were quite acceptable.

A particularly interesting case arises when $x_L(t)$ is a white Gaussian noise, or, in other words, when we know *a priori* that

$$\gamma_L = 0 \ (\text{or } \beta_L = 1), \qquad p_L = 0, \qquad \sigma_L^2 = \sigma_w^2,$$

and, therefore

$$\lambda_L^m = \sigma_w^2.$$

In this case, the maximization of

$$-\frac{1}{2} \sum_{m \in M} N(m) \left\{ \frac{1}{\lambda_L^m} S_L^m(\tilde{\Theta}) + \ln 2\pi \lambda_L^m \right\},$$

is achieved by simply choosing the closed form solution

$$\hat{\lambda}_L = \frac{\displaystyle\sum_{m \in \mathcal{M}} N(m) S_L^m(\tilde{\Theta})}{\displaystyle\sum_{m \in \mathcal{M}} N(m)}.$$

This leads to a computational saving of 1 optimization per iteration.

Figures 4-4 to 4-7 describe the results of simulations performed for the special case where the second generalized fractal process in a second-order generalized fractal process is a white noise. In this experiment, we embedded synthesized first-order generalized fractal in white Gaussian noise, the strength of which we varied. We then applied the EM-based ML estimator to this corrupted signal to jointly estimate the signal parameters (γ, p, σ^2) and the noise strength σ_w^2. The number of data points used was $N = 2^{13}$, and the RMS errors were calculated over 32 trials. We can see that the results here agree very well with the results obtained in the parameter estimation experiment performed on noise-free first-order generalized fractals. More specifically,

Figure 4-4: *Performance of the EM parameter estimation algorithm for a first-order generalized fractal process embedded in white noise (p = .001)*

Figure 4-5: *Performance of the EM parameter estimation algorithm for a first-order generalized fractal process embedded in white noise (p = .001)*

