Scaling analysis of conservative cascades, with applications to network traffic

A. C. Gilbert, W. Willinger, Member, IEEE, and A. Feldmann

Abstract

Recent studies have demonstrated that measured wide-area network traffic such as Internet traffic exhibits locally complex irregularities, consistent with multifractal behavior. It has also been shown that the observed multifractal structure becomes most apparent when analyzing measured network traffic at a particular layer in the well-defined protocol hierarchy that characterizes modern data networks, namely the transport or TCP layer. To investigate this new scaling phenomenon associated with the dynamics of measured network traffic over small time scales, we consider a class of multiplicative processes, the so-called conservative cascades, that serves as a cascade paradigm for and is motivated by the networking application. We present a wavelet-based time/scale analysis of these cascades to determine rigorously their global and local scaling behavior. In particular, we prove that for the class of multifractals generated by these conservative cascades the multifractal formalism applies and is valid, and we illustrate some of the wavelet-based techniques for inferring multifractal scaling behavior by applying them to a set of wide-area traffic traces.

Keywords: Conservative cascades, multifractals, wavelets, data network traffic, TCP/IP protocol.

AT&T Labs-Research, 180 Park Avenue, Florham Park, NJ 07932, Email: {`agilbert, walter, anja`}`@research.att.com`

A. C. Gilbert was supported by the NSF Grant DMS-9705665 at Yale University, New Haven, CT, and AT&T Labs-Research, Florham Park, NJ.

W. Willinger's research was partially supported by the NSF Grant NCR-9628067 at the University of California at Santa Cruz.

1 Introduction

Cascade models or multiplicative processes make especially appealing physical models for turbulence (where they were initially introduced; see for example Mandelbrot [19], Frisch and Parisi [11], Meneveau and Sreenivasan [22], and references therein) and, more recently, for data network traffic (see Feldmann et al. [9]). A cascade is a process which fragments a given set into smaller and smaller pieces according to some geometric rule and, at the same time, divides the measure of the set according to another (possibly random) rule. The generator of the cascade specifies the mass fragmentation rule and depending upon its properties, the cascade either preserves the measure of the initial set or it does so in expectation. The limiting object generated by such a multiplicative process defines, in general, a singular measure or *multifractal* and describes the highly complex way the cascade redistributes the mass of the initial set during this fragmentation procedure.

Multifractal structures have been found in a wide variety of physical systems from turbulence [5] and rain clouds [13, 18], to data network traffic [17, 21, 9]. Multifractals provide a mathematical framework for describing local singularities, for detecting and identifying complex local structure. With time-dependent scaling laws, they are more flexible in describing locally irregular phenomena than monofractals, where the latter are governed by single scaling laws and "look the same across a wide range of scales." Exactly self-similar processes are special cases of monofractals; their degree of local irregularity is the same across all scales and across all points in time and can be captured by a single parameter, the *Hurst parameter* H.

Motivated by the original findings reported by Riedi and Levy-Vehel [17] of multifractal scaling behavior in measured wide-area network (WAN) traffic, Feldmann et al. [9] present a more detailed investigation into the multifractal nature of network traffic using wavelet-based analysis and inference tools tailor-made for a particular class of multiplicatively generated multifractals, *conservative* cascades. This class of cascades was originally introduced by Mandelbrot [20] and preserves the total mass of the initial set at each stage of the cascade construction, not only in expectation, but almost surely. Feldmann et al. bring multifractals into the realm of networking by demonstrating that (i) conservative cascades are inherent to wide-area network traffic, (ii) multiplicative structure becomes apparent when studying data traffic at the TCP layer, and (iii) the cascade paradigm appears to be a traffic invariant for WAN traffic that can co-exist with self-similarity. In fact, they systematically investigate the causes for the observed self-similar and multifractal nature of measured network traffic and identify the former to be an additive property that is mainly caused by the global characteristics of user-initiated sessions (i.e., Poisson arrivals of sessions and heavy-tailed distributions with infinite variance for the sizes or durations of each session) and manifests itself in terms of self-similar scaling behavior over a wide range of sufficiently large time scales. On the other hand, the packet arrival patterns within individual sessions, or even more so, within individual TCP connections within the individual sessions, appear to be consistent with a multiplicative structure that seems to be mainly caused by networking mechanisms operating on small time scales and results in aggregate network traffic that exhibits multifractal scaling behavior over a wide range of small time scales. Moreover, Feldmann et al. suggest that the transition from multifractal to self-similar scaling occurs around time scales on the order of the typical round-trip time of a packet within the network under consideration.

This paper is the technical counterpart to Feldmann et al. [9]. We present a number of tools for exploring the multifractal nature of data network traffic and illustrate them (both their potential and limitations) with a number of different WAN traffic traces. We discuss their development and applicability, showing that these empirical tools are, in fact, rigorous for the class of conservative cascades. We use these wavelet-based techniques to determine the global and local scaling behaviors of this class of cascades and of the limiting multifractals that they generate. Here global scaling behavior corresponds to a single scaling law which holds for all time; that is, how the cascade as a whole changes from one time scale to another. Specifically, we examine the energy contained in each level of the cascade and show that it obeys a simple scaling law, one governed by a linear relation. By local scaling behavior we mean the time-dependent scaling law which governs the intricate local behavior of the cascade. To capture this time-dependent scaling law in a compact form, we compute the multifractal spectrum of the limiting measure generated by a conservative cascade using a discrete wavelet transform-based partition function and prove that the multifractal formalism applies to this particular class of multiplicatively generated measures. We also connect the generator of the conservative cascade to both the local and global scaling behaviors via an invariant of the cascade, the *modified cumulant generating function* or *MKP function* (see Holley and Waymire [14]).

The structure of the paper is as follows. In Section 2 we discuss the general cascade construction procedure and structural properties of the resulting limiting measures such as the dimension of their supports. We also introduce the conservative cascade, discuss its relevance for networking applications, and present an inverse cascade construction as a method for verifying that a given data set conforms to an underlying conservative cascade construction. In Section 3 we present the results for the global scaling analysis for the class of multifractals generated by these conservative cascades and show how the global scaling analysis is applied to measured data network traffic. The corresponding results for their local scaling analysis are presented in Section 4, where we also prove that the wavelet techniques considered for this purpose result in statistically rigorous inference tools that allow for an effective and efficient local scaling analysis of large sets of traffic data. We conclude in Section 5 with a discussion of some of the apparent limitations of the presented wavelet-based inference techniques and with some open problems related to the analysis and inference of multifractal scaling phenomena in the networking context.

Short description of the data sets used in this paper: Throughout this paper, we use traffic traces collected from two different WAN environments. One of the traces was gathered from an FDDI ring that connects an ISP modem bank of roughly 420 modems to the rest of the Internet. The trace, referred to as DATASET 1A, was collected on July 23, 1997 between 19:02 and 23:43 and consists of a total of 12,870,502 packets and 4.212 Gbytes from/to modem users. A 1-hour long subset of this trace (from 22:00 to 23:00), referred to as DATASET 1B, contains a total of 8,719,659 packets, with 2,752,779 packets coming from modem users. To contrast this ISP traffic trace with a corporate WAN trace, the third data set, consisting of 3,903,350 packets and 1.131 Gbytes, was collected off an Ethernet connecting AT&T Labs-Research in Florham Park, NJ to the Internet via a fractional T3 connection; packets were collected on August 29, 1997 between 16:00 and 17:00. We refer to this data as DATASET 2. Note that while DATASET 1A and DATASET 1B were also used in [9], DATASET 2 has not been analyzed before.

2 Cascade construction: Properties and application

A process that fragments a set into smaller and smaller pieces according to some geometric rule and, at the same time, divides the measure of these pieces according to another rule (possibly preserving the measure) is a *multiplicative process* or a *cascade* (see Evertsz and Mandelbrot [8]). In this section we describe the general cascade construction (which Mandelbrot introduced for modeling turbulence) and several other important classes of cascades, including the well-studied deterministic cascades and the conservative cascades. We discuss the structural properties of these cascades, including the dimension of their support, and introduce the modified cumulant generating function associated with a cascade; this function will reappear as a crucial ingredient at the various stages in our scaling analysis of cascades. We also illustrate why conservative cascades arise naturally in the data network context and present an *inverse cascade construction* that provides a simple heuristic for checking whether or not a given data set conforms to an underlying conservative cascade construction. In the following, we will use the notion of conservative cascade to refer to the corresponding cascade construction as well as to the limiting measure or multifractal generated by this construction. The context in which these terms are used will resolve any potential confusion.

2.1 Random cascades

Mandelbrot first introduced the random cascade as a physical model for turbulence (see [19]). In the random cascade construction, we begin with an initial mass M distributed uniformly over the unit interval $I = [0, 1]$. We divide the unit interval into a collection of c, c^2, ..., c^l, ... subintervals; every subinterval of the lth stage we divide into c subintervals to form the $(l+1)$st stage. We denote the intervals generated by this construction process, the c-adic intervals of resolution size c^{-l}, by $I(j_1, \dots, j_l)$,

$$I(j_1, \dots, j_l) = \left[\sum_{k=1}^{l} j_k c^{-k}, \sum_{k=1}^{l} j_k c^{-k} + c^{-l} \right] \quad \text{for } j_k = \{0, \dots, c-1\} \text{ and } l = 1, 2, \dots.$$

The indices $j_1, \dots, j_l$ form the c-adic expansion of the left end-point of the interval $I(j_1, \dots, j_l)$. We assign mass $MV(0), \dots, MV(c-1)$ to the subintervals of the first stage, where V is a non-negative random variable with mean $1/c$ and the random variables $V(0), \dots, V(c-1)$ are independent and have the same distribution as V, the generator of the random cascade. Iterating this procedure generates a collection of random variables $V(j_1), \dots, V(j_1, \dots, j_l)$ indexed by the collection of intervals, all of which are independent and identically distributed as the generator V. See Figure 1 for an example of the random cascade construction with $M = 1$ and $c = 2$.

The measure of the c-adic interval $I(j_1, \dots, j_l)$ generated by this random cascade construction at stage l is given by

$$\mu_l\big(I(j_1, \dots, j_l)\big) = MV(j_1)V(j_1, j_2) \cdots V(j_1, \dots, j_l).$$

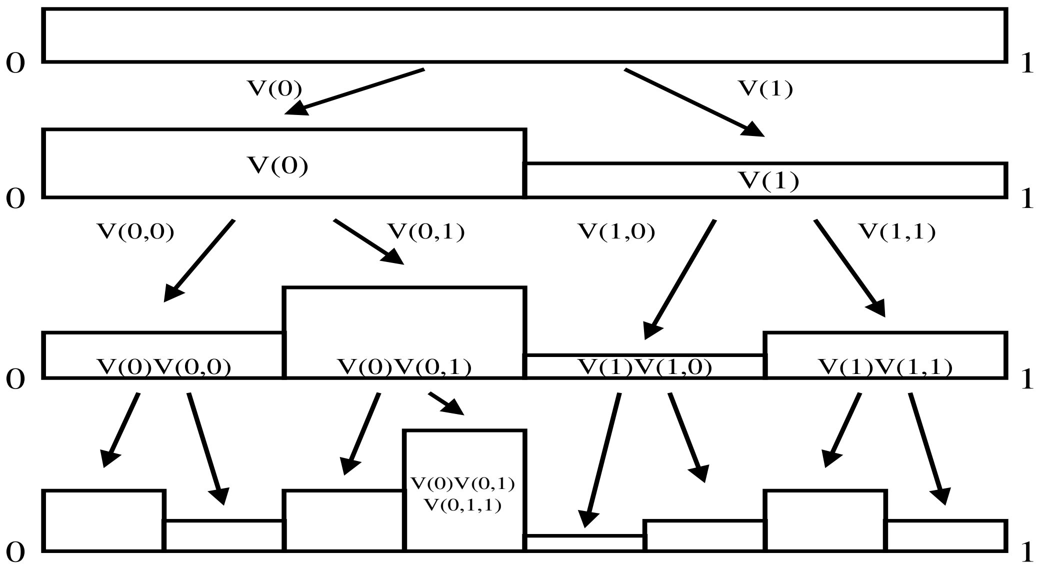

Figure 1: Cascade construction

Note that the random cascade preserves mass only in expectation; i.e., for all $l \geq 1$, we have $\mathbf{E}[\mu_l(I)] = M$. Moreover, since for every measurable set A, the sequence $\{\mu_l(A) \ : \ l = 1, 2, \ldots\}$ is an L^1-bounded martingale with respect to $\mathcal{F}_l$, the sequence of σ-fields generated by the sequence of random variables $V(j_1), \ldots, V(j_1, \ldots, j_l)$, the measures μ_l converge (weakly) to a limit measure μ_∞ (for details see for example [14]). The following theorem of Kahane and Peyrière [15] provides more detailed information about the structural properties of this limit measure generated by a random cascade and relies on the analysis of a modified cumulant generating function given by

$$\chi_c(h) = \log_c \mathbf{E}[V^h] + 1. \tag{1}$$

Theorem 2.1 *Let V denote the generator of the random cascade μ_∞ and let $\chi_c(h) = \log_c \mathbf{E}[V^h] + 1$.*

1. *If $-D = \chi_c'(1^-) = c\mathbf{E}[V \log_c V] < 0$, then $\mathbf{E}[\mu_\infty([0,1])] > 0$, and conversely.*

2. *Let $h > 1$. Then $Y = \mu_\infty([0,1])$ has a finite moment of order h if and only if $\mathbf{E}[V^h] < 1/c$. Moreover,*

$0 < \mathbf{E}[Y^h] < \infty$ *for all* $h > 0$ *if and only if* V *is essentially bounded by* c *and* $\mathrm{P}(V = c) < 1/c$.

