MULTIFRACTAL SIGNAL MODELS WITH APPLICATION TO NETWORK TRAFFIC

Matthew S. Crouse, Rudolf H. Riedi, Vinay J. Ribeiro, and Richard G. Baraniuk

Department of Electrical and Computer Engineering,
Rice University, Houston TX
Email: {mcrouse, riedi, vinay, richb}@rice.edu

ABSTRACT

In this paper, we develop a new multiscale modeling framework for characterizing positive-valued data with long-range-dependent correlations. Using the Haar wavelet transform and a special multiplicative structure on the wavelet coefficients to ensure non-negative results, the model provides a rapid $O(N)$ algorithm for synthesizing N-point data sets. We study the second-order and multifractal properties of the model and derive a scheme for matching it to real data observations. To demonstrate its effectiveness, we apply the model to TCP network traffic synthesis. The flexibility and accuracy of the model and fitting procedure result in a close match to the real data statistics (variance-time plots) and queuing behavior.

1. INTRODUCTION

Fractals models arise frequently in a variety of scientific disciplines, such as physics, chemistry, astronomy, and biology. In DSP, fractals have long proven useful for applications such as computer graphics and texture modeling [1]. More recently, fractal models have had a major impact on data communication, particularly in the arena of data networks such as the Internet. In their landmark paper [2], Leland et al. demonstrated that network traffic exhibits fractal properties such as self-similarity, "burstiness," and long-range dependence (LRD) that are inadequately described by classical traffic models. Characterization of these fractal properties, particularly LRD, has provided exciting new insights into network behavior and performance.

Fractals are geometric objects that exhibit an irregular structure at all resolutions. Most fractals are *self-similar*; if we use a magnifying glass to "zoom" (in or out) of the fractal, we obtain a picture similar to the original. Deterministic fractals usually have a highly specific structure that can be constructed through a few simple steps. Real-world phenomena can rarely be described using such simple models. Nevertheless, "similarity on all scales" can hold in a statistical sense, leading to the notion of random fractals.

As the pre-eminent random fractal model, fractional Brownian motion (fBm) has played a central rôle in many fields [2, 3]. Fbm is the unique Gaussian process with stationary increments and the following scaling property for all $a > 0$

$$B(at) \stackrel{fd}{=} a^H B(t), \tag{1}$$

with the equality in (finite-dimensional) distribution. It follows from the definition that $B(0) = 0$ and that B and its increments are zero-mean. The parameter H, $0 < H < 1$, is known as the *Hurst parameter*. It rules the LRD of fBm, as we will see later, but it also governs its local "spikiness." In particular, for all t

$$B(t+s) - B(t) \simeq s^H, \tag{2}$$

meaning that, for $0 < H < 1$, fBm has "infinite slope" everywhere. FBm must, therefore, oscillate wildly, with the smaller the H, the "wilder" or "rougher-looking" the fBm — whence the term "fractal."

The statistical self-similarity (1) of fBm has proven most useful for signal modeling, since it efficiently captures signal features such as burstiness and LRD, while still allowing tractable theoretical analysis [2]. Nevertheless, models based on fBm can be too restrictive to adequately characterize many types of signals.

First, strictly self-similar scaling behavior is not always realistic. For example, H in (1) may vary when measured over different values of a, or H in (2) may depend on t. Second, fBm models are inherently Gaussian. Many signals have positive increments and, hence, are non-Gaussian.

For applications such as network traffic [4, 5] and turbulence analysis [6], the statistics of such signals can be more accurately characterized using a *multifractal* analysis, which describes how the signal's scaling behavior varies across the signal. In this paper, we develop a multifractal signal model especially suited to positive-valued data with LRD. Modeling of such data is not only vital for networking [2, 4], but it also provides key insights into a host of other applications such as turbulence and geophysics [6].

We call our model the *multifractal wavelet model* (MWM) because of its wavelet-domain formulation. With its simple Haar-wavelet construction, the MWM is simple to apply, yet it can characterize LRD and a number of different multifractal properties. Since we cannot treat the MWM in full detail in this paper, we refer the interested reader to [7] for a more in-depth treatment.