Figure 4-6: *Performance of the EM parameter estimation algorithm for a first-order generalized fractal process embedded in white noise (p = .01)*
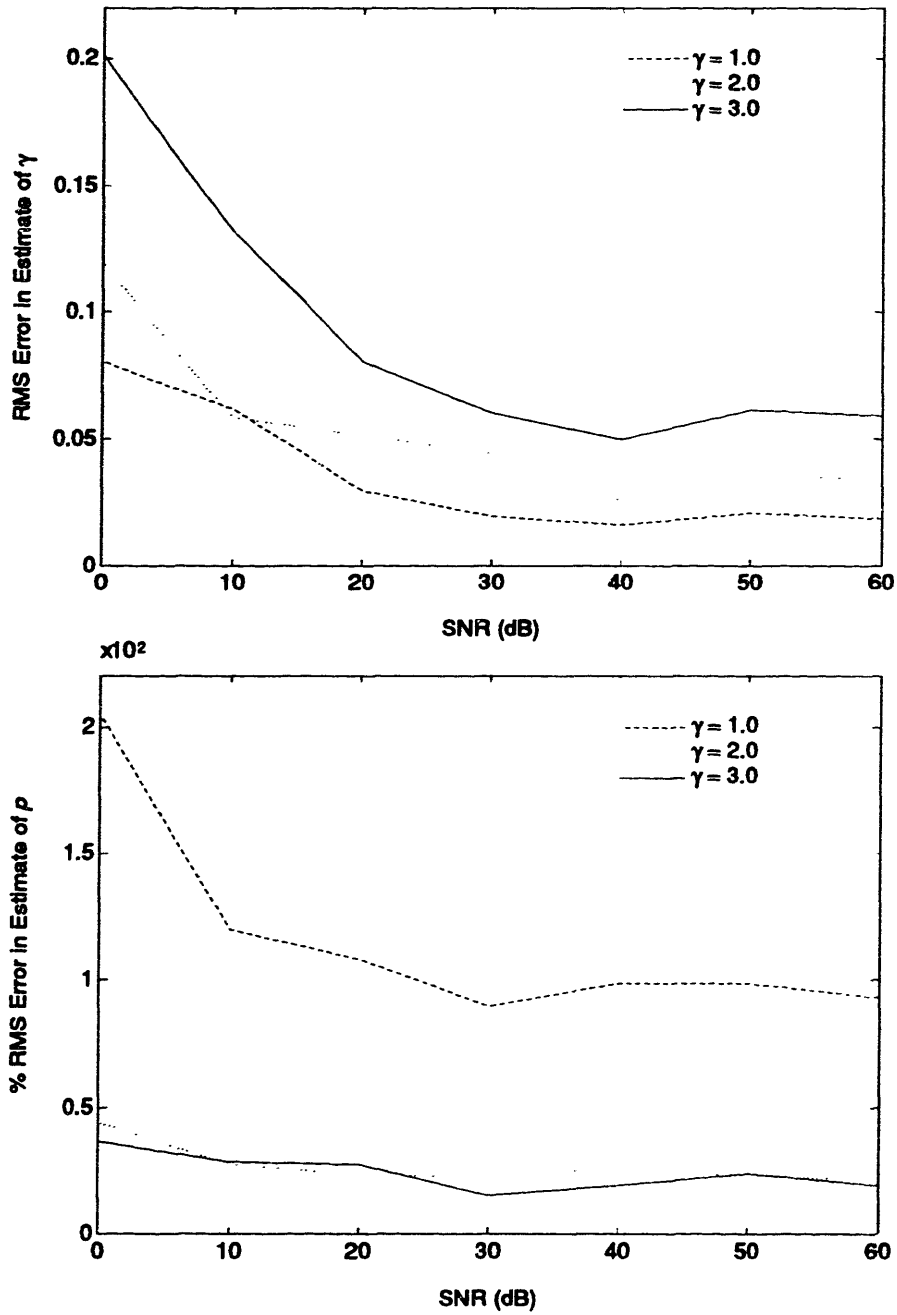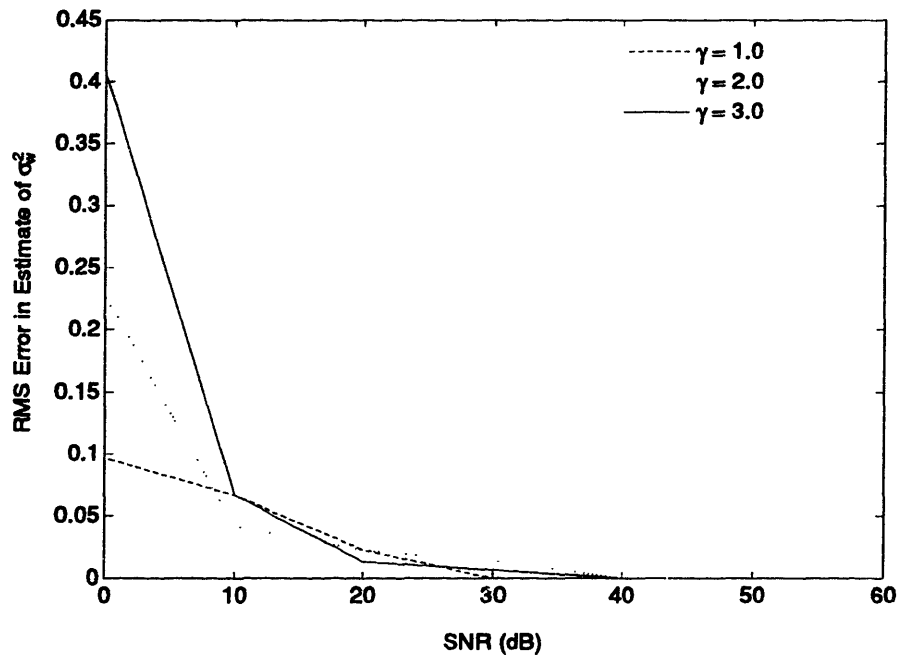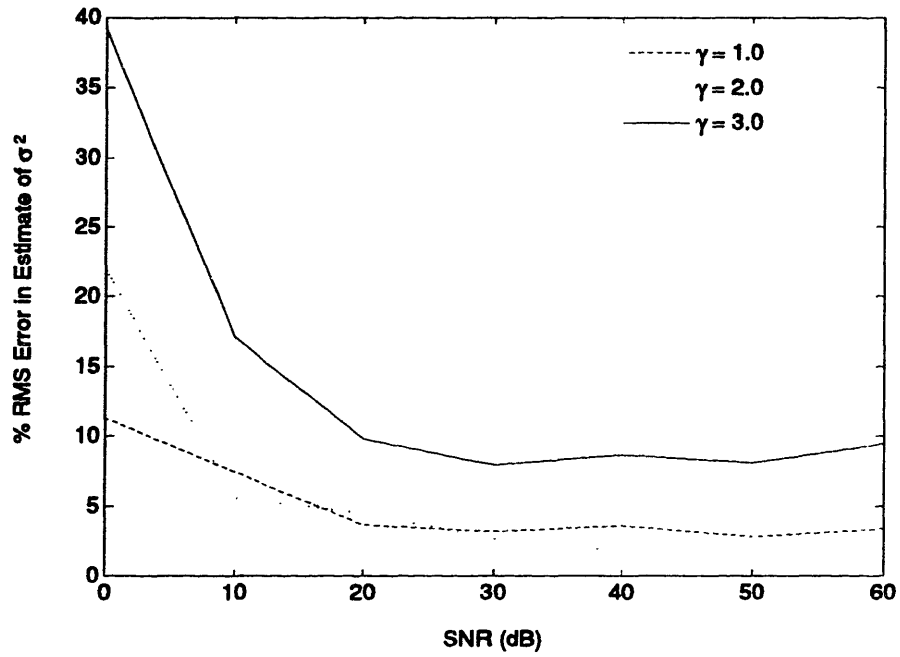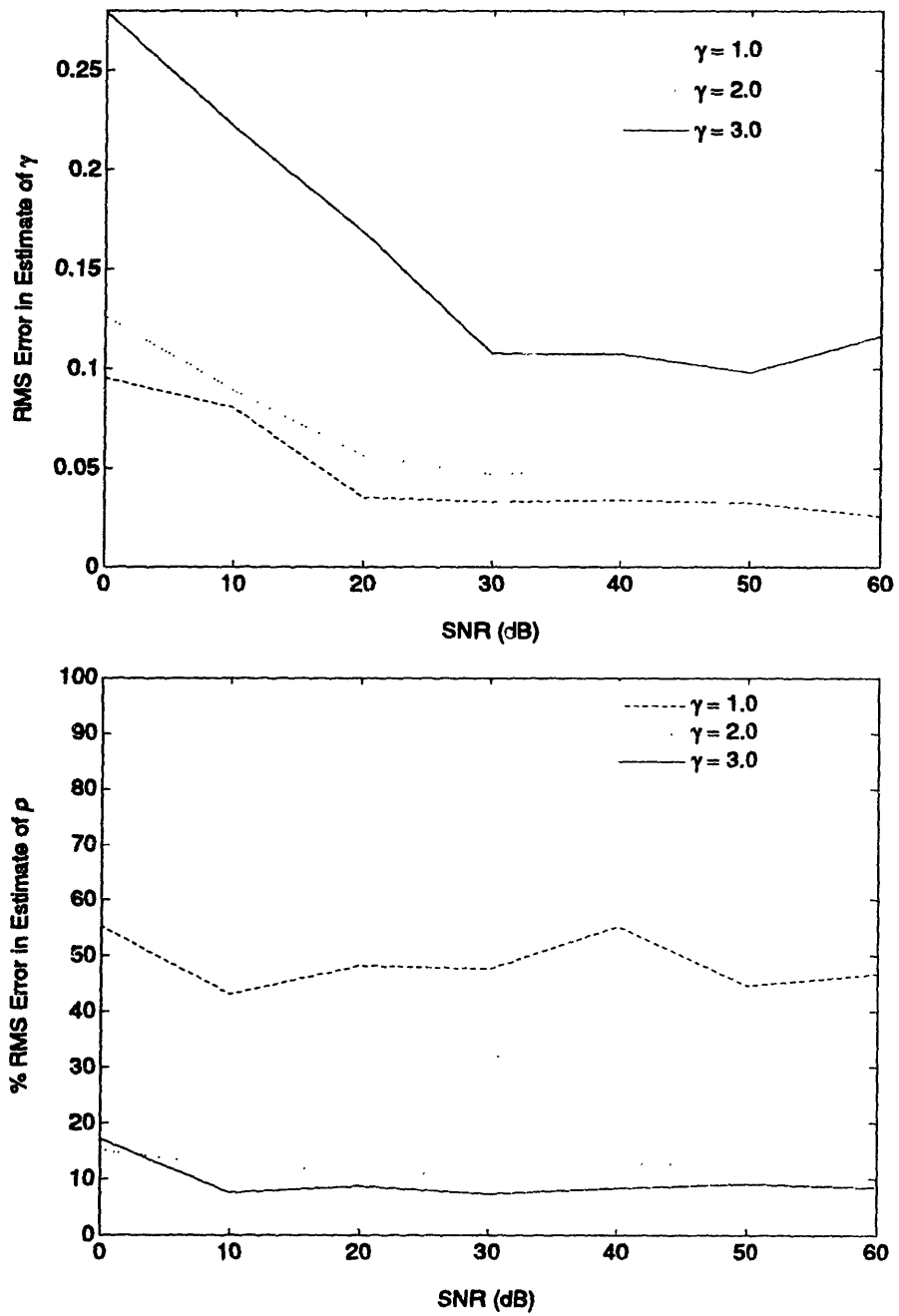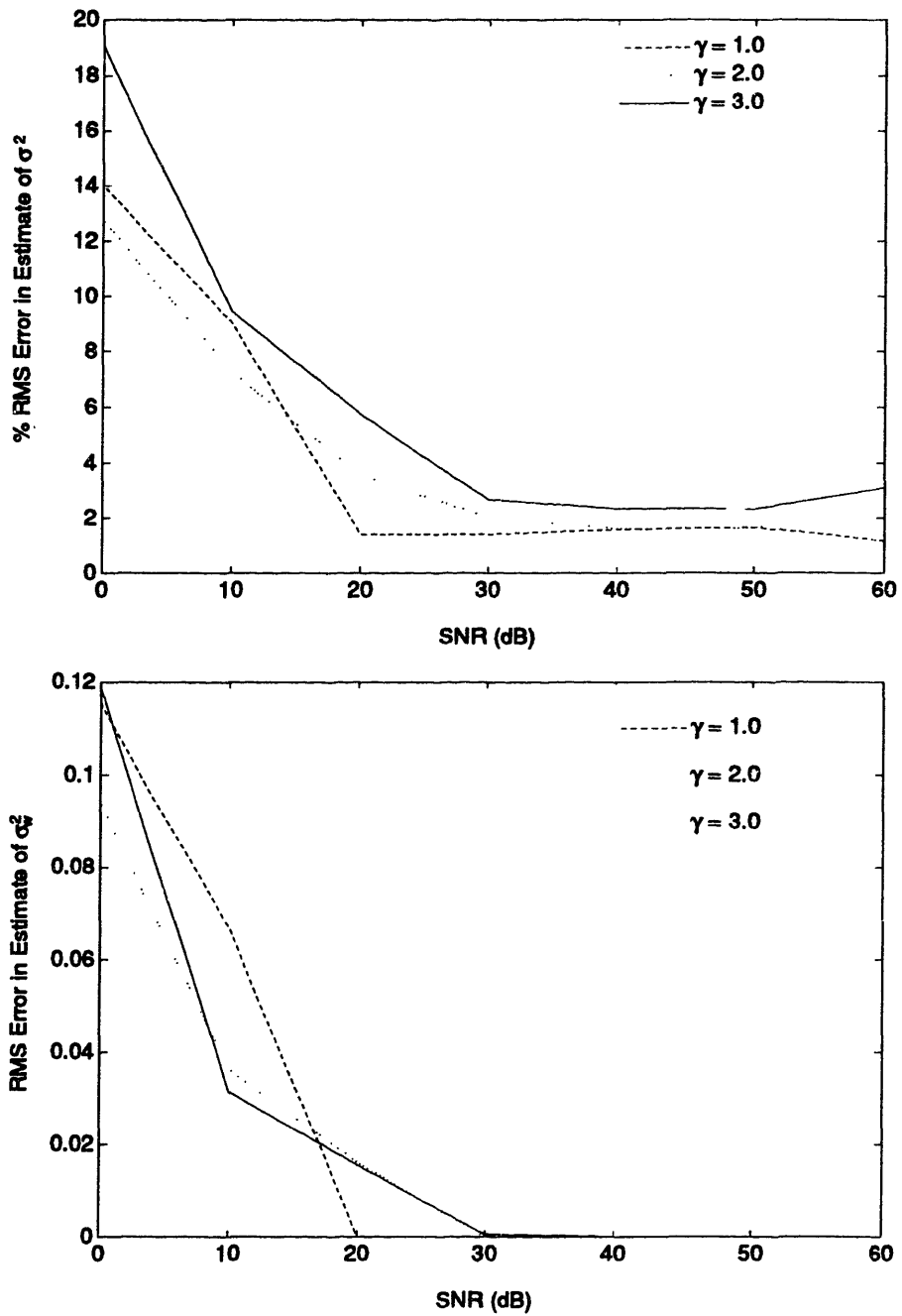
Figure 4-7: *Performance of the EM parameter estimation algorithm for a first-Order generalized fractal process embedded in white noise (p = .01)*
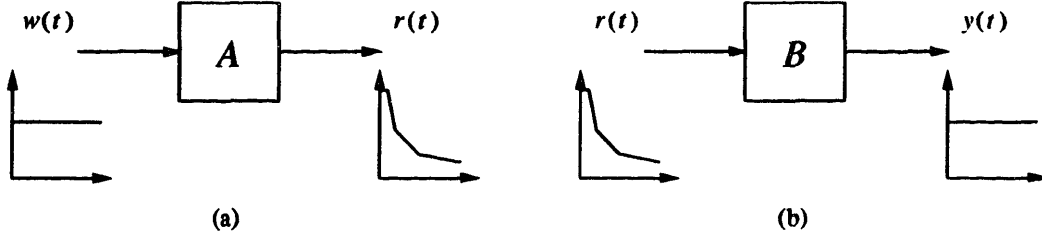
Figure 4-8: *Idea behind the maximum-spectral-flatness-measure estimator: (a) system generating generalized fractal process from white input; (b) inverse system of A*

the estimators for γ and p are not as reliable for larger values of p, while larger values of γ result in better estimates for p and poorer estimates for γ. Also, we observe that as the noise strength decreases, the RMS errors in the estimates of γ and p converge to the values corresponding to $N = 2^{13}$ in the noise-free case. Finally, the accuracy in the estimates even at low SNR implies that this estimation strategy is quite robust.