3. *Assume that* $\mathbf{E}[Y \log Y] < \infty$*; then* μ_∞ *is supported on a Borel set of dimension*

$$D = -c\mathbf{E}[V \log_c V] = -\chi_c'(1^-).$$

2.2 Deterministic cascades

To illustrate the deterministic cascade, we choose $c = 2$ to simplify our notation and to clarify the descriptions of this important example. The deterministic cascade is a special case of the random cascade in that we assign a fixed multiple of the parent mass to each subinterval (regardless of the stage in the cascade construction). We choose a fixed $p \in (0, 1/2]$ and, at the first stage in the construction, we assign mass Mp to the left interval $I(0)$ and mass $M(1-p)$ to the right interval $I(1)$. If we iterate this procedure, the mass of the dyadic interval $I(j_1, \ldots, j_l)$ at the lth stage is an l-factor product of ps and $(1-p)$s; that is,

$$\mu\big(I(j_1, \ldots, j_l)\big) = M\left(\prod_{k=1}^{l} p_{j_k}\right) = Mp^{n_0}(1-p)^{n_1}$$

where $j_k \in \{0, 1\}$, $p_0 = p$, $p_1 = 1-p$, and where n_0 is the number of zeros in $(j_1, \ldots, j_l)$ and $n_1 = l - n_0$ is the number of ones. This process preserves the original mass M at every stage and its limiting measure μ_∞ (the binomial measure) on $I = [0, 1]$ is singular and multifractal if $p \neq 1/2$ (for details see for example [3, 25]).

2.3 Conservative cascades

For convenience, we again restrict the following discussion to the case $c = 2$. While random cascades may be appropriate physical models for turbulence, they are not appropriate in the networking context (nor is the deterministic cascade). In short, the networking context calls for a compromise between the highly flexible random cascade and the rigid deterministic cascade; it requires the mass preservation property of the deterministic rule (see Section 2.4 for details) while, at the same time, aims for inherent randomness to account for extremely heterogeneous aspects of modern data networks. To accommodate these two competing objectives of mass preservation (deterministic cascade) and fully random choice (random cascade), we define a semi-random (or conservative) rule that assigns mass MW to the interval $I(0)$ and mass $M(1-W)$ to $I(1)$. The generator W is a random variable with mean $1/2$, takes on values in $(0, 1)$, and is symmetric about its mean. To iterate this procedure, we consider a sequence of random variables $W(j_1, \ldots, j_l)$, $l \geq 1$, with a dependence structure given by

$$W(j_1, \ldots, j_{l-1}, 1) = 1 - W(j_1, \ldots, j_{l-1}, 0) \tag{2}$$

and where the random variables $W(j_1, \ldots, j_{l-1}, 0)$ and $W(j_1, \ldots, j_{l-1}, 1) = 1 - W(j_1, \ldots, j_{l-1}, 0)$ are identically distributed as W (recall that the generator W is symmetric about its mean). This process constructs a *conservative cascade*[1] and a collection of measures μ_l. For all $l \geq 1$, the measure of the dyadic interval $I(j_1, \ldots, j_l)$ is given by

$$\mu_l(I(j_1, \ldots, j_l)) = MW(j_1)W(j_1, \ldots, j_2) \cdots W(j_1, \ldots, j_l).$$

and because of its multiplicative structure, $\mu_l(I(j_1, \ldots, j_l))$ is approximately lognormal (e.g., see [21]). It is clear that for all $l \geq 1$, we have $\mu_l(I) = M$. The main difference between the random cascade and the conservative cascade is the dependence in the way the mass is distributed from the parent interval to the left and the right subintervals at each stage in the construction process. This additional dependence structure simplifies the structural properties of the limiting measure μ_∞ generated by the conservative cascade. In fact, the limiting measure μ_∞ is non-degenerate and, trivially, all moments of $\mu_\infty([0,1]) = M$ exist. We can use the more general results in Holley and Waymire [14], or Ben Nasr [23] and Waymire and Williams [28] to formalize the following corollary to Theorem 2.1 that states explicitly the structure results for the limiting measure μ_∞ generated by the conservative cascade.

Corollary 2.1 *Let W denote the (fixed) generator of the conservative cascade, let μ_∞ denote its limiting measure; set $\chi(h) = \log_2 \mathbf{E}[W^h] + 1$, and assume $M = 1$.*

1. *(Nondegeneracy) μ_∞ is nondegenerate; in fact, we have $\mu_\infty([0,1]) = 1$ almost surely.*

2. *(Support of μ_∞) μ_∞ is supported on a Borel set of dimension*

$$D = -2\mathbf{E}[W \log_2 W] = -\chi'(1^-).$$

[1] This type of cascade was referred to as a *semi-random cascade* in [9]. We have since learned that the standard term is *conservative cascade*, as coined by Mandelbrot [20].

There are, of course, many generalizations of this basic conservative cascade construction procedure. At each step in the construction process we can, for example, divide the parent interval into c subintervals instead of only two; see for example the discussion in Section 5. Furthermore, we can change the generator W at each stage of the construction and we can impose a more general dependence structure upon the measures of the subintervals. In the following sections, we will focus exclusively on the class of conservative cascades where $c = 2$ and where we allow for possible changes in the variability of the generator W at each stage of the construction.

2.4 Data networks as cascades

To argue for the relevance of cascades in the networking context, we summarize in the following the main findings reported in [9], starting with physical understanding of the self-similar nature of WAN traffic and moving on to the more recent observation of multifractal scaling behavior in measured data traces. In particular, Feldmann et al. observe in [9] that aggregate WAN traffic exhibits self-similar scaling over wide range of sufficiently large time scales, provided user-initiated sessions (e.g., TELNET, FTP, WWW) arrive in a Poisson fashion and their durations (in seconds) or sizes (in bytes or packets) have a heavy-tailed distribution with infinite variance (i.e., range from extremely short or small to extremely long or large). Note that this understanding of the observed self-similarity of aggregate WAN traffic allows one to conclude that asymptotically self-similar behavior (i) is an *additive* property (i.e., aggregate over many heavy-tailed sessions), (ii) is mainly caused by user/session characteristics (e.g., Poisson arrivals of sessions, heavy-tailed distributions with infinite variance for the session sizes), and (iii) has little to with the network; that is, on how the individual packets within a session or connection are sent over the network. In fact, whether the packets within a connection arrive at a constant rate (see Figure 2, left plot) or in a highly bursty fashion (as, for example, illustrated in Figure 2, right plot, which shows the actual traffic rate of a measured TCP connection from DATASET 1A) is irrelevant for the self-similarity property of data traffic over large time scales; for the latter to hold, all that is needed is that the number of packets or bytes per connection is heavy-tailed with infinite variance.

When trying to understand the impact of networks on the traffic that they carry, Feldmann et al. [9] provide empirical evidence that (i) the network shows up when studying traffic over small time scales, (ii) the local

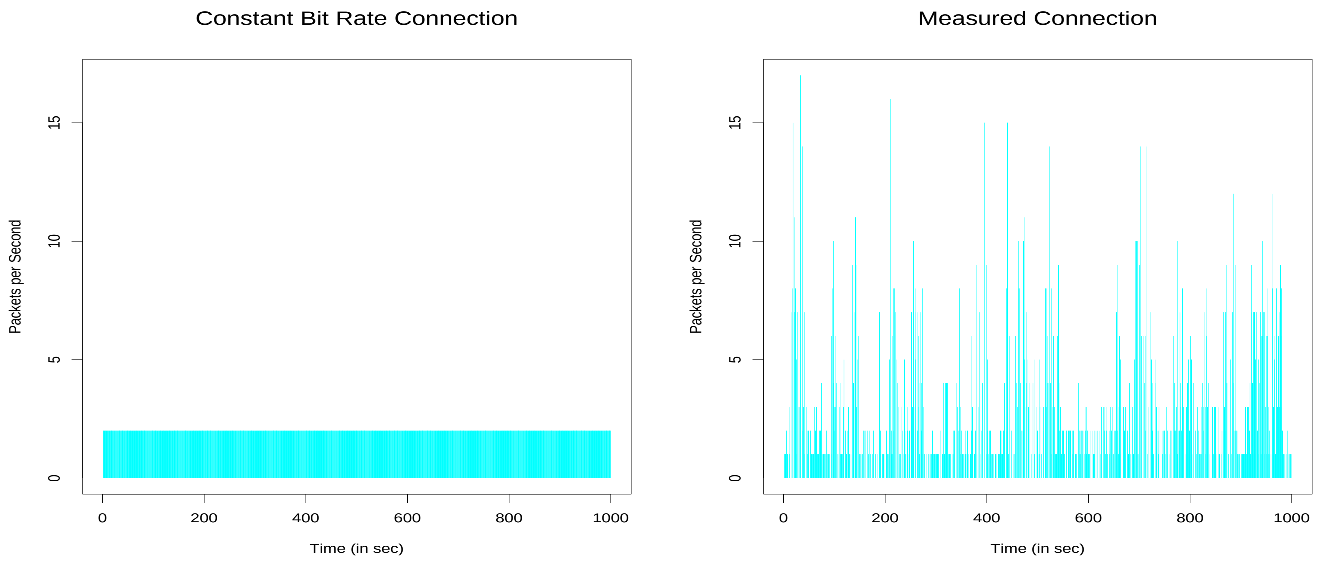

Figure 2: Constant bit-rate connection (left) vs. measured TCP connection (right); both connections have the same duration and the same size (e.g., number of packets).

properties of measured WAN traffic appear to be consistent with multifractals, (iii) multifractal scaling that has little to do with user/session characteristics but seems to be the result of the predominant protocols and end-to-end congestion control mechanisms that determine the flow of the packets at the different layers in the TCP/IP protocol suite. In particular, they suggest that the transition between multifractal and self-similar scaling occurs at time scales on the order of the typical round-trip time of a packet in the network and that measured data traffic over small time scales conforms to conservative cascades. Although it is tempting to invoke the TCP/IP protocol hierarchy of modern data networks for motivating the presence of an underlying conservative cascade construction (e.g., a web session generates requests, each request gives rise to connections, each connection is made up of flows, flows consist of individual packets), Feldmann et al. demonstrate that the multiplicative structure associated with a conservative cascade construction is most apparent when studying network traffic at the TCP layer; i.e., when analyzing the arrival patterns of packets within individual TCP connections or port-to-port flows. To illustrate this finding, we depict in Figure 3 stages 0, 1, 2, 3, 4, and 10 of a simple conservative cascade construction (where we allow the variability of the underlying generator to vary from stage to stage) that turns a constant bit-rate connection into a highly

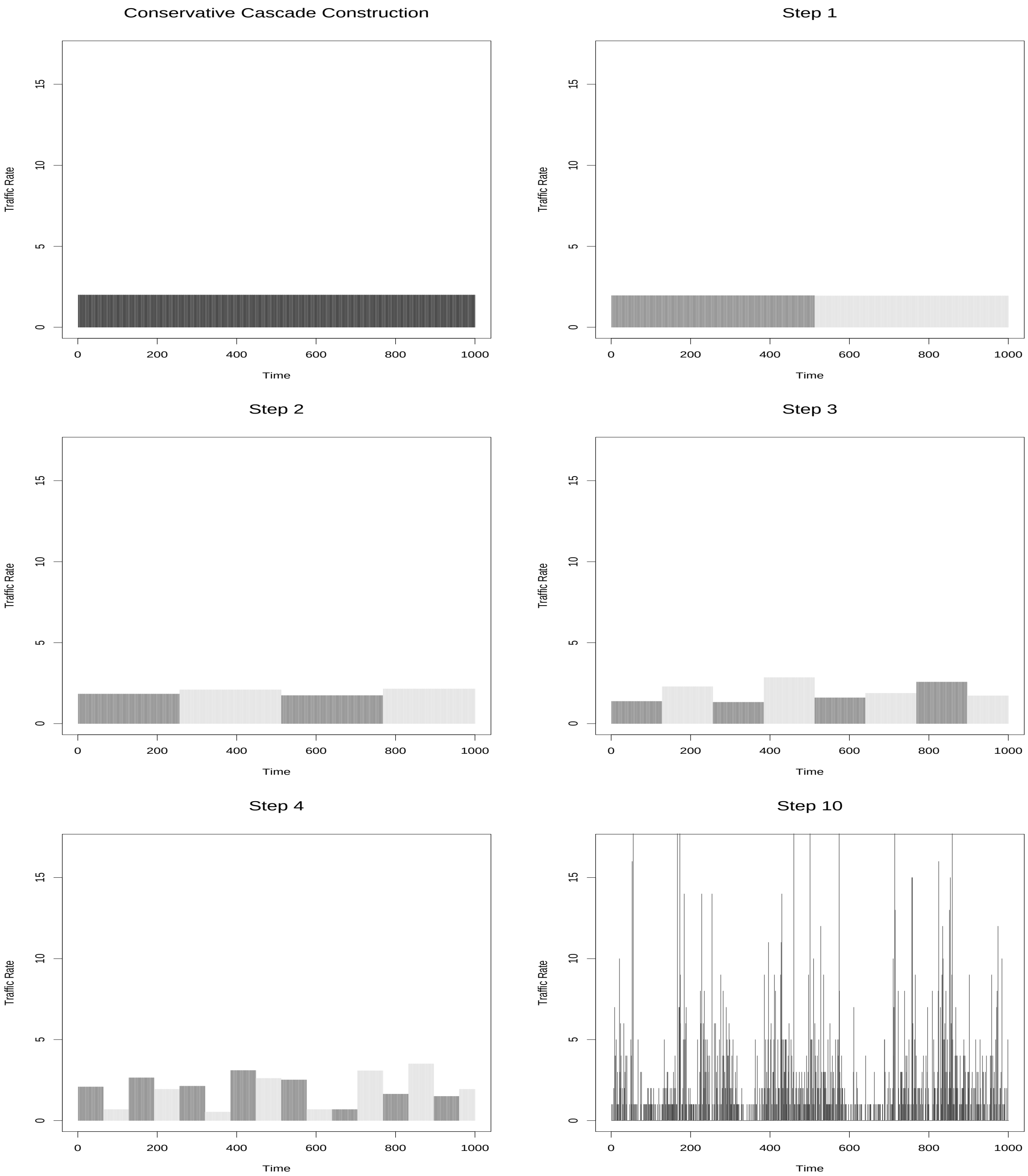

Figure 3: Steps 0,1,2,3,4, and 10 of a conservative cascade construction, starting with the constant bit-rate source in Figure 2.

bursty one (bottom right). Comparing Figures 2 (right plot) and 3 (bottom right) shows that conservative cascades can closely match the way networking mechanisms operating over small time scales (e.g., TCP) determine the flow of packets/bytes over the duration of an actual TCP connection.