2. FBM AND LRD

Although we analyze fBm from a continuous-time point of view, for practical computations and simulations, we often work with sampled continuous-time fBm. The increments process of sampled fBm

$$X[n] := B(n) - B(n-1) \tag{3}$$

defines a stationary Gaussian sequence known as discrete *fractional Gaussian noise* (fGn) with covariance behavior [1]

$$r_X[k] \simeq |k|^{2H-2} \tag{4}$$

for $|k|$ large.

For $1/2 < H < 1$, the covariance of fGn is strictly positive and decays so slowly that it is non-summable (i.e.,

This work was supported by the National Science Foundation, grant no. MIP–9457438, and by DARPA/AFOSR, grant no. F49620-97-1-0513.

$\sum_k r_X[k] = \infty$). This non-summability, corresponding to positive, slowly-decaying covariances over large time lags, defines LRD. LRD is responsible for the "bursty" (fractal) appearance of fBm, even on larger scales (compare with (2)). This can be viewed as a simple consequence of (1): the process "looks the same" on all scales. Due to its LRD property, fBm with $1/2 < H < 1$ has proven most useful for signal modeling, since it permits tractable theoretical analysis due to (1).

The LRD of fBm can be equivalently characterized in terms of how the aggregated processes

$$X^{(m)}(n) := \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(i) \tag{5}$$

behave. It follows from (1) that $X(n) \stackrel{fd}{=} m^{1-H} X^{(m)}(n)$.

Hence, a log-log plot of the variance of $X^{(m)}(n)$ as a function of m — known as a *variance-time plot* — will have a slope of $2 - 2H$. The variance-time plot can characterize LRD in non-Gaussian, non-zero-mean data as well [2].

3. WAVELETS AND LRD

The discrete wavelet transform is a multiscale signal representation of the form [8]

$$\begin{aligned} x(t) &= \sum_k u_k \, 2^{-J_0/2} \, \phi\big(2^{-J_0}t - k\big) \; + \\ & \sum_{j=-\infty}^{J_0} \sum_k w_{j,k} \, 2^{-j/2} \, \psi\big(2^{-j}t - k\big), \qquad j,k \in \mathbb{Z} \end{aligned} \tag{6}$$

with J_0 the coarsest scale and u_k and $w_{j,k}$ the scaling and wavelet coefficients. The scaling coefficients may be viewed as providing a coarse approximation of the signal, with the wavelet coefficients providing higher-frequency "detail" information.

Wavelets serve as an approximate Karhunen-Loève transform for fBm [3], fGn, and more general LRD signals [9]. Thus, highly-correlated, LRD signals become nearly uncorrelated in the wavelet domain. In addition, the energy of the wavelet coefficients of continuous-time fBm decays with scale according to a power law [3]. While for *sampled* fBm the power-law decay is not exact [3], the *Haar* wavelet transform of fGn exhibits power-law scaling of the form[1] [9]

$$\operatorname{var}(W_{j,k}) = \sigma^2 \, 2^{(2H-1)(j-1)} \, (2 - 2^{2H-1}). \tag{7}$$

4. THE MWM

The basic idea behind the multifractal wavelet model is simple. To model non-negativity, we use the Haar wavelet transform with special wavelet-domain constraints. To capture LRD, we characterize the wavelet energy decay as a function of scale.

4.1. Haar Wavelets and Non-Negative Data

Before we can model non-negative signals using the wavelet transform, we must develop conditions on the scaling and wavelet coefficient values for $x(t)$ in (7) to be non-negative.

[1] We use capital letters when we consider the underlying signal X (and, hence, its wavelet and scaling coefficients) to be random.