4.1.2 Maximum-Spectral-Flatness-Measure Estimator

In this section we present an alternative to the ML estimator for solving the parameter estimation problem, namely, the maximum-spectral-flatness-measure or MSFM estimator. The main idea behind this estimation strategy is illustrated in Figure 4-8. Just as we can view a generalized fractal process as the output of a wavelet-based synthesis filter A driven by white noise, we can think of undoing the effects of A by using a hypothetical wavelet-based system B. If B is parametrized appropriately, its output $y(t)$, should be a white noise with unit variance. In other words, its power spectrum should be maximally flat. The MSFM estimator, therefore, uses the spectral-flatness-measure of the output of B as the criterion for the estimation.

A reasonable definition for the spectral flatness measure is

$$\text{sfm } \{y(t)\} = \exp\left(\frac{1}{\sum_{m \in \mathcal{M}} N(m)} \sum_{m,n \in \mathcal{R}} \ln(\tilde{\sigma}_m^2)\right) \Big/ \left(\frac{1}{\sum_{m \in \mathcal{M}} N(m)} \sum_{m,n \in \mathcal{R}} \tilde{\sigma}_m^2\right),$$

$$(4.12)$$

where $\tilde{\sigma}_m^2$ denotes the variance of the y_n^m. Equation (4.12) can be interpreted as the

64

ratio of the geometric mean of the variances $\tilde{\sigma}_m^2$ to their arithmetic mean. Because the geometric mean of a collection of non-negative numbers is always less than or equal to their arithmetic mean, and equality arises only if the numbers are all equal,

$$\text{sfm } \{y(t)\} \leq 1, \tag{4.13}$$

with equality only if all variances are equal. In the case where $y(t)$ is white, the wavelet coefficients y_n^m will all have the same variance because of the completeness and orthonormality of the wavelet transform. The resulting spectral flatness measure as defined in (4.12) is then equal to 1, or maximized. Thus, the definition given in (4.12) does correspond to a reasonable measure of the spectral flatness of $y(t)$.

In our MSFM parameter estimator for first-order generalized fractal process, we use, for $\tilde{\sigma}_m^2$, the sample variances of the sequence y_n^m:

$$\tilde{\sigma}_m^2 = \frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (y_n^m)^2. \tag{4.14}$$

That we can obtain the sample variance via (4.14) is because we model the wavelet transform r_n^m of a first-order generalized fractal as a wide-sense stationary white sequence for each m. Weighting the transform values at each scale by the same constant does not alter the stationarity or the whiteness of the sequence. Moreover, that $r(t)$ is a zero-mean, Gaussian process implies that each y_n^m is a zero-mean Gaussian random variable. Hence, the variance of each y_n^m can still be reasonably estimated using (4.14)

Based on this set-up, a detailed derivation of the MSFM parameter estimator is given in Appendix B. The resulting objective function to be used in this estimator is

$$U(\beta, p) = \sum_{m \in \mathcal{M}} N(m) \ln(\beta^m + p) - \left(\sum_{m \in \mathcal{M}} N(m) \right) \cdot \ln \left(\sum_{m \in \mathcal{M}} N(m) \hat{\sigma}_m^2 (\beta^m + p) \right).$$

Like the ML estimator discussed earlier, this estimator has the desirable property that the objective function involves only β and p, and can therefore be optimized via a two-dimensional search. The independence of σ^2 is reasonable if we consider the log-log plot of the power spectrum. Multiplication by a constant merely shifts the

log-spectrum vertically, and does not affect its shape. This means that the factor σ^2 cannot affect the flatness of the power spectrum and hence its spectral flatness measure. It is, therefore, not surprising that it drops out of the expression for the spectral flatness measure.

We have run a set of simulations to examine the performance of the MSFM parameter estimator. Again, the experimentation involved the estimation of the parameters from synthesized signals, and the results are shown in Figures 4-9 to 4-11. Here, we observe trends very similar to the ones prominent in the ML cases for both noise-free and noise-corrupted first-order generalized fractals. More precisely, the estimators are affected similarly by the values of γ and p used in the synthesis of the signals. This is reasonable since given a large number of data points, errors due to the randomness of the signals will be very small. The main origin of the errors will therefore be due to the lack of observability of the desired spectral characteristics over the range of available scales.

4.2   Signal Separation

In this section, we shall consider some problems of signal separation pertaining to our class of generalized fractal signals. More specifically, given an Lth-order generalized fractal