While the question of "Why do networks act as conservative cascades?" remains unanswered in [9] and remains one of the most interesting and challenging outstanding problems in this area, Feldmann et al. do propose a workload model for data traffic that incorporates both the infinite variance property of measured TCP connection sizes and the conservative cascade construction for distributing the workload over the duration of the connection. Moreover, they hypothesize that the corresponding aggregate packet stream (i.e., aggregated over all connections) will exhibit asymptotic self-similarity as well as multifractal behavior, where the latter is identical in nature to the local scaling properties of the conservative cascade used for the individual connections. Such a workload model would allow for a plausible explanation of the observed multifractal nature of measured WAN traffic in terms of the multiplicative property of the conservative cascade construction underlying the individual TCP connections. More importantly, it makes some of the different aspects associated with the TCP protocol the most likely candidates for ultimately serving as a phenomenological or physical explanation for the observed cascade structure underlying TCP connections.

2.5 The inverse cascade construction

For the tools and techniques to be developed in the later sections to apply, it will be necessary to check whether or not the structural properties of measured data network traffic (at the aggregate level or at the level of individual connections) conform to an underlying conservative cascade procedure. To this end, we consider a simple heuristic, the *inverse cascade construction*. The main objectives of the inverse cascade construction are (i) to check whether or not a semi-random or conservative rule for redistributing mass (which we equate with number of packets per chosen time interval) from a parent interval to its two subintervals is consistent with the data and, if so, (ii) to infer the pertinent statistical properties of the generator W of the underlying conservative cascade. The data set we use is the WAN traffic trace DATASET 2.

We start the inverse cascade procedure by fixing a fine time scale Δ (here, $\Delta = 10$ milliseconds) and considering the time series representing the total number of packets in non-overlapping intervals of size Δ (for convenience, we take the length of the time series to be a power of 2). Next we sum over non-overlapping

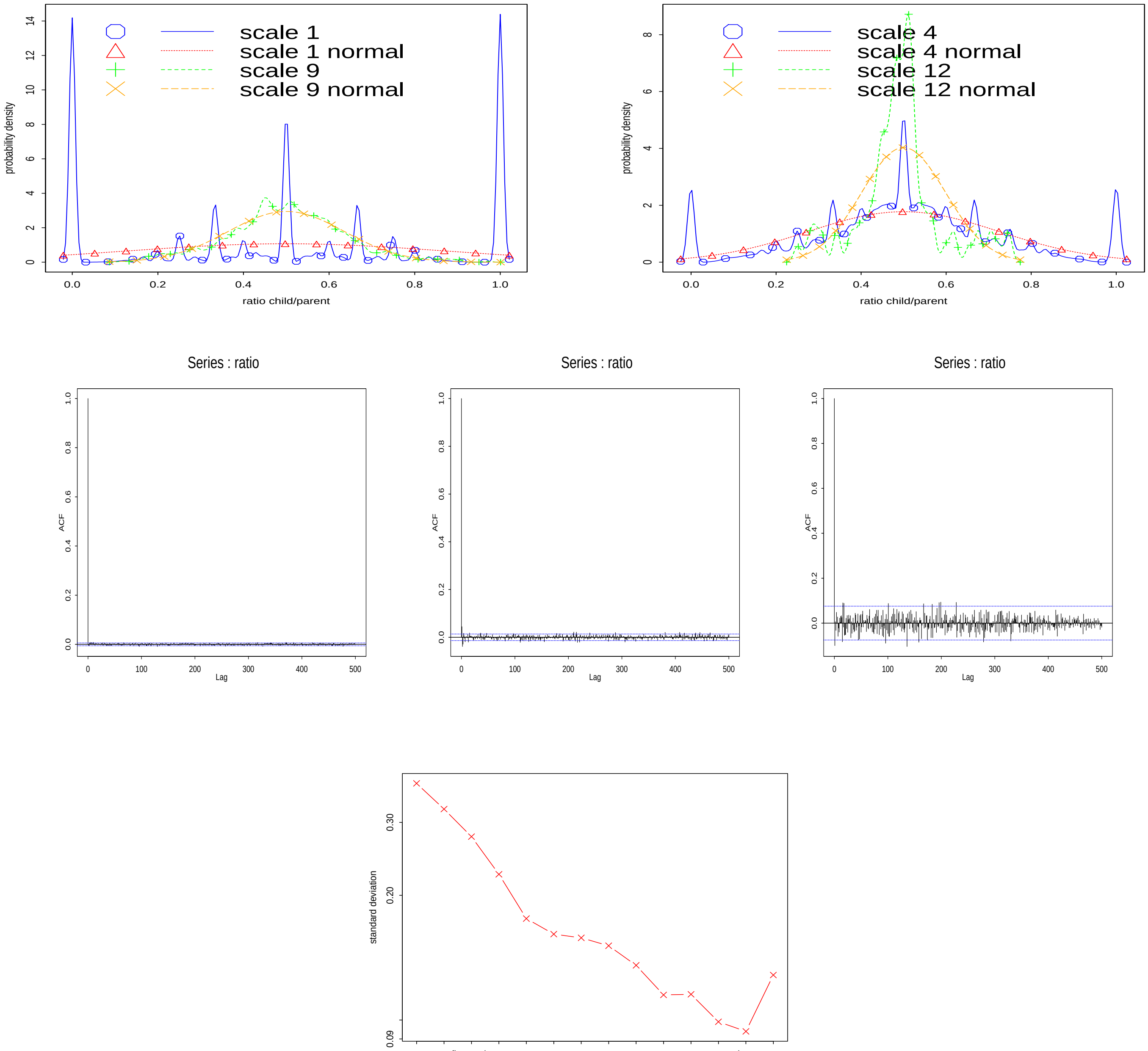

Figure 4: Inferring the generator of a conservative cascade for DATASET 2: density of the ratios for levels 1,4,9, and 12 in the inverse cascade procedure (top); autocorrelations of the ratios for scales 1, 4, and 9, where the horizontal lines denote the 95% confidence bands of statistically significant autocorrelations (middle); variability of the ratios across levels (bottom left).

blocks of size 2 of the intervals (left child and right child) and obtain the total number of packets that each parent distributed to its two children; that is, the time series representing the number of packets per time unit of size 2Δ. Iterating this process and adding now over non-overlapping blocks of size 2 of parents results in a new time series that gives the total number of packets per time unit of size 4Δ, and so on. Given a parent-related time series at scale $2^k\Delta$ and the corresponding children-related time series at scale $2^{k-1}\Delta$, we then check the properties of the empirical distribution of the ratios *number of packets in the left interval* divided by *number of packets in the parent interval.* We use this information to infer the statistical properties of the underlying generator W.

Figure 4 shows the results of applying the inverse cascade procedure to the traffic trace DATASET 2. The top two plots show the empirical probability density functions of the ratios for a number of selected stages ($k = 1, 4, 9, 12$) in the inverse cascade construction together with their fitted truncated normal distributions. For the stages $k = 1, 4, 9$ the middle plots give the empirical autocorrelation functions for the corresponding ratios. Finally, the bottom plot of Figure 4 shows the empirical standard deviation of the ratios as a function of the stage k in the inverse procedure. Figure 4 provides strong empirical evidence in support of a conservative cascade construction underlying the measured WAN trace DATASET 2. Indeed, across the different stages in the inverse construction, the empirically observed properties of the ratios suggest a generator W that follows roughly a truncated normal on $[0, 1]$ with mean $1/2$; i.e., it is symmetric around $1/2$ (top plots) and appears roughly independent when viewed across any given fixed stage (middle plots). Figure 4 (bottom left) also illustrates that the variability of the generator W varies more or less monotonically as a function of the stage in the inverse construction, with a slightly slower decrease in the variability in the late stages (coarse time scales) of the inverse construction than in the earlier stages (fine time scales). In other words, the generator is not fixed but only its variability changes at each level of the conservative cascade construction.

3 Global scaling of cascades

Wavelets provide a tool for time-frequency localization; the wavelet transform divides data into different frequency components and analyzes each component with a resolution matched to its scale. We can use the coefficients of a wavelet decomposition to directly study the scale (or frequency) dependent properties of the data. For example, if we fix a scale j and study a signal at that scale across time (usually by computing certain statistics about the wavelet coefficients at that scale), we can obtain information about the scaling behavior of the signal as a function of j (the "global" scaling behavior). In particular, we examine the scaling behavior of the process across large scales. Alternatively, if we fix a point in time t_0 and examine how the wavelet coefficients within the "cone of influence of t_0" change across scales as we examine finer and finer scales, we can determine the local irregularity (the "local" scaling behavior) of the signal about the point t_0. In this and the following sections we analyze both the global and local scaling behaviors of the class of multifractals generated by conservative cascades, using their discrete wavelet transforms. In this section we focus on the global scaling properties of cascades and their connections to self-similar processes. In particular, we show how the generator of an underlying conservative cascade affects the global scaling behavior of the limiting multifractal.

The key feature of a wavelet expansion is that we can write an approximation of a signal X at scale j (with resolution 2^j) as the sum of a coarser approximation at scale $j+1$ (with resolution 2^{j+1}) and the difference between these two approximations. We may iterate this procedure, writing the approximation at scale $j+1$ as a sum of a coarser approximation and the difference. We write the *wavelet decomposition* of a signal X as

$$X = \sum_{k \in \mathbf{Z}} \langle X, \phi_{0,k} \rangle \, \phi_{0,k} + \sum_{j \le 0} \sum_{k \in \mathbf{Z}} \langle X, \psi_{j,k} \rangle \, \psi_{j,k} = \sum_{j \in \mathbf{Z}} \sum_{k \in \mathbf{Z}} \langle X, \psi_{j,k} \rangle \, \psi_{j,k} .$$

We call the inner products $\langle X, \psi_{j,k} \rangle$ of X with the rescaled and translated copies of the wavelet ψ the *wavelet coefficients* $d_{j,k}$ of X and we refer to the set of all wavelet coefficients as the *discrete wavelet transform* (DWT) of the signal X. The coefficient $|d_{j,k}|^2$ measures the amount of energy in a signal X about the time $t_0 = 2^j k$ and about the frequency $2^{-j}\lambda_0$, where λ_0 is a reference frequency which depends on the wavelet ψ. The formal definitions of the discrete wavelet transform are given in Appendix.

3.1 Self-similar processes

Let X denote a signal that is generated by a finite variance, wide-sense stationary, long-range dependent process with Hurst parameter $H \in (1/2, 1)$; that is, the spectral density $r_X(\lambda)$ of X obeys a power-law for frequencies near the origin [1, 2],

$$r_X(\lambda) \sim c_r |\lambda|^{1-2H} \quad \text{as } \lambda \to 0, \tag{3}$$

where c_r denotes a finite positive constant, independent of λ. The Hurst parameter measures the degree of long-range dependence; short-range dependent processes have $H = 1/2$. Long-range dependence plays an important role in the study of self-similar processes. Here we call a wide-sense stationary process X *exactly self-similar* with self-similarity parameter H if for all integers $m > 0$,

$$X = m^{1-H} X^{(m)}, \tag{4}$$

where the equality holds in the sense of finite-dimensional distributions, and where the aggregated processes $X^{(m)}$ with level of aggregation m are defined by $X^{(m)}(k) = m^{-1} \sum_{i=(k-1)m+1}^{i=km} X_i,\ k \geq 1$. We define *asymptotic self-similarity* similarly but we require that the above equality holds only in the limit as $m \to \infty$. For a less stringent definition, we say that X is *exactly (asymptotically) second-order self-similar* (with self-similarity parameter H) if for all m (in the limit as $m \to \infty$), X and $m^{1-H}X^{(m)}$ are identical with respect to their second-order statistical properties. Notice that for a zero-mean process, exact (resp., asymptotic second-order) self-similarity implies that Equation (4) holds for all frequencies λ (resp., for all λ near the origin), and vice versa (e.g., see [6]). Therefore, to analyze the scaling phenomenon of (second-order) self-similar signals, we can use the information about the behavior of the spectral density r_X near the origin rather than the more delicate estimates for Equation (4).