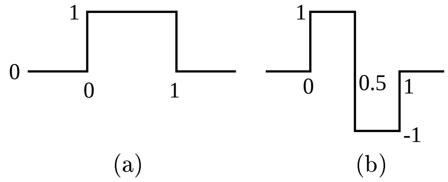


Figure 1: *(a) The Haar scaling function ϕ and (b) the Haar wavelet function ψ.*

While cumbersome for a general wavelet system, these conditions are simple for the Haar system. In a Haar transform (see Figure 1), the scaling and wavelet coefficients can be recursively computed using

$$\begin{aligned} u_{j+1,k} &= 2^{-1/2}(u_{j,2k} + u_{j,2k+1}) \\ w_{j+1,k} &= 2^{-1/2}(u_{j,2k} - u_{j,2k+1}). \end{aligned} \tag{8}$$

Solving (8) for $u_{j,2k}$ and $u_{j,2k+1}$ we find

$$\begin{aligned} u_{j,2k} &= 2^{-1/2}(u_{j+1,k} + w_{j+1,k}) \\ u_{j,2k+1} &= 2^{-1/2}(u_{j+1,k} - w_{j+1,k}). \end{aligned} \tag{9}$$

For non-negative signals, $u_{j,k} \geq 0, \; \forall \; j,k$, which with (9) implies that

$$|w_{j,k}| \leq u_{j,k}, \qquad \forall \; j,k. \tag{10}$$

4.2. Multiplicative Model

The positivity constraints (10) on the Haar wavelet coefficients lead us to a very simple multiscale, multiplicative signal model for positive processes. (See [10] for a similar model used as an intensity prior for wavelet-based image estimation.) Let $A_{j,k}$ be a random variable supported on the interval $[-1,1]$ and define the wavelet coefficients recursively by

$$W_{j,k} = A_{j,k} \; U_{j,k}. \tag{11}$$

Together with (9) we obtain

$$\begin{aligned} U_{j,2k} &= 2^{-1/2}(1 + A_{j+1,k})\, U_{j+1,k} \\ U_{j,2k+1} &= 2^{-1/2}(1 - A_{j+1,k})\, U_{j+1,k}. \end{aligned} \tag{12}$$

The above construction can be visualized as a course-to-fine synthesis (see Figure 2). Starting from the coarsest scale $j = J_0$, we can synthesize a realization of a process by iteratively applying (11) to obtain the wavelet coefficients at scale j and then applying (8) to obtain the scaling coefficients at the next finest scale $j-1$. In essence the algorithm simultaneously synthesizes the wavelet coefficients and inverts the wavelet transform, requiring only $O(N)$ operations to create a length-N signal.

This construction has two attractive features. First, we ensure that $|W_{j,k}| \leq U_{j,k}$ and, hence, that the corresponding process is non-negative. Second, by specifying the densities for the $A_{j,k}$'s, we can model the time-domain LRD or covariance structure of a signal through the energy decay of its wavelet coefficients with scale j. This modeling would be exact if the decorrelation property of wavelets were exact. Typically, the correlation of LRD between the wavelet coefficients of LRD processes are small [9], and therefore we can approximate the time-domain behavior of such LRD processes quite accurately.

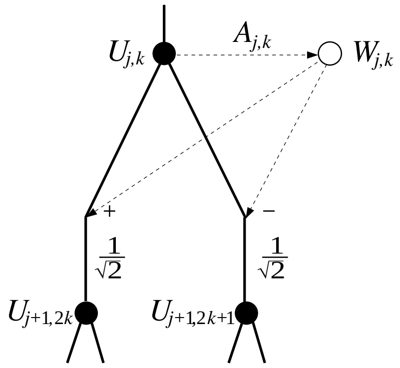


Figure 2: *MWM construction: At scale j, we form the wavelet coefficient as the product $W_{j,k} = A_{j,k}U_{j,k}$. Then, at scale $j-1$, we form the scaling coefficients $U_{j-1,2k}$ and $U_{j-1,2k+1}$ as sums and differences of $U_{j,k}$ and $W_{j,k}$ (normalized by $1/\sqrt{2}$).*

4.3. β multipliers

What remains is to choose the distribution for the multiplier $A_{j,k}$. First, we will assume that $A_{j,k}$ is independent of $U_{j,k}$. Second, we will assume that $A_{j,k}$ is symmetric about 0; it is easily shown this symmetry is necessary for the resulting process to be stationary [7].