$$r(t) = \sum_{i=1}^{L} x_i(t), \tag{4.15}$$

we are interested in estimating the signal component

$$a(t) = \sum_{i \in \Gamma} x_i(t)$$

where $\Gamma = \{l_1, l_2, \ldots, l_k\}$, with $1 \leq l_1 < l_2 \cdots < l_k \leq L$. For this problem, we shall use a least-squares-error criterion for the estimation.
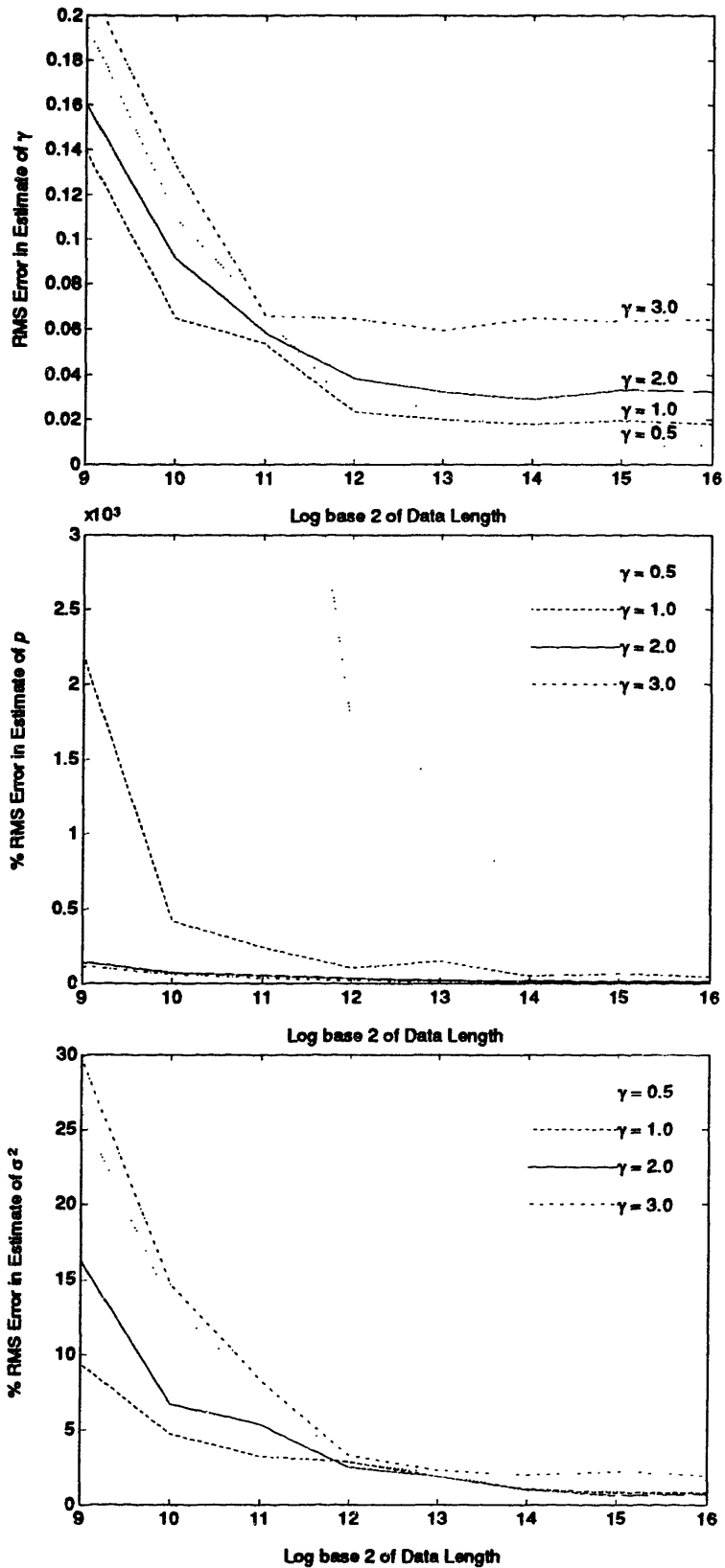
Figure 4-9: *Performance of the MSFM parameter estimation algorithm for a first-order generalized fractal process (p = .001)*
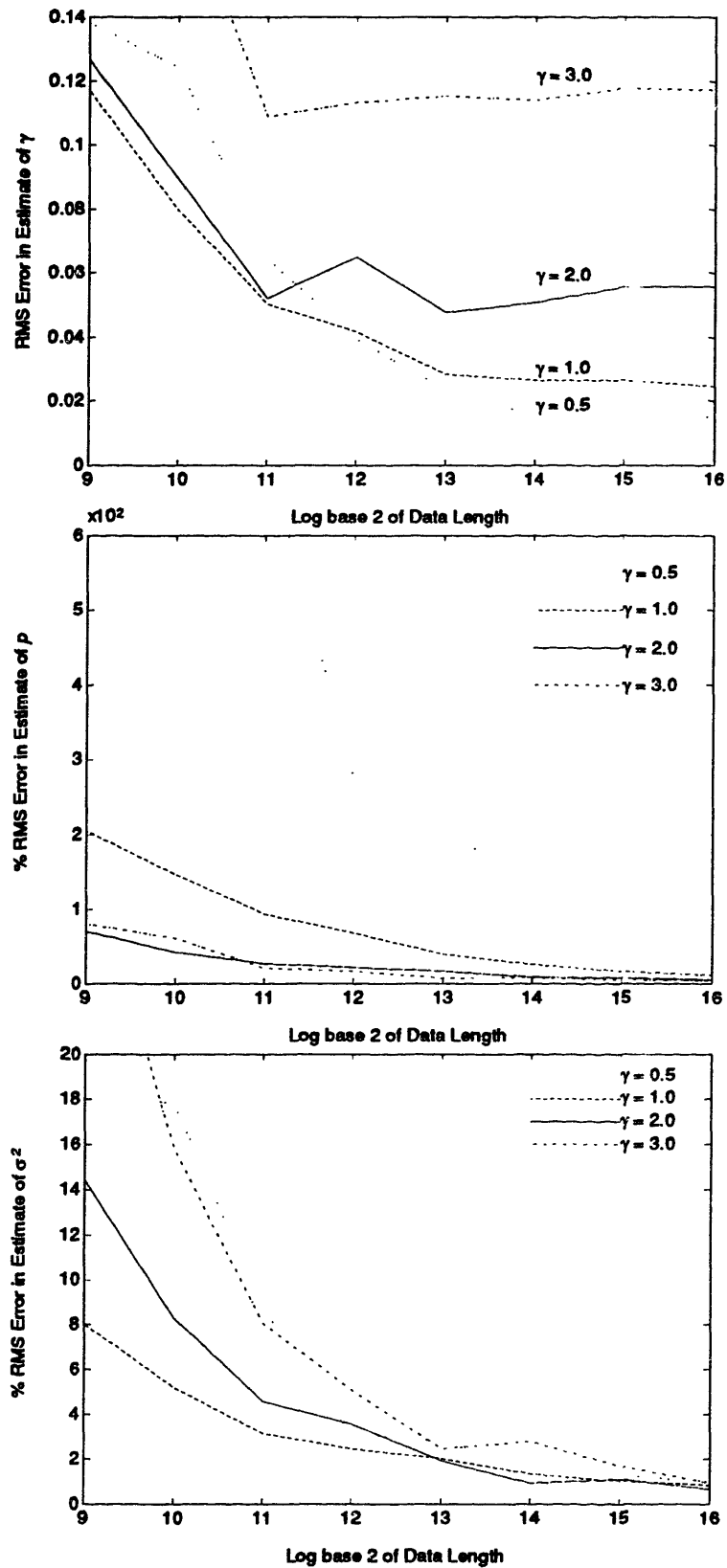
67

Figure 4-10: *Performance of the MSFM parameter estimation algorithm for a first-order generalized fractal process (p = .01)*

Figure 4-11: *Performance of the MSFM parameter estimation algorithm for a first-order generalized fractal process (p = .1)*
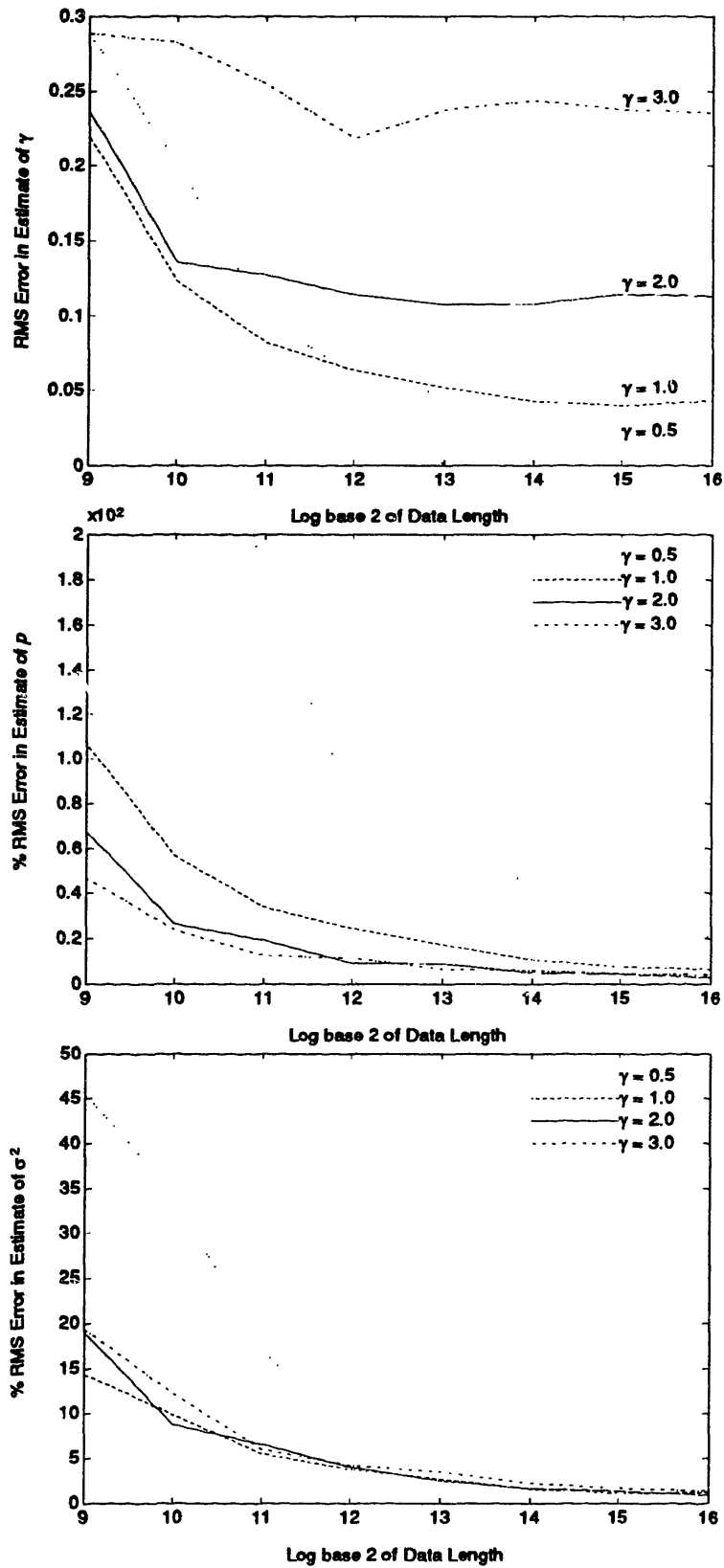
Because

$$\int E\{[\tilde{a}(t) - \hat{a}(t)]^2\}\, dt = \sum_{m,n \in \mathcal{R}} E\{(a_n^m - \hat{a}_n^m)^2\},$$

where

$$\tilde{a}(t) = \sum_{m,n \in \mathcal{R}} a_n^m\, \psi_n^m(t)$$

denotes a scale-limited representation of $a(t)$, the inverse dyadic wavelet transform of the least-squares-error estimates of a_n^m is the least-squares-error estimate of $\tilde{a}(t)$. In other words, the solution obtained using a wavelet transformation gives rise to the optimal estimate of a scale-limited version of $a(t)$. This justifies a wavelet-based approach to this problem.