3.2 Energy at scale j

Abry and Veitch recently proposed a wavelet-based method for analyzing long-range dependent time series arising in network traffic measurements [2] (see [1] for a general discussion of the wavelet-based spectral analysis of $1/f$ processes). In particular, they showed that the average of $|d_{j,k}|^2$ at each scale j is a useful spectral estimator. In fact, if E_j denotes the average of $|d_{j,k}|^2$ at each scale,

$$E_j = \frac{1}{N_j} \sum_k |d_{j,k}|^2$$

(N_j is the number of wavelet coefficients at each scale j), then E_j is a measure of the energy that lies within a given bandwidth 2^{-j} around frequency $2^{-j}\lambda_0$. Furthermore, the expectation of E_j is given by

$$\mathbf{E}[E_j] = \int f(\lambda) 2^j |\hat{\psi}(2^j\lambda)|^2 \, d\lambda = c_f |2^{-j}\lambda_0|^{1-2H} \int |\lambda|^{1-2H} |\hat{\psi}(\lambda)|^2 \, d\lambda, \tag{5}$$

where $\hat{\psi}(\lambda)$ is the Fourier transform of $\psi(t)$. By plotting $\log_2 E_j$ against scale j, we obtain an unbiased scaling analysis of X that is both effective and efficient. We can use this scaling analysis to identify scaling regions, breakpoints, and non-scaling behavior. For example, the scaling analysis of a signal which is asymptotically self-similar will, for large scales, show a linear relationship between $\log_2 E_j$ and the scale j. If the signal is exactly self-similar, a plot of $\log_2 E_j$ versus j will show a linear relationship for all scales. Feldmann et al. [10] used this technique to provide additional empirical evidence that local-area network (LAN) traffic seems consistent with exact self-similarity while WAN traffic (over the same range of time scales as LAN traffic) tends to exhibit asymptotic self-similarity; i.e., large-time linear scaling regions, in addition to pronounced non-linear scaling behavior for small time scales.

3.3 Scaling analysis of cascades

We now apply the wavelet-based global scaling analysis to cascades and we begin with the scaling analysis of the deterministic binomial cascade as an exercise (for simplicity, we use the Haar wavelet basis). Because the Haar wavelet $\psi_{-l,n}$ is (up to the normalization by a factor of $2^{l/2}$) the characteristic function of the left subinterval $I(j_1, \dots, j_l, 0)$ minus the characteristic function of the right subinterval $I(j_1, \dots, j_l, 1)$, the Haar wavelet coefficient $d_{-l,n}$ of a measure μ is the (normalized) difference in measure of two adjoining dyadic intervals, $2^{l/2}\big(\mu(I(j_1, \dots, j_l, 0)) - \mu(I(j_1, \dots, j_l, 1))\big)$. In particular, the wavelet coefficient $d_{-l,n}$ of the (deterministic binomial) measure μ_∞ at level l (or scale $j = -l$) is

$$d_{-l,n} = 2^{l/2} \left(\prod_{k=1}^{l} p_{j_k} \right) (2p - 1),$$

where the factors are given by $p_0 = p$ and $p_1 = 1 - p$. If we sum $|d_{-l,n}|^2$ over all wavelet coefficients at a fixed level l, the (normalized) energy concentrated at that level l is

$$E_l = \frac{1}{2^l} \sum_{\alpha \in A} 2^l \alpha (2p-1)^2 = (2p-1)^2 \big(p^2 + (1-p)^2\big)^l$$

where the sum is taken over all possible words α (of length l) in the alphabet $A = \{pp, (1-p)(1-p)\}^l$. The key observation is that the sum

$$\begin{aligned}\sum_{\alpha \in A} \alpha &= p^{2l} + lp^{2(l-1)}(1-p)^2 + \cdots + lp^2(1-p)^{2(l-1)} + (1-p)^{2l} \\ &= \sum_{r=0}^{l} \binom{l}{r} p^{2(l-r)}(1-p)^{2r} = \left(p^2 + (1-p)^2\right)^l\end{aligned}$$

is simply the binomial expansion of $\left(p^2 + (1-p)^2\right)^l$. If we take the logarithm of E_l, we find that $\log_2(E_l)$ depends linearly on the level l (or scale $j = -l$):

$$\begin{aligned}\log_2(E_l) &= \log_2\left((2p-1)^2\left(p^2 + (1-p)^2\right)^l\right) \\ &= \log_2\left((2p-1)^2\right) + \log_2\left(\left(p^2 + (1-p)^2\right)^l\right) \\ &= \left(\log_2\left(p^2 + (1-p)^2\right)\right)l + \log_2\left((2p-1)^2\right).\end{aligned}$$

If we plot $\log_2(E_l)$ as a function of l, we get a line with slope $\log_2\left(p^2 + (1-p)^2\right)$ which indicates simple scaling (or one scaling region). In other words, the deterministic binomial cascade has global linear scaling. Furthermore, the intercept of the line $\log_2(E_l)$ is the only term which depends on our choice of the Haar wavelet. We summarize the above calculations as:

Theorem 3.1 *Let μ_∞ be the limiting measure generated by the deterministic binomial cascade with generator $p \in (0, \frac{1}{2}]$. Then μ_∞ has global linear scaling; i.e, the logarithm of the energy in μ_∞ contained around level l depends linearly on l with the form:*

$$\log_2(E_l) = \left(\log_2(p^2 + (1-p)^2)\right)l + \log_2\left((2p-1)^2\right).$$

These arguments serve as a warm-up to the following theorem which characterizes the global scaling of conservative cascades. In fact, the proof of this theorem is no different in spirit from the above calculations.

Theorem 3.2 *Let μ_∞ be the conservative cascade with fixed generator W. The measure μ_∞ has global linear scaling and the logarithm of the expected value of the energy in μ_∞ contained around level l depends linearly on l with the form:*

$$\log_2 \mathbf{E}[E_l] = l\left(1 + \log_2 \mathbf{E}[W^2]\right) + \log_2 \mathbf{E}\left[(2W-1)^2\right] = \chi(2)l + \log_2 \mathbf{E}\left[(2W-1)^2\right],$$

where $\chi(\cdot)$ denotes the modified cumulant generating function defined by Equation (1).

Proof. The wavelet coefficient $d_{-l,n}$ of μ_∞ at scale $-l$ is the product

$$d_{-l,n} = 2^{l/2} W(j_1) \cdots W(j_1, \dots, j_l)\big(2W(j_1, \dots, j_l, 0) - 1\big)$$

where the indices $j_1, \dots, j_l$ form the dyadic expansion of the point $2^{-l}(n)$. The energy E_l is the average of $|d_{-l,n}|^2$ over all wavelet coefficients at scale $-l$,

$$E_l = \sum_{(j_1,\dots,j_l)\in\{0,1\}^l} W^2(j_1) \cdots W^2(j_1, \dots, j_l)\big(2W(j_1, \dots, j_l, 0) - 1\big)^2$$

and its expected value is given by

$$\mathbf{E}[E_l] = \sum_{(j_1,\dots,j_l)\in\{0,1\}^l} \mathbf{E}\Big[W^2(j_1) \cdots W^2(j_1, \dots, j_l)\big(2W(j_1, \dots, j_l, 0) - 1\big)^2\Big]$$

Recall that the random variables $W(j_1), \dots, W(j_1, \dots, j_l)$, $l \geq 1$, are identically distributed as the generator W and they have dependence structure (2). Moreover, W and $1 - W$ have the same distribution because W is symmetric about its mean. Thus we may simplify the expression for $\mathbf{E}[E_l]$:

$$\begin{aligned}
\mathbf{E}[E_l] &= \sum_{(j_1,\dots,j_l)\in\{0,1\}^l} \mathbf{E}\Big[W^2(j_1) \cdots W^2(j_1, \dots, j_l)\big(2W(j_1, \dots, j_l, 0) - 1\big)^2\Big] \\
&= \sum_{(j_1,\dots,j_l)\in\{0,1\}^l} \mathbf{E}[W^2]^l \mathbf{E}[|2W - 1|^2] \\
&= 2^l \mathbf{E}[W^2]^l \mathbf{E}[|2W - 1|^2].
\end{aligned}$$

Taking the logarithm of $\mathbf{E}[E_l]$, we obtain

$$\log_2 \mathbf{E}[E_l] = l\big(1 + \log_2 \mathbf{E}[W^2]\big) + \log_2 \mathbf{E}[|2W - 1|^2] = \chi(2)l + \log_2 \mathbf{E}[|2W - 1|^2],$$

the desired result. $\square$

Note that the slope in the formula in Theorem 3.2 depends only on $\mathbf{E}[W^2]$, the second moment of the generator and is given by the modified cumulant generating function $\chi(2)$. In fact, if we choose the generator W to be p with probability $1/2$ and $1 - p$ with probability $1/2$ so that we generate a conservative binomial cascade, then the slope of the scaling analysis of this semi-random binomial cascade is the same as the slope of the scaling analysis of the deterministic binomial cascade with the same value of p since $\log_2 \mathbf{E}[W^2] + 1 = \log_2(p^2 + (1-p)^2)$. As in the deterministic binomial cascade, the intercept of the line $\log_2 \mathbf{E}[E_l]$ is the only term which depends on our choice of the Haar wavelet. If we use other compactly supported wavelets (with 2^k tap filters) to analyze the global scaling behavior of cascades, we get similar results: the scaling analysis shows a linear relationship between the logarithm of the energy at level l contained in the cascade and the level l, the slope of this line depends only on the second moment of the generator, and the only term which depends upon our choice of wavelets is the intercept of the line $\log_2 \mathbf{E}[E_l]$. If the wavelet corresponds to a 2^k tap filter, then there are 2^{k-1} terms in the intercept which depend on the filter.

Theorem 3.2 implies that a cascade with a fixed generator W will have a linear global scaling analysis and that only the second moment of the generator W determines the slope of the line. To allow for a more flexible (e.g., non-linear) global scaling behavior within the conservative cascade paradigm, we can change the second moment of the generator W at each level in our cascade construction (or within a range of scales). The following theorem shows how a change in the second moment of the generator affects the scaling analysis. Let $W(j_1, \ldots, j_l)$, $l \geq 1$, denote the sequence of random variables associated with the conservative cascade construction; that is, a sequence of random variables with dependence structure (2) and identically distributed as the generator W up to the normalization factor $\lambda_l^2 = \mathbf{Var}\big(W(j_1, \ldots, j_l)\big)/\mathbf{Var}(W)$. In other words, $W(j_1, \ldots, j_l)$ is equal in distribution to $\lambda_l W + 1/2(1-\lambda_l)$ which means at each stage in the conservative cascade construction process we change only the variance of the generator and we keep $\mathbf{E}[W(j_1, \ldots, j_l)] = 1/2$ and the dependence structure (2). We assume that $\mathbf{Var}(W) \geq \mathbf{Var}\big(W(j_1, \ldots, j_l)\big)$ so that $\lambda_l^2 \leq 1$.

Theorem 3.3 *The limiting measure μ_∞ of the conservative cascade with variable generator $\lambda_l W + 1/2(1-\lambda_l)$, $l \geq 1$, exhibits non-linear global scaling, and the logarithm of the expected value of E_l is given by*

$$\log_2 \mathbf{E}[E_l] = l\big(1 + \log_2 \mathbf{E}[W^2]\big) + \sum_{k=1}^{l} \log_2\left(\lambda_k^2 + \frac{1-\lambda_k^2}{4\mathbf{E}[W^2]}\right) + 2\log_2 \lambda_{l+1} + \log_2 \mathbf{E}[|2W-1|^2].$$

The global scaling analysis $\log_2 \mathbf{E}[E_l]$ is finite provided λ_{l+1} is not equal to zero.

Proof. We begin with the expectation of the energy E_l contained in the measure μ_∞ at level l

$$\mathbf{E}[E_l] = \sum_{(j_1,\ldots,j_l)\in\{0,1\}^l} \mathbf{E}\Big[W^2(j_1)\cdots W^2(j_1,\ldots,j_l)\big(2W(j_1,\ldots,j_l,0)-1\big)^2\Big].$$

Observe that because we have rescaled the random variables $W(j_1,\dots,j_l) = \lambda_l W + 1/2(1-\lambda_l)$ (that is, they are equal in distribution to a renormalization of the generator W), we can simplify the expectation of E_l (we also use the arguments in the proof of the previous theorem):

$$\begin{aligned}\mathbf{E}[E_l] &= \sum_{(j_1,\dots,j_l)\in\{0,1\}^l} \left(\prod_{k=1}^{l} \mathbf{E}\big[|\lambda_k W + 1/2(1-\lambda_k)|^2\big]\right)\left(\lambda_{l+1}^2 \mathbf{E}[|2W-1|^2]\right) \\ &= 2^l \lambda_{l+1}^2 \mathbf{E}[|2W-1|^2] \prod_{k=1}^{l} \mathbf{E}\big[|\lambda_k W + 1/2(1-\lambda_k)|^2\big].\end{aligned}$$

Thus, the logarithm of $\mathbf{E}[E_l]$ is given by

$$\log_2 \mathbf{E}[E_l] = l + \sum_{k=1}^{l} \log_2 \mathbf{E}\big[|\lambda_k W + 1/2(1-\lambda_k)|^2\big] + 2\log_2 \lambda_{l+1} + \log_2 \mathbf{E}[|2W-1|^2]$$

We simplify the sum in the above expression, using the relation

$$\mathbf{E}\big[|\lambda_k W + 1/2(1-\lambda_k)|^2\big] = \lambda_k^2 \mathbf{E}[W^2] + 1/4(1-\lambda_k^2) = \mathbf{E}[W^2]\left(\lambda_k^2 + \frac{1-\lambda_k^2}{4\mathbf{E}[W^2]}\right) = \mathbf{E}[W^2](z_k),$$

to obtain the following expression for the logarithm of $\mathbf{E}[E_l]$:

$$\log_2 \mathbf{E}[E_l] = l\big(1 + \log_2 \mathbf{E}[W^2]\big) + \sum_{k=1}^{l} \log_2(z_k) + 2\log_2 \lambda_{l+1} + \log_2 \mathbf{E}[|2W-1|^2].$$

If $\lambda_{l+1} \neq 0$, then $\log_2 \mathbf{E}[E_l]$ is finite. $\square$

The value of $\log_2 \mathbf{E}[E_l]$ is finite as l tends to infinity when the sum

$$2\log_2 \lambda_{l+1} + \sum_{k=1}^{l} \log_2(z_k) \leq \sum_{k=1}^{l+1} \log_2(z_k)$$

converges as l tends to infinity. Because the terms $z_k = \lambda_k^2 + \frac{1-\lambda_k^2}{4\mathbf{E}[W^2]}$ are all greater than zero (recall $\lambda_k^2 \leq 1$), the sum $\sum_{k=1}^{\infty} \log_2(z_k)$ converges if and only if the sum $\sum_{k=1}^{\infty} z_k - 1$ converges absolutely, where

$$z_k - 1 = \lambda_k^2 + \frac{1-\lambda_k^2}{4\mathbf{E}[W^2]} - 1 = \frac{\big(4\mathbf{E}[W^2]-1\big)\big(\lambda_k^2-1\big)}{4\mathbf{E}[W^2]}.$$

Therefore, the sum $\sum_{k=1}^{\infty} \log_2(z_k)$ converges if and only if $\frac{4\mathbf{E}[W^2]-1}{4\mathbf{E}[W^2]} \sum_{k=1}^{\infty} \big|\lambda_k^2 - 1\big|$ converges. Note that λ_k^2 cannot tend to 1 too slowly. If $\lambda_k^2 = 1 - 1/k$, then the sum $\sum_{k=1}^{\infty} \big|\lambda_k^2 - 1\big|$ equals $\sum_{k=1}^{\infty} 1/k$ which diverges.