Because of its simplicity and flexibility, we will use a symmetric beta distribution, $\beta(p,p)$, for the $A_{j,k}$'s and hence christen the resulting model the βMWM. The parameter p is a shape factor, with $p = 1$ giving rise to the uniform density on $(-1,1)$, $0 < p < 1$ corresponding to convex densities on $(-1,1)$, and $p > 1$ leading to concave densities on $(-1,1)$. The variance of a random variable $A \sim \beta(p,p)$ is

$$\mathrm{var}[A] = \frac{1}{2p+1}. \tag{13}$$

Since we assume the multipliers are identically distributed within scale, i.e., $A_{j,k} \sim \beta(p_j, p_j)$, for our Haar wavelet construction we can control the wavelet energy decay across scale via

$$\frac{\mathrm{var}(W_{j+1,k})}{\mathrm{var}(W_{j,k})} = \frac{2\ \mathrm{var}\left[A_{j+1,k}\right]}{\mathrm{var}\left[A_{j,k}\right]\ \left(1+\mathrm{var}\left[A_{j+1,k}\right]\right)}. \tag{14}$$

Equations (13) and (14) provide a link between the LRD or covariance behavior of a signal and the βMWM parameters, which consist of the shape parameters p_j, one for each scale j. To model a given process with the βMWM, we can select the parameters via (13) and (14) to match the signal's theoretical wavelet-domain energy decay, such as (7). Or, given training data, we can select the parameters to match the sample variances of the wavelet coefficients as a function of scale.

Of course, this analysis addresses only the second-order statistics of our signal. Higher-order properties of the MWM are the subject of a multifractal analysis.

5. THE MWM IS A MULTIFRACTAL

Multifractals offer a wealth of processes that are novel in many respects. The backbone of a multifractal is typically, but not necessarily, a construction where one starts at a coarse scale and develops details of the process on finer scales iteratively in a multiplicative fashion. It follows from (12) that the MWM is a *binomial cascade*,[2] one of the simplest multifractals. The name binomial cascade is explained by applying (12) iteratively and writing $U_{j,k}$ as the product of the coarsest scale $U_{J_0,k}$ and the multipliers $2^{-1/2}(1 + A_{j+1,k})$.

Multiplicative structures, in particular the product representation of $U_{j,k}$, bear various consequences. First, if all multipliers $1 + A_{j+1,k}$ in (12) are log-normal, then the marginals $U_{j,k}$ will be log-normal as well. Similarly, if the $1 + A_{j,k}$ are all identically distributed, $U_{j,k}$ will be approximately log-normal by the central limit theorem.

Second, interpreting $U_{j,k}$ as the increment of a limiting process Y over the interval $[k2^{-n}, (k+1)2^{-n}]$, we find for Y a local behavior of the type (2). To see this, note that $\log|U_{j,k}|/\log 2^j$ can be written as the sum of approximately j factors of the form $\log_2 2^{-1/2}(1 \pm A_{l,m})$ normalized by $1/j$. So, we expect this number to converge to some limiting value H. It is essential to note, however, that this value H depends now on t — whence the term *multifractal* for Y. A rigorous argument for this fact is too involved to be included here. However, an intuitive reason is as follows.

Due to the iterative construction, a binomial cascade Y will display a "scale invariance" similar to fBm in (1), with the difference that the binary construction now produces *two* terms on the right. If we assume that all the $A_{j,k}$ are i.i.d. then

$$Y(t) \overset{fd}{=} Y_1(t)(1 + A_1) + Y_2(t)(1 - A_2), \tag{15}$$

where $Y_1(t), Y_2(t)$ are mutually independent and identically distributed to $Y(t)$, and A_1, A_2 are mutually independent and identically distributed to the $A_{j,k}$. In contrast to (1), the scaling involves now *two* versions of the process with random scaling factors. This explains both the more involved local behavior of Y and the greater versatility of the MWM.