We define the difference between $r(t)$ and $a(t)$ to be $b(t)$, so that

$$r(t) = a(t) + b(t). \tag{4.16}$$

Therefore,

$$b(t) = \sum_{i \in \Gamma'} x_i(t)$$

with Γ' defined as

$$\Gamma' = \{1, \ldots, L\} \backslash \Gamma.$$

Using the eigenfunction property of the wavelet functions, we have the wavelet transform relation

$$r_n^m = a_n^m + b_n^m,$$

70

where $\{a_n^m, b_n^m; \; m, n \in \mathcal{R}\}$ is a set of independent random variables, with

$$\text{Var } a_n^m = \lambda_a^m = \sum_{i \in \Gamma} \frac{\sigma_i^2}{\beta_i^m + p_i}$$

and

$$\text{Var } b_n^m = \lambda_b^m = \sum_{i \in \Gamma'} \frac{\sigma_i^2}{\beta_i^m + p_i}.$$

Because a_n^m and b_n^m are all zero-mean Gaussian, the Bayesian least-squares estimate of each a_n^m given r is

$$\hat{a}_n^m = \frac{\lambda_a^m}{\lambda_a^m + \lambda_b^m} r_n^m$$

Using these transform coefficients, then, we can find the estimate of $a(t)$ by using the inverse wavelet transform. The overall processing can be summarized more compactly as

$$\hat{a}(t) = \mathcal{W}^{-1}\left\{ \frac{\lambda_a^m}{\lambda_a^m + \lambda_b^m} \mathcal{W}\{r(t)\} \right\}.$$

Note that in the derivation of this estimate, we have assumed the knowledge of the parameters $\{\gamma_i, p_i, \sigma_i^2; \; i = 1, \ldots, L\}$. In principle, these parameters can be estimated using, for instance, the algorithms given in section 4.1.

Shown in Figure 4-12 is the results of applying the signal recovery algorithm to first-order zero-mean Gaussian generalized fractals corrupted by white Gaussian noise. The data length in this experiment is $N = 2^{11}$. Various choices of noise strength were used and the SNR gain in each case was the average of 32 trials. Note that in this experiment, the parameters $\{\gamma, p, \sigma^2, \sigma_w^2\}$ are estimated using the EM algorithm given in section 4.1. As can be seen from the plot, the performance of the algorithm is best when γ_1 is large and when p_1 is small. This is because the large difference in the shape of the spectra of signal and the corrupting noise in these cases lead to a large attenuation of the noise at high frequencies. Notice however that the
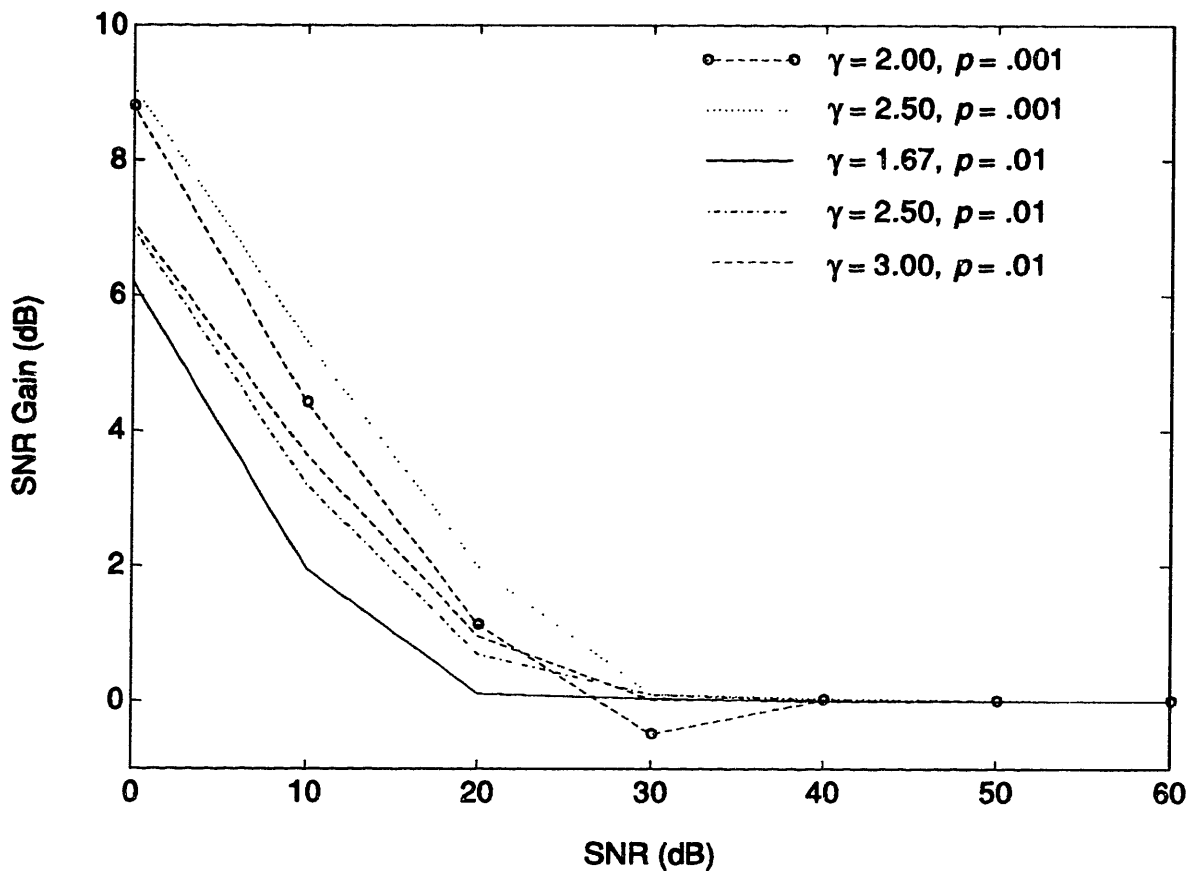
71

Figure 4-12: *Performance of the signal smoothing algorithm*

SNR gain can be negative at very high initial SNR. This is because of the removal of high-frequency information by the smoothing. Shown in Figure 4-13 is an example of signal recovery. It is a 400-point section of a 2^{15}-point generalized fractal which has undergone corruption and recovery. The result is quite remarkable as the main features of the signal are recovered despite the strong additive noise. Note that a 400-point scale is chosen here because it appears to be the longest length scale at which the correlation between points are still strong and features still simple. Length scales longer than this correspond to frequencies below the corner frequency of the power spectrum of the signal. In such regimes, the observations of the signal are too uncorrelated and random for the sake of this visual demonstration.

Original Signal ($\gamma = 3.0$, $p = .001$, $\sigma = 1.0$, $N = 2^{15}$):
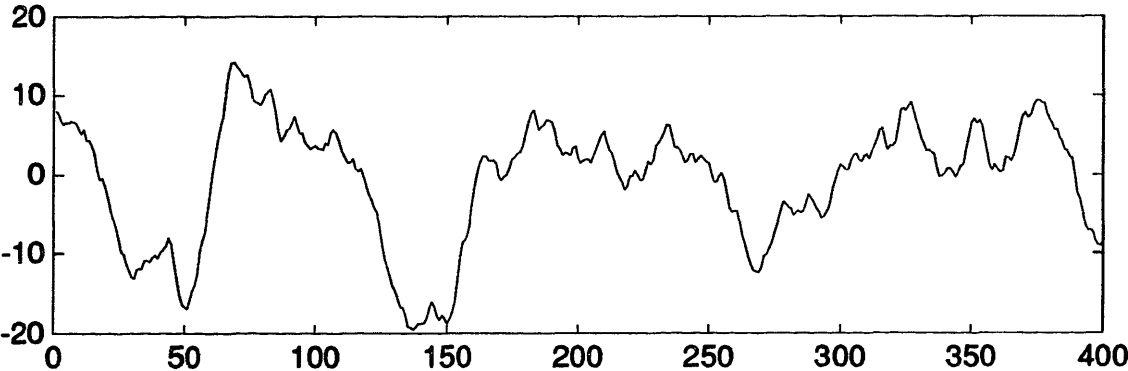
Original Signal Corrupted by White Noise (SNR = 0dB):
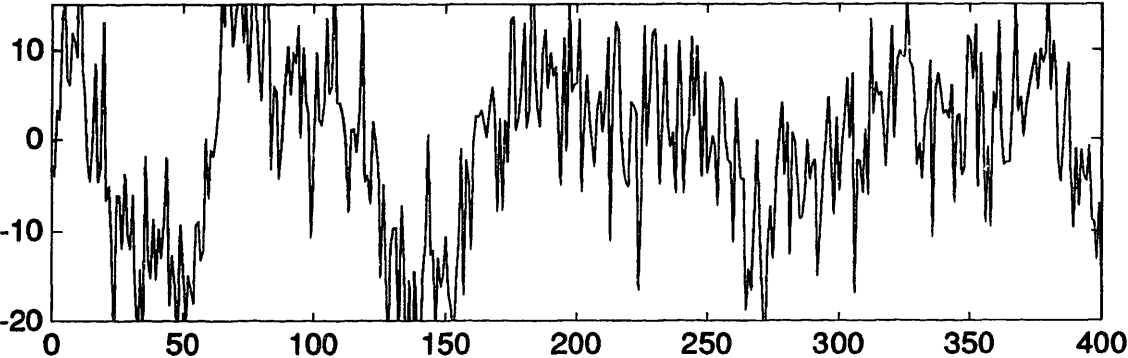
Smoothed Signal (SNR = 8.5dB):