For the global scaling analysis, we associate level l in the cascade with scale $j = -l$ (or resolution 2^{-l}) and we plot $\log_2 \mathbf{E}[E_l]$ from the finest scales to the coarsest; that is, the global scaling analysis is the graph

$\{(-l, \log_2 \mathbf{E}[E_l])\}$. Using the first two terms in the Taylor expansion of $\log_2\left(\lambda_{l+1}^2 + \frac{1-\lambda_{l+1}^2}{4\mathbf{E}[W^2]}\right)$, it is easy to see that the slope of the line connecting two adjacent values of $\log_2 \mathbf{E}[E_l]$ in the global scaling analysis is approximately the difference

$$\log_2 \mathbf{E}[E_l] - \log_2 \mathbf{E}[E_{l+1}] = -(1 + \log_2 \mathbf{E}[W^2]) + \log_2 \lambda_{l+1}^2 - \log_2 \lambda_{l+2}^2 - \log_2\left(\lambda_{l+1}^2 + \frac{1-\lambda_{l+1}^2}{4\mathbf{E}[W^2]}\right)$$

$$\approx -(1 + \log_2 \mathbf{E}[W^2]) - 2\log_2 \lambda_{l+2} - \frac{1}{4\mathbf{E}[W^2]\ln 2}\left(\frac{1-\lambda_{l+1}^2}{\lambda_{l+1}^2}\right).$$

Thus, if the renormalization factors $\lambda_l^2 = \mathbf{Var}\big(W(j_1, \ldots, j_l)\big)/\mathbf{Var}(W)$ of the generators at each stage in the construction process increase (decrease) monotonically (as we go from the coarsest scale to the finest), then the slope of the global scaling analysis increases (decreases) monotonically from the finest scale to the coarsest.

We use Theorem 3.3 to interpret the global scaling behavior of the traffic trace DATASET 2. In Figure 5, we show (left) the empirical standard deviation of the ratios as a function of scale obtained from the inverse cascade construction discussed earlier (same as bottom plot in Figure 4). Observe that the standard deviation decreases almost monotonically as a function of scale from the finest scales to the coarsest, with a significant "bend" around the fifth finest scale. On the right in Figure 5 is the plot of the global scaling analysis for DATASET 2. Note the corresponding "bend" in the global scaling behavior around the fifth finest scale, where the slope increases as the scales get coarser. While Figure 4 suggests that DATASET 2 is consistent with an underlying conservative cascade construction with a varying generator W, Theorem 3.3 tells us that the only property of the cascade generator W that determines the slope of the global scaling analysis is its variability and that a decrease in variability of W from fine to coarse scales yields an increase in the slope of the global scaling analysis—just as shown in Figure 5.

3.4 Distribution of wavelet coefficients at scale j

Instead of relying on statistics such as the energy E_j in scale j that simply take the average of the $|d_{j,k}|^2$, the amount of energy in the signal X about the time $t_0 = 2^j k$ and about the frequency $2^{-j}\lambda_0$, across time for fixed j, we can use more sophisticated statistics of the wavelet coefficients that give a more informative scaling analysis of the underlying signal X. For example, given the DWT of a self-similar signal X, we fix scale j and consider the marginal distribution or, equivalently, the probability density function f_j of the stationary

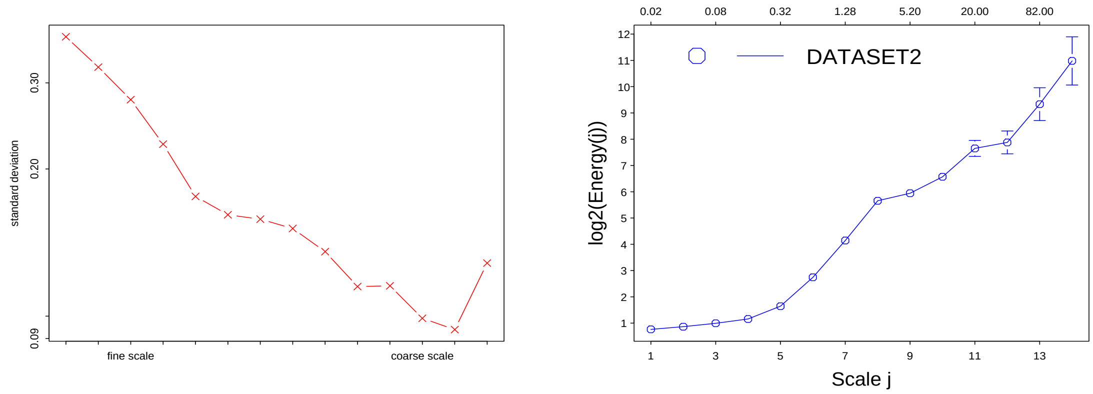

Figure 5: Variability of the generator across levels (left), global scaling analysis (right) for DATASET 2.

process $d_j = (d_{j,k} : k \in \mathbf{Z})$. Because wavelets are orthogonal to constants, d_j has mean zero; moreover, because of property (3), we can show that the variance of d_j is of the form $2^{j(2H-1)}$ (see [2]). Therefore, if we rescale the wavelet coefficients at scale j by the factor $2^{-j(H-1/2)}$ (that is, consider the normalized wavelet coefficients $\tilde{d}_{j,k} = 2^{-j(H-1/2)} d_{j,k}$) and we look at the corresponding probability density distributions $\tilde{f}_j$ (or equivalently, appropriate QQ-plots), we obtain a scaling analysis of the underlying signal X that provides information about the distributions of the wavelet coefficients across different scales. For example, if X is approximately Gaussian and the probability density functions $\tilde{f}_j$ "collapse" onto a single distribution, we can conclude that X is consistent with exact (second-order) self-similarity with scaling exponent H. On the other hand, for an asymptotically (second-order) self-similar signal X with self-similarity parameter H, we expect the large-scale (i.e., j large) density functions $\tilde{f}_j$ to collapse, but to differ significantly from the small-scale (i.e., small j) $\tilde{f}_j$'s.

In fact, if the signal X is Gaussian with zero mean, then the densities $\tilde{f}_j$ collapse if and only if X is exactly self-similar. If we restrict our attention to a fractional Brownian motion process B_H with mean zero and Hurst parameter H, then the mean of the wavelet coefficients at each scale j is also zero. In addition, the distribution of the wavelet coefficients $d_{j,k}$ at each scale j is Gaussian because B_H is Gaussian in nature. Equation (5) tells us that in this case the only factor in the standard deviation of the process $d_{j,k}$ which depends on j is the term $2^{j(H-1/2)}$. Therefore, if we re-scale each (Gaussian) distribution of wavelet

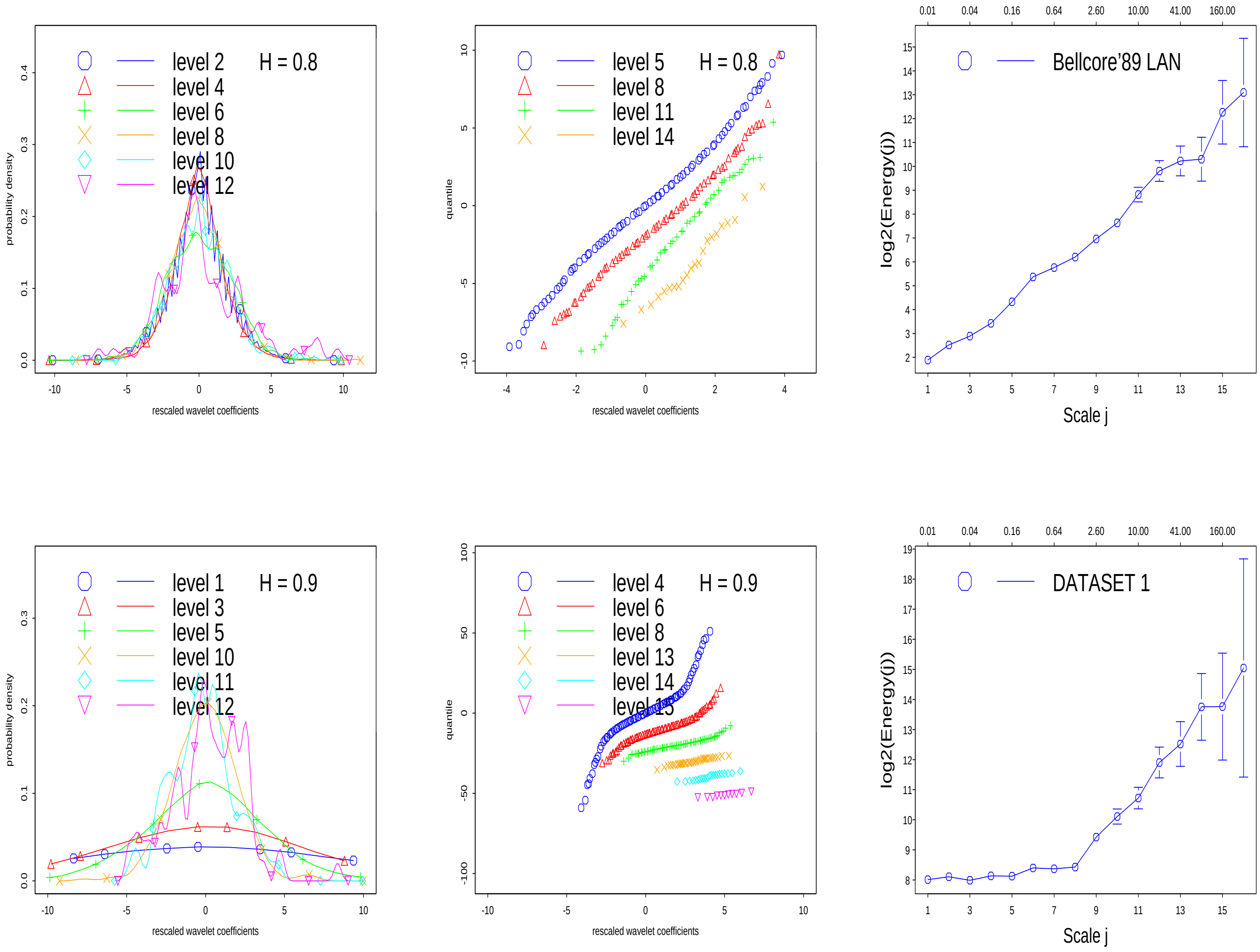

Figure 6: Probability density plots (left) and QQ-plots (middle) of the rescaled wavelet coefficients, for different scales: Bellcore'89 LAN (top) and DATASET 1B (bottom); global scaling plot (right).

coefficients, setting $\tilde{d}_{j,k} = 2^{-j(H-1/2)} d_{j,k}$, then the distributions of the coefficients $\tilde{d}_{j,k}$ will all be Gaussian with zero mean and the same variance; i.e., a plot of all the distributions will "collapse" onto one single Gaussian distribution.

Figure 6 shows the results of our global scaling analysis using the probability density functions $\tilde{f}_j$ of the rescaled wavelet coefficients for different scales and for two different traces. The top plots are for the well-studied August'89 Bellcore Ethernet LAN trace (see for example, [2, 16, 9]), the bottom plots are for DATASET 1B. The plots on the left-hand side show the rescaled densities $\tilde{f}_j$ overlaid for a number of different scales, and the plots in the middle depict the QQ-plots corresponding to the different $\tilde{f}_j$'s (for ease of comparison, the QQ-plots of the different $\tilde{f}_j$'s are offset to avoid overstriking). While the plots for the LAN trace provide convincing evidence for a collapse of the wavelet coefficient densities and are yet another indication of the exactly self-similar nature of LAN traffic, the plots corresponding to the WAN trace illustrate that even though the densities $\tilde{f}_j$ for large-time scales collapse, the large-time scale densities are significantly different from the small-time scale densities. The latter properties are clear indications that the underlying trace is asymptotically rather than exactly self-similar. Note that these plots are fully consistent with the scaling analysis based on the energy-statistics (plots on the right-hand side), but provide more detailed information about the wavelet coefficients than the global scaling analysis plots of $\log_2 E_j$ vs. j. The scaling analysis plot of the LAN trace is almost linear while the scaling analysis plot of DATASET 1 shows linear scaling at large time scales and a distinct break or bend at scale $j = 8$, roughly corresponding to one second.