As a further feature of interest, depending on the moments of the multipliers, the marginals $U_{j,k}$ of binomial cascades may have diverging moments of order q larger than some q_{crit}, where q_{crit} can be arbitrarily large. This broadens the realm of "heavy tailed" processes considerably.

The exact multifractal properties of the MWM are studied in detail in [7].

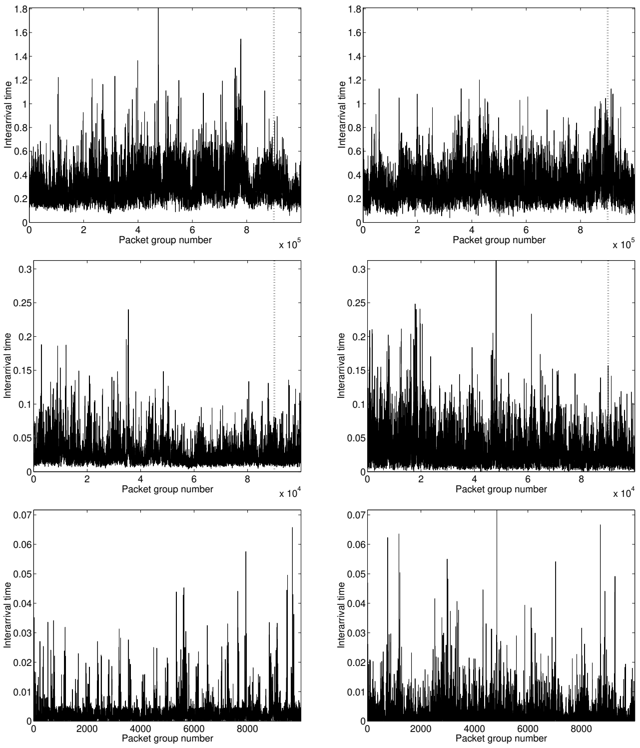
6. APPLICATION TO NETWORK TRAFFIC

We demonstrate the power of the βMWM for a problem of considerable practical interest — network traffic modeling. For concreteness, we focus on the August 1989 Bellcore Ethernet trace *pAug*, a record of one-million inter-arrival times (Figure 3(a)), as measured by Leland et al. [2]. From Figure (3)(a), it is clear that the data has non-Gaussian marginals contradictory to the hypothesis of an fBm or fGn model.

We analyze the LRD properties of the trace by estimating the variance-time plot, as shown in Figure 4(a). Although the data exhibits LRD (average slope corresponding to $H = 0.8$), the data does not appear to be strictly second-order self-similar, as evidenced by the "kink" in the slope. Again, an fBm or fGn model would be somewhat inaccurate.

We now model this data using the βMWM. To train the βMWM, we split the Bellcore inter-arrival times into a series of 2^{16} length blocks, take a 16-scale Haar DWT

[2] For a detailed introduction to binomial cascades and multifractal theory, see [4, 7, 11].



(a) (b)

Figure 3: *Inter-arrival times of groups of packets of (a) Bellcore August 1989 [2] and (b) one realization of the beta multifractal wavelet model (βMWM) synthesis at different three resolutions, from coarsest (top) to finest (bottom).*

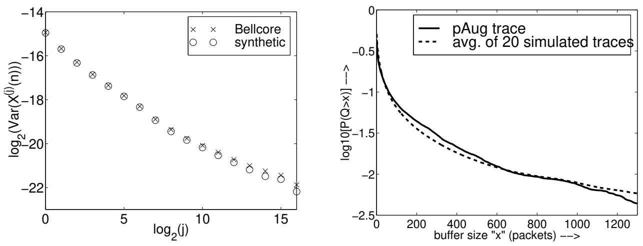
for each block (truncating any leftover data), and calculate statistics for the scaling and wavelet coefficients. Using (13) and (14), we choose the $\beta(p_j, p_j)$ distribution used at each scale j so that the theoretical variances of our synthesized wavelet coefficients match the measured variances of the Bellcore wavelet data.