Figure 4-13: *Signal smoothing example*

Chapter 5

Conclusions and Future Directions

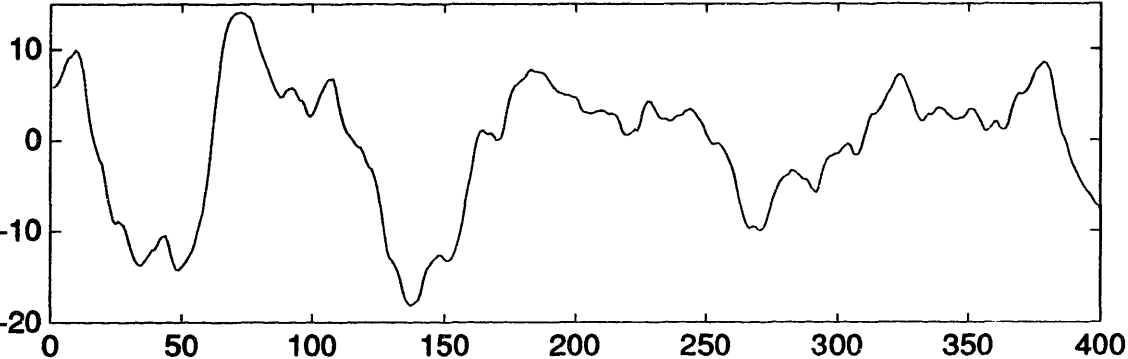In this work, we have focused on the development of a generalized class of fractal signals, a useful extension of the $1/f$ signal model. The generalized fractal model that we studied in this work was shown to be capable of modeling varying self-similar behavior over scales. It also allows the modeling of phenomena which show $1/f$ type structures at short scales but appear white at large length scales. Having established that the dyadic wavelet basis functions constitute as an approximate Karhunen-Loève basis for these generalized fractal signals, we formulated a number of wavelet-based algorithms for solving several fundamental signal processing problems involving such signals. Included in the package were a number of parameter estimation algorithms for classification purposes and a signal separation algorithm which finds application in problems such as signal recovery. In general, results of the simulations confirmed that wavelet transforms are well-suited for the analysis of these generalized fractal signals. In the parameter estimation problem involving clean first-order generalized fractal signals, it was demonstrated that both the maximum-likelihood and maximum-spectral-flatness-measure schemes give rise to consistent estimators with very good performance. The parameter estimation algorithm involving first-order generalized fractal signals embedded in white noise yielded good results even at very low signal-to-noise ratios. Finally, the signal separation problem involving the recovery of a first-order generalized fractal from observations corrupted by additive white noise gave reasonable SNR gain as well as recovery of features.

Although the generalized fractal signal model developed in this work constitutes as one class of generalization of the $1/f$ signal model, we note that much richer class of such generalizations exist. The cascade scheme shown in Figure 3-8, for example, gives rise to a significantly different class of generalized fractal signals. The development of even broader generalizations of the $1/f$ signal model and the associated signal processing algorithms offers a direction for future research effort.

Appendix A

Proof of Theorem 3.1

In this appendix, we present a proof of Theorem 3.1.

Proof:

We start by defining $x_M(t)$ as a resolution-limited approximation of $x(t)$:

$$x_M(t) = \sum_{m \geq M} \sum_n x_n^m \psi_n^m(t).$$

Since for $m \neq m'$, x_n^m and $x_n^{m'}$ are uncorrelated, and x_n^m is wide-sense stationary for each m, $x_M(t)$ is cyclostationary with period 2^{-M} [20]. Moreover, since for each m, x_n^m is white, with variance

$$\text{Var } x_n^m = \frac{\sigma^2}{2^{\gamma m} + p},$$

the time-averaged power spectral density of $x_M(t)$ is

$$S_M(\Omega) = \sum_{m \geq M} \frac{\sigma^2}{2^{\gamma m} + p} |\Psi(2^{-m}\Omega)|^2.$$

Because in practice, an observer is always restricted to a finite range of frequencies, a measured spectrum is always consistent with the spectral expression

$$S_{xx}(\Omega) = S(\Omega; p) \triangleq \lim_{M \to -\infty} S_M(\Omega) = \sum_m \frac{\sigma^2}{2^{\gamma m} + p} |\Psi(2^{-m}\Omega)|^2,$$

and we shall use this expression in the rest of this proof. Also, we shall set $\sigma^2 = 1$

without loss of generality. Using (2.29), (2.26), (2.27), and (2.28), in order, we can express $S(\Omega; p)$ in terms of the Fourier transform of the scaling function, i.e.,

$$S(\Omega; p) = \sum_m \left(\frac{1}{2^{\gamma(m-1)} + p} - \frac{1}{2^{\gamma m} + p} \right) |\Phi(2^{-m}\Omega)|^2. \qquad (A.1)$$

Now for any $\Omega \in \mathbf{R}$, we can choose $1 \le \Omega_0 < 2$, and $m_0 \in \mathbf{Z}$ such that $\Omega = 2^{m_0}\Omega_0$. After a change of variables, it follows from (A.1) that

$$S(\Omega; p) = 2^{-\gamma m_0} S\left(\Omega_0; \frac{p}{2^{\gamma m_0}} \right). \qquad (A.2)$$

Multiplying both sides of (A.2) by $(\Omega^\gamma + p)$, we obtain

$$(\Omega^\gamma + p)S(\Omega; p) = (\Omega_0^\gamma + \frac{p}{2^{\gamma m_0}})S\left(\Omega_0; \frac{p}{2^{\gamma m_0}} \right),$$

which leads to the bounds for $S(\Omega; p)$:

$$\inf_{\substack{1 \le \Omega_0 < 2 \\ m_0 \in \mathbf{Z}}} \left[(1^\gamma + \frac{p}{2^{\gamma m_0}})S\left(\Omega_0; \frac{p}{2^{\gamma m_0}} \right) \right] \frac{1}{\Omega^\gamma + p} \le S(\Omega; p)$$

$$\le \sup_{\substack{1 \le \Omega_0 < 2 \\ m_0 \in \mathbf{Z}}} \left[(2^\gamma + \frac{p}{2^{\gamma m_0}})S\left(\Omega_0; \frac{p}{2^{\gamma m_0}} \right) \right] \frac{1}{\Omega^\gamma + p}.$$

It remains, therefore, to show that

$$\sup_{\substack{1 \le \Omega_0 < 2 \\ m_0 \in \mathbf{Z}}} \left[(2^\gamma + \frac{p}{2^{\gamma m_0}})S\left(\Omega_0; \frac{p}{2^{\gamma m_0}} \right) \right] < \infty \qquad (A.3)$$

and

$$\inf_{\substack{1 \le \Omega_0 < 2 \\ m_0 \in \mathbf{Z}}} \left[(1^\gamma + \frac{p}{2^{\gamma m_0}})S\left(\Omega_0; \frac{p}{2^{\gamma m_0}} \right) \right] > 0. \qquad (A.4)$$

To show (A.3), we first note that by breaking down the summation, we can rewrite (A.1) as

$$S(\Omega;p) = \sum_{m=0}^{\infty} \left(\frac{1}{2^{\gamma(m-1)} + p} - \frac{1}{2^{\gamma m} + p} \right) |\Phi(2^{-m}\Omega)|^2$$
$$+ \sum_{m=1}^{\infty} \left(\frac{1}{2^{-\gamma(m+1)} + p} - \frac{1}{2^{-\gamma m} + p} \right) |\Phi(2^{m}\Omega)|^2 \qquad (A.5)$$

where the second summation has been rewritten via a change of variables. Equation (2.13) leads to a very simple upper bound for the first summation:

$$\sum_{m=0}^{\infty} \left(\frac{1}{2^{\gamma(m-1)} + p} - \frac{1}{2^{\gamma m} + p} \right) |\Phi(2^{-m}\Omega)|^2 \leq \sum_{m=0}^{\infty} \left(\frac{1}{2^{\gamma(m-1)} + p} - \frac{1}{2^{\gamma m} + p} \right)$$
$$= \frac{1}{2^{-\gamma} + p}, \qquad (A.6)$$

where the equality arises from the summation of the telescoping sum.