4 Local scaling analysis

The global scaling analysis measures a global property of a signal within each scale; e.g., a measure of the average energy of a signal at a given scale and how this average energy changes as a function of scale (as we look at coarser scales). The global scaling analysis gives us a partial picture of the scaling in data; it does not give information about the local behavior of the data. The possibility that a signal can exhibit non-trivial scaling analysis at small time scales motivates a detailed investigation into the local irregularities or local scaling behavior of a given signal. In this section we define multifractals (and monofractals) which provide a mathematical framework for describing local singularities. We discuss the local scaling behavior of a measure and introduce a DWT-based structure function with which we compute the spectrum of local scaling exponents (the multifractal spectrum) of a measure. We also analyze the local scaling behavior of the class of multifractals that are generated by the conservative cascades. Note that even though structure functions and multifractal spectra are global statistics (i.e., they only provide information about the frequency with which certain scaling exponents occur within the signal; in particular, they say nothing about where in the signal a certain scaling exponent occurs), we use here the term *local scaling analysis* to contrast it from the *global scaling analysis* discussed earlier. As illustrated in Gilbert et al. [12], this terminology is appropriate and justified because multifractal spectra can be thought of as arising from a genuinely localized (in time) analysis of a signal.

4.1 Monofractals and multifractals

To define multifractals (and monofractals), we restrict our attention to non-negative measures μ on the unit interval and we consider dyadic partitions of $[0,1]$ without loss of generality. We classify the singularities of μ by measuring the singularity exponent $\alpha(t_0)$ at the point t_0:

$$\lim_{l\to\infty} \frac{\log\big(\mu(B_{2^{-l}}(t_0))\big)}{-l\log 2} = \alpha(t_0).$$

where $B_{2^{-l}}(t_0)$ are the dyadic intervals of size 2^{-l} that contain the point t_0. If the above limit does not exist, we leave $\alpha(t_0)$ undefined. We say that μ is *multifractal* if the scaling behavior of μ at t_0 depends on the point t_0; that is, if $\alpha(t_0)$ varies as t_0 varies. A measure μ is *monofractal* if a single global scaling exponent α describes the singularities of μ. Self-similar processes with self-similarity parameter H (renormalized so that they can be considered as measures) are a special example of monofractals for which $\alpha(t_0) = H$, for all t_0.

4.2 Multifractal spectrum of a measure

While it is possible to calculate the local scaling exponents $\alpha(t)$ of a measure μ for each point $t \in [0,1]$, it is not always the best way to examine the local properties of the measure; it's simply too detailed a perspective. Instead, we can develop a more refined approach from which we may draw statistically sound conclusions about how frequently certain scaling exponents appear. We refer to the statistical distribution of scaling exponents $\alpha(t)$ and the frequency with which $\alpha(t)$ takes on a specified value α as the multifractal spectrum of a measure μ. In general, the points in $[0,1]$ with equal singularity strength form subsets K_α of $[0,1]$

that themselves have fractal (geometric) properties (hence the notion "multifractal"). In other words, the multifractal structure of a measure μ refers to the Hausdorff dimensions (d_h) of the sets where the measure μ has scaling exponent α:

$$K_\alpha = \left\{ t \,\middle|\, \lim_{l \to \infty} \frac{\log \mu(B_{2^{-l}}(t))}{l \log 2} = \alpha \right\}$$

The function $f(\alpha) = d_h(K_\alpha)$ is called the *multifractal spectrum of* μ. If μ has a continuous positive density on $[0, 1]$, then the dimension of K_α as a function of α is a spiked function which takes on the value 1 at $\alpha = 1$. The function $f(\alpha)$ for a monofractal measure will also be a spiked function of α. For further examples and details, see for example Arneodo [3].

Unfortunately, it is often quite difficult to calculate the Hausdorff dimension of a fractal set; whereas, it is easier to calculate the moments of the measure of small intervals. Another way to obtain the multifractal spectrum of a measure μ which is more statistical rather than geometric is to consider the *structure function* $\tilde{\tau}(q)$, for $q \in \mathbf{R}$, defined by

$$\tilde{\tau}(q) = \lim_{l \to \infty} -\frac{1}{l \log 2} \log \sum_k \left(\mu([2^{-l}(k), 2^{-l}(k+1)]) \right)^q.$$

We can think of $\tilde{\tau}(q)$ as a scaling analysis of the higher order moments of μ. The multifractal formalism asserts that the inverse Legendre transform of $\tilde{\tau}(q)$ estimates the multifractal spectrum $f(\alpha)$:

$$f(\alpha) = \min_q (q\alpha - \tilde{\tau}(q)).$$

In other words, we may formally equate the statistical and the geometric methods for computing the multifractal structure of a measure μ. However, we must be careful here; this equality has been established rigorously only for random (and deterministic) cascades and it is more accurate to say that

$$f(\alpha) \le \min_q (q\alpha - \tilde{\tau}(q)).$$

(In fact, Riedi and Mandelbrot [27] show there are exceptions to the multifractal formalism.)

In Section 3 we used the wavelet coefficients of a signal (or a measure) to study its scale-dependent properties. We fixed a scale l and computed certain statistics about the wavelet coefficients at that scale. Now we construct a structure function to analyze the multifractal structure (or the local rather than the global scaling properties) of a measure which is based upon the (Haar) DWT of the measure. Let the *partition function* $Z(q,l)$ for $q>0$ be the sum of the absolute values of the (Haar) wavelet coefficients of μ at scale l (see Appendix) raised to the qth power; that is,

$$Z(q,l)=\sum_{i=0}^{2^l-1}|d_{-l,i}|^q. \tag{6}$$

We define the *DWT-based structure function* $\tau(q)$ as follows

$$\tau(q)=\lim_{l\to\infty}\frac{\log Z(q,l)}{-l\log 2}+q/2.$$

4.3 Multifractal spectrum of conservative cascades

Using the DWT-based structure function $\tau(q)$, we compute the multifractal spectrum of the limiting measure generated by a conservative cascade. We relate the multifractal spectrum to the generator of the conservative cascade and to the modified cumulant generating function $\chi(h)$. We assume that μ_∞ is the limiting measure generated by a conservative cascade with a variable generator $\lambda_l W+1/2(1-\lambda_l)$, $l\geq 1$. We begin by observing that because μ_∞ is a conservative cascade, the power law decay of the Haar wavelet coefficients will match the power law decay of μ_∞ locally (i.e., the coefficients will have the same local scaling exponent as μ_∞ as the resolution size tends to zero). This observation motivates the construction of our partition function using the Haar wavelet coefficients of μ_∞.

We modify each wavelet coefficient $d_{-l,k}$ of the conservative cascade by replacing the factor $2W(j_1,\ldots,j_l)-1$ with $\lambda_l\left(2W(-1)-1\right)$ where $W(-1)$ is equal in distribution to W. One can show that the structure function constructed with the modified wavelet coefficients is equal almost surely to the original DWT-based structure function. We shall work with the modified wavelet coefficients in what follows.

Next, we observe that the distributions of the modified wavelet coefficients of the conservative cascade are "self similar." We make this observation precise in the following lemma:

Lemma 4.1 *Let μ_∞ denote the limiting measure generated by a conservative cascade on $[0,1]$ with variable generator $W_l=\lambda_l W+1/2(1-\lambda_l)$. The modified wavelet coefficients of μ_∞ at level l are equal in distribution to the renormalized wavelet coefficients at level $l-1$, where the renormalization depends only on the generator*

W_l; *that is,*

$$d_{-l,2k} = d_{-l,2k+1} = \frac{\sqrt{2}\lambda_{l+1}}{\lambda_l}\big(\lambda_l W + 1/2(1-\lambda_l)\big)d_{-l+1,k} = \frac{\sqrt{2}\lambda_{l+1}}{\lambda_l} W_l d_{-l+1,k}$$

for $k = 0, \ldots, 2^{l-1} - 1$ *and where the equality holds in distribution.*

Proof. The modified wavelet coefficients $d_{-l,2k}$ and $d_{-l,2k+1}$ (for $k = 0, \ldots, 2^{l-1} - 1$) are given by

$$d_{-l,2k} = 2^{l/2}\lambda_{l+1} W(j_1)\cdots W(j_1, \ldots, j_{l-1}, 0)\big(2W(-1) - 1\big)$$
$$d_{-l,2k+1} = 2^{l/2}\lambda_{l+1} W(j_1)\cdots W(j_1, \ldots, j_{l-1}, 1)\big(2W(1) - 1\big)$$

and the wavelet coefficients for $d_{-l+1,k}$ are given by the product

$$d_{-l+1,k} = 2^{(l-1)/2}\lambda_l W(j_1)\cdots W(j_1, \ldots, j_{l-1})\big(2W(-1) - 1\big).$$

Finally, we can conclude that because the variable generator is symmetric about its mean and because of the dependence structure (2), the modified wavelet coefficients at level l are equal in distribution to the renormalized coefficients at level $l-1$:

$$d_{-l,2k} = d_{-l,2k+1} = \frac{\sqrt{2}\lambda_{l+1}}{\lambda_l}\big(\lambda_l W + 1/2(1-\lambda_l)\big)d_{-l+1,k} = \frac{\sqrt{2}\lambda_{l+1}}{\lambda_l} W_l d_{-l+1,k}.$$

□

Because μ_∞ is the limiting multifractal generated by a conservative cascade and because the modified wavelet coefficients of μ_∞ are "self similar," the partition function $Z(q,l)$ defined in (6) satisfies a renormalization property:

Lemma 4.2 *The (modified) partition function* $Z(q,l)$ *of the multifractal* μ_∞ *generated by a conservative cascade with (variable) generator* $W_l = \lambda_l W + 1/2(1-\lambda_l)$ *is equal in distribution to*

$$Z(q,l) = 2^{1+q/2}\left(\frac{\lambda_{l+1}}{\lambda_l}\right)^q W_l^q Z(q, l-1)$$

where $Z(q,l-1)$ *is independent of* W_l.

Proof. We split $Z(q,l)$ into two sums, one sum over the even-indexed wavelet coefficients and the second sum over the odd-indexed coefficients:

$$Z(q,l) = \sum_{i=0}^{2^l-1} |d_{-l,i}|^q = \sum_{i'=0}^{2^{l-1}-1} |d_{-l,2i'}|^q + \sum_{i'=0}^{2^{l-1}-1} |d_{-l,2i'+1}|^q.$$

Now we apply Lemma 4.1 and we recall that the modified coefficients at level l are equal in distribution to the renormalized coefficients at level $l-1$. Therefore, we can renormalize the partition function $Z(q,l)$ by the factor $2^{q/2+1}\left(\frac{\lambda_{l+1}}{\lambda_l}\right)^q W_l^q$:

$$
\begin{aligned}
Z(q,l) &= \sum_{i'=0}^{2^{l-1}-1} 2^{q/2}\left(\frac{\lambda_{l+1}}{\lambda_l}\right)^q W_l^q |d_{-l+1,i'}|^q + \sum_{i'=0}^{2^{l-1}-1} 2^{q/2}\left(\frac{\lambda_{l+1}}{\lambda_l}\right)^q W_l^q |d_{-l+1,i'}|^q \\
&= 2^{q/2+1}\left(\frac{\lambda_{l+1}}{\lambda_l}\right)^q W_l^q \sum_{i=0}^{2^{l-1}-1} |d_{-l+1,i}|^q \\
&= 2^{q/2+1}\left(\frac{\lambda_{l+1}}{\lambda_l}\right)^q W_l^q Z(q,l-1).
\end{aligned}
$$

□

We turn now to the proof of the theorem about the multifractal spectrum of μ_∞.