In Figure 3 and Figure 4(a), we see that the synthesized data captures much of the gross structure of the Bellcore data, both in terms of marginal densities (definitely non-Gaussian) and of LRD, as evidenced through the variance-time plot. In [7], we further provide an empirical multifractal analysis of the synthesized data and compare it that of the Bellcore data.

To assess the accuracy and usefulness of the βMWM for traffic modeling, we compare the queuing behavior of the simulated traffic traces against that of the actual Bellcore data. In our simulation experiments, we consider the performance of an infinite length single server queue with a single trace as input (see [7] for more details). In Figure 4(b) the queuing performance of the real trace and those synthesized by the βMWM are roughly equivalent, demonstrating the potential usefulness of the βMWM for network modeling and simulation.

7. CONCLUSIONS

The multiplicative wavelet model (MWM) combines the power of multifractals with the efficiency of the wavelet transform to form a flexible framework natural for characterizing and synthesizing positive-valued data with LRD. As our numerical experiments have shown, the MWM is particularly suited to the analysis and synthesis of TCP network traffic data. In addition, the model could find application in areas as diverse as financial time-series characterization, geophysics (using 2-d and 3-d wavelets), and texture modeling. The parameters of the MWM are simple enough to be easily inferred from observed data or chosen a priori. Computations involving the MWM are extremely efficient — synthesis of a trace of N sample points requires only $O(N)$ computations. Finally, several extensions to the MWM are straightforward. The choice of β-distributed wavelet multipliers $A_{j,k}$ is not essential. Alternatively, we can employ mixtures of β's or even purely discrete distributions to fit higher-order multifractal moments.



(a) (b)

Figure 4: *(a) Variance-time plot of the Bellcore data "×" and one realization of the βMWM synthesis "∘". (b) Comparison of the queuing behavior of the Bellcore pAug with that of 20 synthesized βMWM traces. Displayed are the tail probabilities of buffer occupancy against buffer size. A constant service rate of 500 packets/sec is assumed.*

REFERENCES

[1] T. Lundahl, W. Ohley, S. Kay, and R. Siffert, "Fractional Brownian motion: A maximum likelihood estimator and its application to image texture," *IEEE Trans. on Medical Imaging*, vol. 5, pp. 152–161, Sep. 1986.

[2] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, pp. 1–15, 1994.

[3] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–916, Mar. 1992.

[4] R. Riedi and J. L. Véhel, "Multifractal properties of TCP traffic: A numerical study," *Technical Report No 3129, INRIA Rocquencourt, France*, Feb, 1997. Available at `www.dsp.rice.edu/~riedi`.

[5] J. L. Véhel and R. Riedi, "Fractional Brownian motion and data traffic modeling: The other end of the spectrum," *Fractals in Engineering*, pp. 185–202, Springer 1997.

[6] B. B. Mandelbrot, "Intermittent turbulence in self similar cascades: Divergence of high moments and dimension of the carrier," *J. Fluid. Mech.*, vol. 62, p. 331, 1974.

[7] R. H. Riedi, M. S. Crouse, V. Ribieri, and R. G. Baraniuk, "A multifractal wavelet model with application to TCP network traffic," *Submitted to IEEE Trans. Inform. Theory*, 1998. Available at www.dsp.rice.edu.

[8] I. Daubechies, *Ten Lectures on Wavelets*. New York: SIAM, 1992.

[9] L. M. Kaplan and C.-C. J. Kuo, "Extending self-similarity for fractional Brownian motion," *IEEE Trans. Signal Proc.*, vol. 42, pp. 3526–3530, Dec. 1994.

[10] K. E. Timmerman and R. D. Nowak, "Multiscale Bayesian estimation of Poisson intensities," in *Proc. 31st Asilomar Conf.*, (Pacific Grove, CA), Nov. 1997.

[11] R. H. Riedi and I. Scheuring, "Conditional and relative multifractal spectra," *Fractals. An Interdisciplinary Journal*, vol. 5, no. 1, pp. 153–168, 1997.