Since $\Phi(\Omega)$ decays at least as fast as $1/\Omega$ and is bounded, there exists a constant $C \geq 1$ such that

$$|\Phi(\Omega)| \leq \frac{C}{1 + |\Omega|}. \qquad (A.7)$$

Using (A.7), together with (A.6) and the expression in (A.5), we obtain an upper bound for $S\left(\Omega_0; \frac{p}{2^{\gamma m_0}}\right)$:

$$S\left(\Omega_0; \frac{p}{2^{\gamma m_0}}\right) \leq \frac{1}{2^{-\gamma} + p/2^{\gamma m_0}}$$
$$+ \sum_{m=1}^{\infty} \left(\frac{1}{2^{-\gamma(m+1)} + p/2^{\gamma m_0}} - \frac{1}{2^{-\gamma m} + p/2^{\gamma m_0}} \right) C^2 2^{-2m}$$
$$= \frac{1}{2^{-\gamma} + p/2^{\gamma m_0}}$$
$$+ C^2(1 - 2^{-\gamma}) \sum_{m=1}^{\infty} \frac{2^{-\gamma m} 2^{-2m}}{(2^{-\gamma(m+1)} + p/2^{\gamma m_0})(2^{-\gamma m} + p/2^{\gamma m_0})} \qquad (A.8)$$

Since $\gamma \geq 0$,

$$2^{-\gamma m} + \frac{p}{2^{\gamma m_0}} \geq 2^{-\gamma m} + 2^{-\gamma(m+1)} \frac{p}{2^{\gamma m_0}} \qquad (A.9)$$

78

for $m \geq 0$. Using (A.9) in (A.8), we get that

$$
S\left(\Omega_0; \frac{p}{2^{\gamma m_0}}\right) \leq \frac{1}{2^{-\gamma} + p/2^{\gamma m_0}}
$$
$$
+ C^2(1 - 2^{-\gamma}) \sum_{m=1}^{\infty} \frac{2^{-\gamma m}2^{-2m}}{(2^{-\gamma(m+1)})(2^{-\gamma m} + 2^{-\gamma(m+1)}p/2^{\gamma m_0})}
$$
$$
= \frac{1}{2^{-\gamma} + p/2^{\gamma m_0}} + \frac{C^2(1 - 2^{-\gamma})}{2^{\gamma} + p/2^{\gamma m_0}}2^{2\gamma} \sum_{m=1}^{\infty} 2^{(\gamma-2)m}
$$
$$
= \frac{1}{2^{-\gamma} + p/2^{\gamma m_0}} + \frac{C^2(1 - 2^{-\gamma})}{2^{\gamma} + p/2^{\gamma m_0}}2^{2\gamma}\frac{2^{\gamma-2}}{1 - 2^{\gamma-2}}.
$$

So,

$$
(2^{\gamma} + \frac{p}{2^{\gamma m_0}})S\left(\Omega_0; \frac{p}{2^{\gamma m_0}}\right) \leq 2^{2\gamma}\left(1 + C^2(1 - 2^{-\gamma})\frac{2^{\gamma-2}}{1 - 2^{\gamma-2}}\right) < \infty,
$$

and (A.3) is thus established.

Proceeding next to show (A.4), we first note that since $\Phi(\Omega)$ is continuous at 0 and $|\Phi(0)| = 1$, there exists an integer k_0 such that $|\Phi(\Omega)| > 1/2$ whenever $|\Omega| < 2^{-k_0}$. Hence, choosing $n_0 = \max(0, k_0)$, we have that

$$
|\Phi(2^{-n_0-1}\Omega_0)| > 1/2
$$

from which the lower bound

$$
S\left(\Omega_0; \frac{p}{2^{\gamma m_0}}\right) = \sum_{m}\left(\frac{1}{2^{\gamma(m-1)} + p/2^{\gamma m_0}} - \frac{1}{2^{\gamma m} + p/2^{\gamma m_0}}\right)|\Phi(2^{-m}\Omega_0)|^2
$$
$$
\geq \sum_{m=n_0+1}^{\infty}\left(\frac{1}{2^{\gamma(m-1)} + p/2^{\gamma m_0}} - \frac{1}{2^{\gamma m} + p/2^{\gamma m_0}}\right)|\Phi(2^{-m}\Omega_0)|^2
$$
$$
\geq \frac{1}{4}\sum_{m=n_0+1}^{\infty}\left(\frac{1}{2^{\gamma(m-1)} + p/2^{\gamma m_0}} - \frac{1}{2^{\gamma m} + p/2^{\gamma m_0}}\right) \qquad \text{(A.10)}
$$

Summing the telecoping sum in (A.10), we obtain

$$
S\left(\Omega_0; \frac{p}{2^{\gamma m_0}}\right) \geq \frac{1}{4}\left(\frac{1}{2^{\gamma n_0} + p/2^{\gamma m_0}} - \lim_{m \to \infty} \frac{1}{2^{\gamma m} + p/2^{\gamma m_0}}\right)
$$
$$
= \frac{1}{4}\frac{1}{2^{\gamma n_0} + p/2^{\gamma m_0}},
$$

and it follows that

$$
\begin{aligned}
(1^\gamma + \frac{p}{2^{\gamma m_0}}) S\left(\Omega_0; \frac{p}{2^{\gamma m_0}}\right) &\geq \frac{1}{4}(1 + \frac{p}{2^{\gamma m_0}}) \frac{1}{2^{\gamma n_0}(1 + p/2^{\gamma m_0})} \\
&= 2^{-(2+\gamma n_0)} > 0.
\end{aligned}
$$

Hence (A.4). $\square$

Appendix B

Derivations of Algorthms in Chapter 4

B.1 EM Algorithm for Parameter Estimation for Higher-Order Generalized Fractals

Each iteration of the EM algorithm consists of two main steps, the E step and the M step. In the E step, an objective function is set up based on the current estimates of the parameters. In the M step, this objective function is maximized to obtain new estimates. In this section, we derive the objective function that must be maximized in each iteration of our EM-based ML parameter estimation algorithm. Note that although we assume that the wavelet transforms r_n^m and $(x_i)_n^m$ are zero-mean Gaussian random variables, we are not restricting this derivation to any specific form of the variance of r_n^m and $(x_i)_n^m$.

We define the incomplete data or the observed data to be

$$\mathbf{r} = \{r_n^m; \quad m, n \in \mathcal{R}\},$$

and the complete data to be $(\mathbf{x}, \mathbf{r})$, where

$$\mathbf{x} = \{(x_i)_n^m; \quad m, n \in \mathcal{R}, \ i = 1, \ldots, L - 1\}.$$

The variance of $(x_i)_n^m$ will be represented by λ_i^m.

Our task in the E step is to set up the function

$$U(\Theta, \tilde{\Theta}) \triangleq \mathrm{E}\{\ln p_{\mathbf{r},\mathbf{x}}(\mathbf{r},\mathbf{x};\Theta)|\mathbf{r};\tilde{\Theta}\}, \tag{B.1}$$

where Θ is the vector of parameters $(\beta_1, p_1, \sigma_1^2, \ldots, \beta_L, p_L, \sigma_L^2)$. This function is then maximized as a function of Θ in the M step.

Noting that

$$p_{\mathbf{r},\mathbf{x}}(\mathbf{r},\mathbf{x};\Theta) = p_{\mathbf{r}|\mathbf{x}}(\mathbf{r}|\mathbf{x};\Theta) \cdot p_{\mathbf{x}}(\mathbf{x};\Theta),$$

we obtain, from (B.1),

$$U(\Theta, \tilde{\Theta}) = \mathrm{E}\{\ln p_{\mathbf{r}|\mathbf{x}}(\mathbf{r}|\mathbf{x};\Theta) + \ln p_{\mathbf{x}}(\mathbf{x};\Theta)|\mathbf{r};\tilde{\Theta}\}. \tag{B.2}$$

Now, $(x_i)_n^m$ are independent zero-mean Gaussian random variables. This implies that

$$p_{\mathbf{x}}(\mathbf{x};\Theta) = \prod_{i=1}^{L-1} \prod_{m,n\in\mathcal{R}} \frac{1}{\sqrt{2\pi\lambda_i^m}} \exp -\frac{[(x_i)_n^m]^2}{2\lambda_i^m}, \tag{B.3}$$

and

$$\begin{aligned}
p_{\mathbf{r}|\mathbf{x}}(\mathbf{r}|\mathbf{x};\Theta) &= \prod_{m,n\in\mathcal{R}} \frac{1}{\sqrt{2\pi\lambda_L^m}} \exp -\frac{\left(r_n^m - \sum_{i=1}^{L-1}(x_i)_n^m\right)^2}{2\lambda_L^m} \\
&= \prod_{m,n\in\mathcal{R}} \frac{1}{\sqrt{2\pi\lambda_L^m}} \exp -\frac{[(x_L)_n^m]^2}{2\lambda_L^m}.
\end{aligned} \tag{B.4}$$

Substituting (B.3) and (B.4) into (B.2), and grouping terms, we have that

$$U(\Theta, \tilde{\Theta}) = \sum_{i=1}^{L} \left(-\frac{1}{2} \sum_{m\in M} N(m) \left\{ \frac{1}{\lambda_i^m} S_i^m(\tilde{\Theta}) + \ln 2\pi\lambda_i^m \right\} \right), \tag{B.5}$$

where

$$S_i^m(\Theta) = \frac{1}{N(m)} \sum_{n\in\mathcal{N}(m)} \mathrm{E}\{[(x_i)_n^m]^2|\mathbf{r};\Theta\}.$$

Using the properties of Gaussian random variables, we get thet

$$S_i^m(\Theta) = A_i^m(\Theta) + B_i^m(\Theta)\,\hat{\sigma}_m^2,$$

where

$$A_i^m(\Theta) = \lambda_i^m\left(1 - \frac{\lambda_i^m}{\lambda_1^m + \lambda_2^m + \cdots + \lambda_L^m}\right)$$

$$B_i^m(\Theta) = \left(\frac{\lambda_i^m}{\lambda_1^m + \lambda_2^m + \cdots + \lambda_L^m}\right)^2,$$

and

$$\hat{\sigma}_m^2 = \frac{1}{N(m)}\sum_{n\in\mathcal{N}(m)} (r_n^m)^2.$$

We recognize the objective function given in (B.5) as merely a sum of L terms, each having the same algebraic form as (4.5). Moreover, each set of parameters $(\beta_i, p_i, \sigma_i^2)$ exists only in one such term. It follows, then, that the maximization of (B.5) can be achieved by performing the maximization of

$$-\frac{1}{2}\sum_{m\in M} N(m)\left\{\frac{1}{\lambda_i^m}S_i^m(\tilde{\Theta}) + \ln 2\pi\lambda_i^m\right\},$$

for $i = 1, \ldots, L$.