Theorem 4.1 *Let μ_∞ be the limiting multifractal generated by a conservative cascade with variable generator $\lambda_l W + 1/2(1-\lambda_l) = W_l$ and assume that the limit*

$$\lim_{l\to\infty} \frac{1}{l}\left(q\log_2 \lambda_{l+1} + \sum_{k=1}^{l} \log_2 \mathbf{E}[W_k^q]\right)$$

exists and is finite. The structure function $\tau(q)$ of μ_∞ defined as the scaling exponent of the partition function $Z(q,l)$ is given by

$$\lim_{l\to\infty} \frac{\log Z(q,l)}{-l\log 2} + q/2 = -1 - \lim_{l\to\infty} \frac{1}{l}\left(q\log_2 \lambda_{l+1} + \sum_{k=1}^{l} \log_2 \mathbf{E}[W_k^q]\right) = \tau(q) \quad \textit{with probability one.}$$

Proof. As a warm-up exercise, we compute the expected value of $Z(q,l)$ (the indices $j_1,\ldots,j_l$ on the random variables W are the coefficients in the dyadic expansion of the sum index i):

$$
\begin{aligned}
\mathbf{E}[Z(q,l)] &= \mathbf{E}\left[\sum_{i=0}^{2^l-1} |d_{l,i}|^q\right] \\
&= 2^{lq/2} \sum_{(j_1,\cdots,j_l)\in\{0,1\}^l} \mathbf{E}\left[W(j_1)^q \cdots W(j_1,\ldots,j_l)^q |2W(j_1,\ldots,j_{l+1})-1|^q\right] \\
&= 2^{lq/2} \sum_{(j_1,\cdots,j_l)\in\{0,1\}^l} \left(\lambda_{l+1}^q \mathbf{E}[|2W-1|^q] \prod_{k=1}^{l} \mathbf{E}[W_k^q]\right) \\
&= 2^{l(1+q/2)} \lambda_{l+1}^q \mathbf{E}[|2W-1|^q] \prod_{k=1}^{l} \mathbf{E}[W_k^q].
\end{aligned}
$$

Let

$$Y_l = \frac{Z(q,l)}{2^{(1+q/2)l}\lambda_{l+1}^q \prod_{k=1}^l \mathbf{E}[W_k^q]}, \quad \text{for } l = 1, 2, \ldots.$$

The warm-up exercise shows us that as a function of the scale l, the random variables Y_l are uniformly bounded in $L^1([0,1])$. We claim that the collection of random variables $\{Y_l : l = 1, 2, \ldots\}$ is a martingale with respect to $\mathcal{F}_l$, the sequence of σ-fields generated by the suite of random variables $W(l_1), \ldots, W(j_1, \ldots, j_l)$. To validate this claim, we compute the conditional expectation of Y_l given $\mathcal{F}_{l-1}$ and we use Lemma 4.2 to simplify the expectations:

$$\begin{aligned}
\mathbf{E}\big[Y_l|\mathcal{F}_{l-1}\big] &= \mathbf{E}\left[\frac{Z(q,l)}{2^{(1+q/2)l}\lambda_{l+1}^q \prod_{k=1}^l \mathbf{E}[W_k^q]}\middle|\mathcal{F}_{l-1}\right] \\
&= \mathbf{E}\left[\frac{2^{1+q/2}\frac{\lambda_{l+1}^q}{\lambda_l^q} W_l^q Z(q,l-1)}{2^{(1+q/2)l}\lambda_{l+1}^q \prod_{k=1}^l \mathbf{E}[W_k^q]}\middle|\mathcal{F}_{l-1}\right] \\
&= \frac{2^{1+q/2}\frac{\lambda_{l+1}^q}{\lambda_l^q}\mathbf{E}[W_l^q] Z(q,l-1)}{2^{(1+q/2)l}\lambda_{l+1}^q \prod_{k=1}^l \mathbf{E}[W_k^q]} \\
&= \frac{Z(q,l-1)}{2^{(1+q/2)(l-1)}\lambda_l^q \prod_{k=1}^{l-1} \mathbf{E}[W_k^q]} = Y_{l-1}.
\end{aligned}$$

Because $\{Y_l : l = 1, 2, \ldots\}$ is an $L^1([0,1])$-bounded martingale, it converges almost surely to a random variable Y with $\mathbf{E}[|Y|] \leq \sup_l \mathbf{E}[|Y_l|] \leq 1/2\mathbf{E}[|2W-1|^q]$. In addition, the (super)-martingale $\{\log Y_l : l = 1, 2, \ldots\}$ converges almost surely to a random variable $\tilde{Y} \in L^1$.

Finally, we compute the scaling behavior of the structure function as l tends to infinity:

$$\begin{aligned}
\lim_{l\to\infty} \frac{\log Z(q,l)}{-l\log 2} + q/2 &= \lim_{l\to\infty} -\frac{1}{l}\left(\log_2\left(\frac{Z(q,l)}{2^{(1+q/2)l}\lambda_{l+1}^q \prod_{k=1}^l \mathbf{E}[W_k^q]}\right) + \log_2\left(2^{(1+q/2)l}\lambda_{l+1}^q \prod_{k=1}^l \mathbf{E}[W_k^q]\right)\right) + q/2 \\
&= \lim_{l\to\infty}\left(-\frac{1}{l}\log_2 Y_l - 1 - \frac{q}{l}\log_2 \lambda_{l+1} - \frac{1}{l}\sum_{k=1}^l \log_2 \mathbf{E}[W_k^q]\right) \\
&= -1 - \lim_{l\to\infty} \frac{1}{l}\left(q\log_2 \lambda_{l+1} + \sum_{k=1}^l \log_2 \mathbf{E}[W_k^q]\right) = \tau(q)
\end{aligned}$$

The first term in the limit is negligible, $\lim_{l\to\infty} \frac{-1}{l}\log_2 Y_l = 0$, because $\log Y_l$ converges almost surely to the random variable $\tilde{Y}$ which is in $L^1([0,1])$. If, in particular, the normalization factors λ_l tend to a constant (non-zero) value as l tends to infinity then

$$\tau(q) = -1 - \lim_{l\to\infty} \frac{1}{l} \sum_{k=1}^{l} \log_2 \mathbf{E}[W_k^q] \quad \text{almost surely.}$$

$\square$

Observe that if the measure μ_∞ is generated by a fixed generator W (i.e., $\lambda_l = 1$ for all l), the expression for $\tau(q)$ is greatly simplified. Thus, we have the following corollary:

Corollary 4.1 *Let μ_∞ be the limiting multifractal generated by a conservative cascade with fixed generator W. The structure function $\tau(q)$ of μ_∞ defined as the scaling exponent of the partition function $Z(q,l)$ is given by*

$$\lim_{l\to\infty} \frac{\log Z(q,l)}{-l \log_2} + q/2 = -1 - \log_2 \mathbf{E}[W^q] = -\chi(q) = \tau(q) \quad \textit{with probability one.}$$

Assuming the above result for all values of q, we have, in addition, a Large Deviation Principle (LDP) for the wavelet-based "rate function" f_w for the sequence of random variables $S_l = \log |2^{-l/2} d_{-l,k_l}|$ (where the location t in μ is encoded by k_l and is random). Set $\alpha_l(t) = \frac{S_l}{l \log 2}$ and

$$f_w(\alpha, l, \epsilon) = 2^l \mathrm{P}\big(\alpha_l(t) \in (\alpha - \epsilon, \alpha + \epsilon)\big).$$

The LDP says that if $\tau(q) = q/2 + \lim_{l\to\infty} \frac{\log Z(q,l)}{-l \log_2}$ exists and is differentiable for all $q \in \mathbf{R}$, then the double limit

$$f_w(\alpha) = \lim_{\epsilon\to 0} \lim_{l\to\infty} \frac{1}{l} \log_2 f_w(\alpha, l, \epsilon)$$

exists and is equal to the inverse Legendre transform of $\tau(q)$:

$$f_w(\alpha) = \min_\alpha \{(q\alpha - \tau(q)\}.$$

(see [24, 26] for a similar application.)

We apply the multifractal formalism and obtain an approximation of the (Hausdorff) multifractal spectrum $f(\alpha)$ through the inverse Legendre transform of $\tau(q)$:

$$f(\alpha) \le f_w(\alpha) = \min_\alpha \{(q\alpha - \tau(q)\}.$$

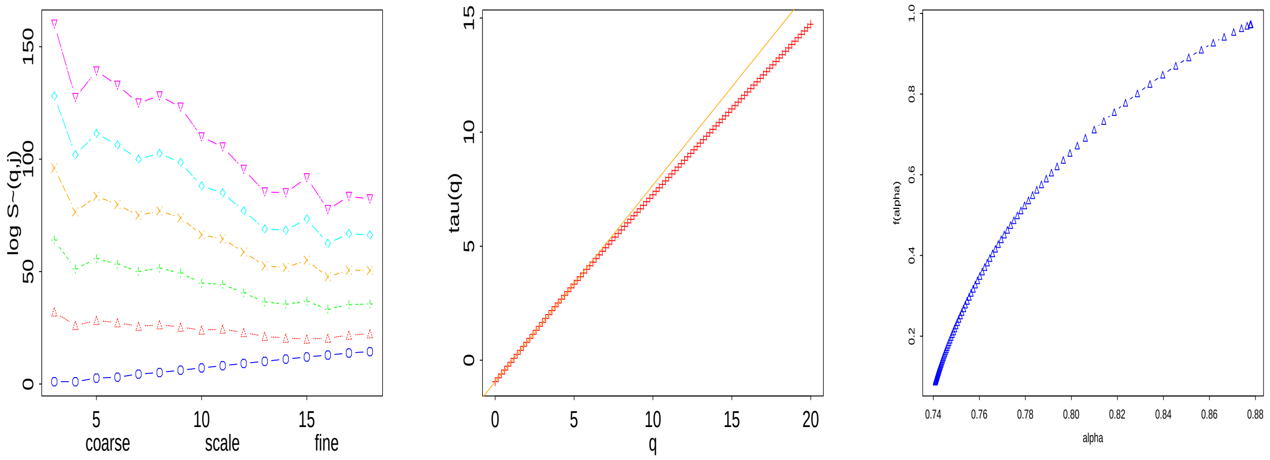

Figure 7: Local scaling analysis for DATASET 2: Partition function $Z(q, l)$ for a number of different q-values (left); structure function $\tau(q), q > 0$ (middle); and (one half of) multifractal spectrum $f(\alpha)$ (right).

To illustrate the use of these DWT-based wavelet techniques for analyzing the multifractal nature of measured WAN traces, we consider again the trace DATASET 2. We have seen earlier that this trace conforms to an underlying conservative cascade construction and hence, the DWT-based partition function $Z(q, l)$ and structure function $\tau(q)$, $q > 0$, can be used to infer the multifractal scaling behavior of this data set, in particular in estimating (one half of) the corresponding multifractal spectrum; note that since the theoretical properties of the DWT-based analysis have only been shown for non-negative q-values, we do not display the part of the estimated multifractal spectrum that corresponds to negative q-values. The left plot in Figure 7 depicts the (logarithm of the) partition function $Z(q, l)$, for $q = 0, 4, 8, 12, 16$, and 20 (from bottom to top). For a given $q > 0$, to obtain an estimate of the corresponding structure function $\tau(q)$; that is, of the scaling exponent of $Z(q, l)$ for fine time scales l, we fit a least squares line through the points $\{(l, \log Z(q, l)) : l_c < l < l_f\}$, where l_c and l_f are cutoffs at the coarse and fine time scales, respectively. In Figure 7 (middle plot), we chose $l_c = 5$ and $l_f = 16$, but the resulting $\tau(q)$-function remains relatively robust under other reasonable choices for the cutoffs. For each q-value, the structure function estimate $\tau(q)$ is obtained as the slope of the corresponding least squares line. Note that the observed non-linear (i.e., concave) shape of the structure function provides empirical evidence that the data set at hand exhibits multifractal scaling properties. That same information is also contained in the right-hand plot in Figure 7 which depicts one half (corresponding to the non-negative q-values) of the estimated multifractal spectrum.

5 Discussion and some open problems

5.1 Related work

The combination of cascades, wavelet analysis and synthesis, and the applications thereof (especially to network traffic) is an active research area with recent contributions by Arneodo et al. [4] and Riedi et al. [26]. Both Arneodo et al. and Riedi et al. propose using wavelets directly to synthesize random multifractal signals and both use wavelet-based tools for analyzing the singularity structure of these signals. Both constructions are built recursively in the (orthogonal) wavelet transform domain, starting with a coarse scale and "cascading" to finer scales. Riedi et al. develop a more specific construction tailored to modeling network traffic which is equivalent to the conservative cascade construction; they use the Haar wavelet to synthesize multifractal signals and impose a simple constraint to insure that the synthesized signal is positive. They demonstrate how to match the multifractal properties of the model to network traffic traces and what the queueing behavior of the synthesized traffic is.

5.2 On the inverse cascade heuristics

One of the original contributions of this paper is the introduction of the inverse cascade construction as a simple heuristic for checking whether or not a given data set conforms to an underlying conservative cascade construction. In this context, the obvious question arises how to use this procedure to identify data sets that are not generated via an underlying conservative cascade construction and how to distinguish them from those that are indeed consistent with such an inherent structure. To this end, we experimented in [9] with three different data sets: a WAN trace that had been shown, using different methods, to exhibit multifractal scaling; a trace of a Poisson process; and a fractional Gaussian noise trace with positive mean. Note that neither the Poisson nor the self-similar trace are multifractal. To each data set, we applied the inverse cascade construction to infer the variability of the generator, generated a synthetic trace using a conservative cascade with a generator that matches the variability of the inferred one, and then compared the statistical properties of the original and synthesized trace. Only in the case of the measured WAN traffic did the traces compare favorably, not only with respect to their first- and second-order statistics, but also with respect, for example, their structure function $\tau(q)$ that captures the multifractal scaling properties. In the Poisson and self-similar case, the original and synthesized traces showed obvious differences across the board, even with respect to the traditionally and often exclusively used first- and second-order statistics. In this sense, the inverse cascade heuristics presented here appears to be an effective technique for investigating data sets for underlying conservative cascade structures.

Recent work by Arneodo et al. [4] suggests a more refined method for uncovering the possible underlying cascade structure, the self-similarity kernel. This method captures how the details of a signal at scale j_1 are similar to the details at scale j_2 (up to a normalization factor) and it does so by relating the distributions of the wavelet coefficients at scales j_1 and j_2 (rather than the averages of the signal at two adjacent scales $j_2 = j_1 + 1$, as the inverse cascade construction does). To build the self-similarity kernel, let f_j be the probability density function of the stationary process $d_j = (d_{j,k} : k \in \mathbf{Z})$ and let $\tilde{f}_j$ be the probability density function of $\log |d_j|$. For a conservative cascade, the wavelet coefficients at scale j_2 are related to those at scale $j_1 < j_2$ with the convolution $\tilde{f}_{j_2} = \tilde{f}_{j_1} * G_{j_1,j_2}$ where $G_{j_1,j_2} = G * \ldots * G$ and G is the probability density function of $\log |W|$. In Fourier space the kernel relation is

$$\widehat{\tilde{f}_{j_2}} = \widehat{\tilde{f}_{j_1}} \widehat{G}^{s(j_1,j_2)}$$

and in physical space, we rewrite the self-similarity property as

$$f_{j_2}(e^x) = \int G_{j_1,j_2}(u) e^{-u} f_{j_1}(e^{x-u})\, du.$$

(Note that $\tilde{f}_j(x) = e^x f_j(e^x)$.) For conservative cascades, the kernel G_{j_1,j_2} depends only on the distance in scales $j_2 - j_1$ (i.e., $s(j_1, j_2) = j_2 - j_1$) and one can show that $\widehat{G}(iq) = \mathbf{E}[W^q]$ so that the self-similarity kernel does indeed capture the statistical properties of the generator W, assuming an underlying cascade construction. Numerically, we can compute the self-similarity kernel by deconvolving $\tilde{f}_{j_2}$ and $\tilde{f}_{j_1}$ (or, equivalently, dividing their Fourier transforms). Arneodo et al. suggest that we may also use the continuous wavelet transform of a signal for a richer and more stable numerical method for computing the self-similarity kernel.