B.2 The Maximum-Spectral-Flatness-Measure Estimator

The expression we use for the spectral flatness measure of a function $y(t)$ with dyadic wavelet transform y_n^m is

$$\begin{aligned}
\text{sfm}\,\{y(t)\} &= \exp\left(\frac{1}{\sum_{m\in\mathcal{M}} N(m)}\sum_{m,n\in\mathcal{R}}\ln(\hat{\sigma}_m^2)\right)\Big/\left(\frac{1}{\sum_{m\in\mathcal{M}} N(m)}\sum_{m,n\in\mathcal{R}}\hat{\sigma}_m^2\right)\\
&= \exp\left(\frac{1}{\sum_{m\in\mathcal{M}} N(m)}\sum_{m\in\mathcal{M}} N(m)\ln(\hat{\sigma}_m^2)\right)\Big/\left(\frac{1}{\sum_{m\in\mathcal{M}} N(m)}\sum_{m\in\mathcal{M}} N(m)\hat{\sigma}_m^2\right)
\end{aligned}$$

where

$$\tilde{\sigma}_m^2 = \frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (y_n^m)^2.$$

In the context of the maximum-spectral-flatness-measure estimation algorithm, we view $y(t)$ as the "whitened" first-order generalized fractal process $r(t)$ with "normalized" variance, i.e.,

$$y(t) = \mathcal{W}^{-1}\{\mathcal{W}\{r(t)\}\sqrt{\frac{\beta'^m + p'}{\sigma'^2}}\},$$

The sample variance of $y_n^m(t)$ at each scale is just

$$\tilde{\sigma}_m^2 = \frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (r_n^m)^2 \frac{\beta'^m + p'}{\sigma'^2}.$$

In terms of the $\Theta' = (\beta', p', \sigma'^2)$, the numerator of the spectral flatness measure is

$$
\begin{aligned}
n(\Theta') &= \exp\left\{\frac{1}{\sum_{m \in \mathcal{M}} N(m)} \sum_{m \in \mathcal{M}} N(m) \ln\left(\frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (r_n^m)^2 \frac{\beta'^m + p'}{\sigma'^2}\right)\right\} \\
&= \exp\left\{-\ln \sigma'^2 + \frac{1}{\sum_{m \in \mathcal{M}} N(m)} \sum_{m \in \mathcal{M}} N(m) \ln\left(\frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (r_n^m)^2 (\beta'^m + p')\right)\right\} \\
&= \frac{1}{\sigma^2} \exp\left\{\frac{1}{\sum_{m \in \mathcal{M}} N(m)} \sum_{m \in \mathcal{M}} N(m) \ln\left(\frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (r_n^m)^2 (\beta'^m + p')\right)\right\},
\end{aligned}
$$

(B.6)

and the denominator is

$$
\begin{aligned}
d(\Theta') &= \frac{1}{\sum_{m \in \mathcal{M}} N(m)} \sum_{m \in \mathcal{M}} N(m) \frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (r_n^m)^2 \frac{\beta'^m + p'}{\sigma'^2} \\
&= \frac{1}{\sigma^2} \frac{1}{\sum_{m \in \mathcal{M}} N(m)} \sum_{m,n \in \mathcal{R}} (r_n^m)^2 (\beta'^m + p').
\end{aligned}
$$

(B.7)

Combining (B.6) and (B.7), we get

$$\mathcal{F}(\Theta') \stackrel{\Delta}{=} \text{sfm}\{y(t)\}$$

84

$$= \frac{\exp\left\{\dfrac{1}{\sum_{m\in\mathcal{M}} N(m)} \sum_{m\in\mathcal{M}} N(m)\ln\left(\dfrac{1}{N(m)}\sum_{n\in\mathcal{N}(m)} (r_n^m)^2(\beta'^m + p')\right)\right\}}{\dfrac{1}{\sum_{m\in\mathcal{M}} N(m)} \sum_{m,n\in\mathcal{R}} (r_n^m)^2(\beta'^m + p')}$$

$$= \frac{\exp\left\{\dfrac{1}{\sum_{m\in\mathcal{M}} N(m)} \sum_{m\in\mathcal{M}} N(m)\ln\left(\hat{\sigma}_m^2(\beta'^m + p')\right)\right\}}{\dfrac{1}{\sum_{m\in\mathcal{M}} N(m)} \sum_{m\in\mathcal{M}} N(m)\hat{\sigma}_m^2(\beta'^m + p')} \tag{B.8}$$

where

$$\hat{\sigma}_m^2 = \frac{1}{N(m)} \sum_{n\in\mathcal{N}(m)} (r_n^m)^2.$$

Taking the logarithm of (B.8), and simplifying, we see that the objective function whose maximum yields the desired parameter estimates is

$$U(\beta', p') = \sum_{m\in\mathcal{M}} N(m)\ln(\beta'^m + p') - \left(\sum_{m\in\mathcal{M}} N(m)\right) \cdot \ln\left(\sum_{m\in\mathcal{M}} N(m)\hat{\sigma}_m^2(\beta'^m + p')\right).$$

Bibliography

[1] I. Daubechies. "Orthonormal Bases of Compactly Supported Wavelets". *Commun. Pure Applied Math.*, 41:909–996, November 1988.

[2] M. Feder and E. Weinstein. "Parameter Estimation of Superimposed Signals Using the EM Algorithm". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(4):477–489, April 1988.

[3] M. Fleischmann, D. J. Tildesley, and R. C. Ball, editors. *Fractals in the Natural Sciences*. Princeton University Press, New Jersey, 1989.

[4] J. A. Goff and T. H. Jordan. "Stochastic Modeling of Seafloor Morphology: Inversion of Sea Beam Data for Second-Order Statistics". *Journal of Geophysical Research*, 93(B11):13589–13608, November 1988.

[5] N. Jayant and P. Noll. *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

[6] M. S. Keshner. "$1/f$ noise". *Proceedings of the IEEE*, 70:212–218, March 1982.

[7] N. M. Laird, A. P. Dempster, and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm". *Ann. Roy. Stat. Soc.*, pages 1–38, December 1977.

[8] D. G. Luenberger. *Linear and nonlinear programming*. Addison-Wesley, Reading, Mass., 1984.

[9] S. G. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):675–693, July 1989.

[10] B. B. Mandelbrot. *Fractal Geometry of Nature*. W. H. Freeman, San Francisco, 1982.

[11] A. V. Oppenheim and R. W. Schafer. *Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, N.J., 1975.

[12] A. Papoulis. *Probability, random variables, and stochastic processes*. McGraw-Hill, New York, 2nd edition, 1984.

[13] H. Peitgen, D. Saupe, and M. F. Barnsley. *The Science of Fractal Images*. Springer-Verlag, New York, 1988.

[14] S. Peleg, J. Naor, R. Hartley, and D. Avnir. "Multiple Resolution Texture and Analysis and Classification". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(4):518–523, July 1984.

[15] O. Rioul and M. Vetterli. "Wavelets and Signal Processing". *IEEE Signal Processing Magazine*, 8(4):14–38, October 1991.

[16] R. Saletti. "A Comparison between Two Methods to Generate $1/f^\gamma$ Noise". *Proceedings of the IEEE*, 74(11):1595–1596, November 1986.

[17] G. Strang. *Wavelets and dilation equations: a brief introduction*. Center for Intelligent Control Systerms, Providence, Rhode Island, 1989.

[18] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part 1*. Wiley, New York, 1968-71.

[19] P. P. Vaidyanathan. "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial". *Proceedings of the IEEE*, 78(1):56–93, January 1990.

[20] G. W. Wornell. "A Karhunen-Loève-like Expansion for $1/f$ Processes via Wavelets". *IEEE Transactions on Information Theory*, 36(4):859–861, July 1990.

[21] G. W. Wornell. *Synthesis, Analysis, and Processing of Fractal Signals*. PhD thesis, M.I.T., Cambridge, Massachusetts, September 1991.

[22] G. W. Wornell and A. V. Oppenheim. "Estimation of Fractal Signals from Noisy Measurements using Wavelets". *IEEE Transactions on Signal Processing*, 40(3):611–623, March 1992.