5.3 On the use of DWT-based structure functions

Instead of the partition function $Z(q,l)$, $q > 0$ as defined in (6), we could also use the *modified partition function* $\tilde{Z}(q,l)$ which is similar in spirit to Arneodo's [3] partition function(s),

$$\tilde{Z}(q,l) = \sum_{\max} |d_{-l,i}|^q.$$

Instead of summing over all the wavelet coefficients at scale l, we sum over only the local maxima of the coefficients. In practice, we find the maximum of the coefficients within a sliding window. We can exhibit all of the above results for $\tilde{Z}(q,l)$ with little change in the proofs (we only have to check that $\tilde{Z}(q,l)$ is "self-similar" but since the wavelet coefficients $d_{-l,i}$ at scale l are "self-similar", this is straightforward). However, we cannot demonstrate that either partition function gives rise to a stable numerical procedure for $q < 0$. In fact, our experience suggests that the modified partition function $\tilde{Z}(q,l)$ is not obviously unstable for negative q (and might produce spurious, misleading results for $q < 0$) while the original partition function $Z(q,l)$ is obviously unstable.

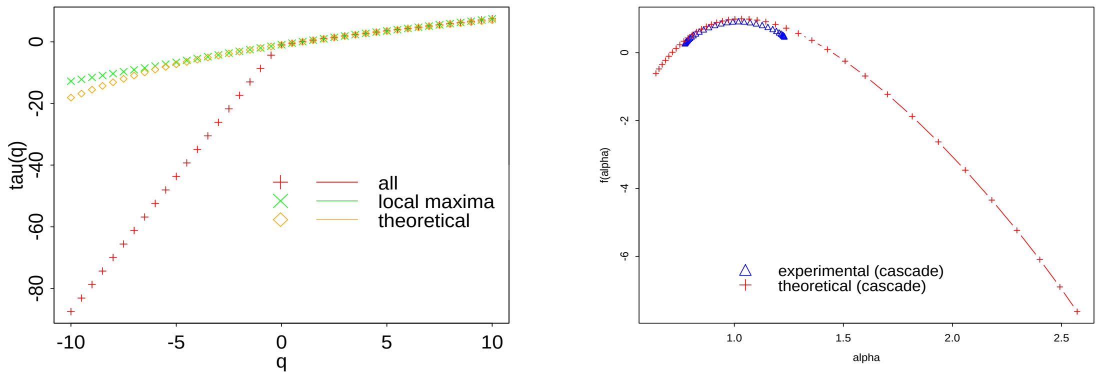

Figure 8: Conservative cascade with a truncated normal generator W (mean $1/2$ and variance 0.01): structure functions $\tau(q)$ and $\tilde{\tau}(q)$ (right) for the two partition functions $Z(q,l)$ (sum over all wavelet coefficients) and $\tilde{Z}(q,l)$ (sum over the local maxima of the wavelet coefficients); multifractal spectrum (left).

To illustrate the differences between the two definitions of the partition function ($Z(q,l)$ and $\tilde{Z}(q,l)$) and to check the accuracy of the definitions, we generate a conservative cascade with a fixed generator W (a truncated normal distribution on $(0,1)$ with mean $1/2$ and variance 0.01) and plot the theoretical structure function $\tau(q) = -1 - \log_2 \mathbf{E}[W^q]$ and the two structure functions corresponding to the two definitions of the partition function. These structure functions are shown in Figure 8 (left). All three structure functions agree for values of $q \geq 0$ but the partition function $Z(q,l) = \sum_i |d_{-l,i}|^q$ produces a structure functions that dramatically disagrees with both the modified and the theoretical structure functions for $q < 0$. The original structure function veers sharply negative for $q < 0$ while the modified structure function tracks the theoretical function more closely (although it does deviate slightly). The inverse Legendre transforms of both the modified and the theoretical structure functions are also shown in Figure 8 (right). In this plot we can see the effects of the slight variation between the modified and the theoretical structure functions. On the left-side of the spectrum, the two functions give a roughly consistent estimate for the multifractal spectrum; while on the right-side of the spectrum, the two estimates diverge (a result of the divergence in the structure functions for $q < 0$). The modified partition function $\tilde{Z}(q,l) = \sum_{\max} |d_{-l,i}|^q$ appears to yield a structure function and an estimate for the multifractal spectrum which are consistent with theoretical results for positive values of q and for the left-side of the multifractal spectrum. This example shows that while the modified partition function does not give obvious inconsistent results for negative values of q, we must be careful to not read too much into the estimate of the right-side of the multifractal spectrum.

5.4 From DWT to CWT

Intuitively, the time-localization capability of wavelets acts as a "mathematical microscope" (see Arneodo [3]) that allows us to zoom in on the local structure of a singular signal in greater and greater detail. Mathematically, this intuition is made precise by relying on the *continuous wavelet transform* (CWT) of a signal. Compared to the DWT considered so far, the CWT provides a highly redundant transformation of a given signal by unfolding the information contained in a 1-dimensional signal into a 2-dimensional state space, the time-scale plane. The appealing features of the CWT technique from a network traffic analysis perspective are (i) its ability to provide more detailed information (than the non-redundant DWT) about the local irregularities present in a given signal, and (ii) its potential for uncovering hierarchical structures (or their main characteristics) hidden underneath the measurements. This latter property makes the CWT an alternative candidate to the above-mentioned inverse cascade construction for identifying or recovering underlying cascade constructions in measured data network traffic. As is the case with the time-scale analysis technique, much theoretical as well as experimental work is required to be able to fully exploit the potential of the CWT as a visualization tool in effectively and intelligently analyzing and interpreting the wide range of network-related measurements.

5.5 Binomial vs. multinomial cascades

As far as the cascades considered in this paper are concerned, we emphasize the inherent assumption that the underlying (conservative) cascade constructions as well as the associated inverse cascade procedures are all binomial rather than trinomial, for example. While this choice has clearly been dictated by our quest for simplicity when it comes to trying to understand network traffic dynamics, we can certainly define a conservative trinomial cascade (again, allowing a variable generator) and a corresponding inverse trinomial cascade procedure (where we examine the distribution of the ratios of left, middle, and right children to the parent). Preliminary results suggest that the inverse trinomial cascade procedure reveals the data to also be consistent with a trinomial conservative cascade construction but the implications of these results have yet to be determined. We do note that it is also possible to generate a binomial conservative cascade and with the inverse trinomial cascade procedure show it to be consistent with a conservative trinomial cascade.

In Figure 9 we verify the accuracy of our modified structure function (and resulting multifractal spectrum) with a simple deterministic binomial cascade ($c = 2$ and $p_0 = 0.4, p_1 = 0.6$) and we illustrate several of the pitfalls of these methods with a deterministic trinomial cascade ($c = 3$ and $p_0 = 0.2, p_1 = 0.4, p_2 = 0.4$). The plots on the left (top and bottom) in Figure 9 show the (modified) partition functions $\tilde{Z}(q, l)$ (of the binomial and trinomial cascades) as a function of scale l (each line corresponds to a different value of q). Because the value of the structure function $\tau(q)$ is computed for each q as the slope of the line $\tilde{Z}(q, l)$, it is important to identify the "scaling regions" (i.e., the range of scales over which $\tilde{Z}(q, l)$ is linear) over which to calculate the slope. For the binomial cascade (top), there is little doubt that the scaling region encompasses all the scales shown; while for the trinomial cascade (bottom), the scaling region is not so clear. We take as the scaling region the entire range of scales for the binomial cascade and we calculate the structure function $\tau(q)$. The middle top plot shows that the calculated structure function agrees almost perfectly (even for $q < 0$) with the theoretical function $\tau(q) = -1 - \log_2(\frac{p_0^q + p_1^q}{2})$. We take two regions for comparison as the scaling region for the trinomial cascade: the entire range of scales and a "cut" range from scale 8 to 14 (the finest). In the middle bottom plot we show the two structure functions corresponding to the two choices

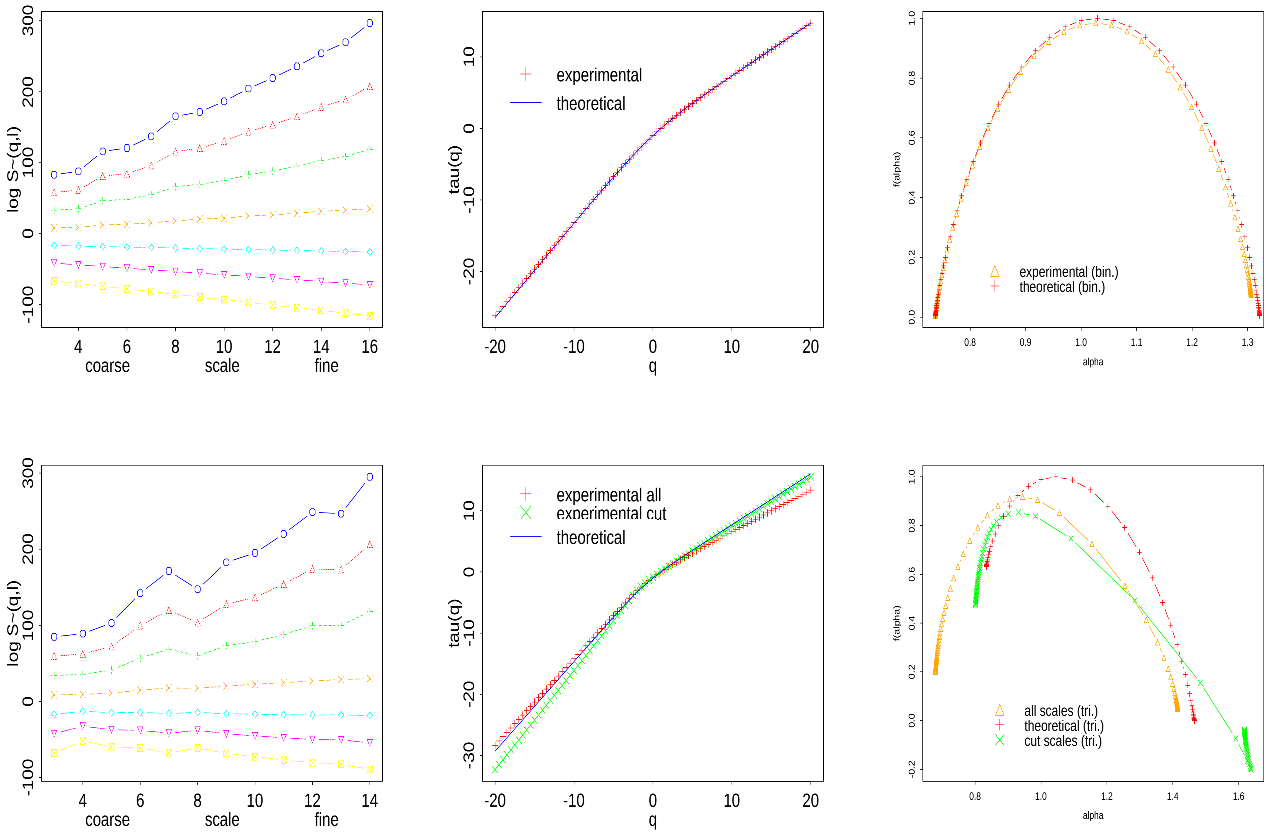

Figure 9: Partition functions $\tilde{Z}(q, l)$ (left, top and bottom); both theoretical and experimental structure functions (middle, top and bottom); experimental and theoretical multifractal spectra (right, top and bottom) for a deterministic binomial cascade (top) and a deterministic trinomial cascade (bottom).

in scaling regions and compare them with the theoretical function $\tau(q) = -1 - \log_3(\frac{p_0^q+p_1^q+p_2^q}{3})$. The "cut" structure function agrees with the theoretical structure function more closely than the "entire" function. On the other hand, neither choice in scaling region yields a highly accurate structure function. We can see in the right bottom plot how these inaccuracies manifest themselves in the multifractal spectrum estimate. In the case of deterministic cascades we know that the multifractal spectrum $f(\alpha)$ equals the inverse Legendre transform of $\tau(q)$ so the right plots (top and bottom) show that for a cascade that is inherently trinomial in nature, our "binary" methods are misleading while for an inherently binomial cascade, our methods agree with theoretical predictions. A more thorough discussion and analysis of these pitfalls is beyond the scope of this paper and is the subject of future work.

5.6 Multifractals, cascades, and TCP/IP

Motivated by the observed multifractal scaling phenomena in measured data network traffic, and realizing that it is difficult to think of any other area in the sciences where the available data provide such detailed information about so many different facets of behavior, there exists great potential for coming up with intuitively appealing, conceptually simple and mathematically rigorous statements as to the causes and effects of multifractals in data networking. Put differently, for multifractals to have a genuine impact on networking, their application has to move beyond the traditional descriptive stage and has to be able to answer question as to why network traffic is multifractal (i.e., physical explanation in the network context) and how it may or may not impact network performance (i.e., engineering).

In this paper and in [9], we move beyond the mere empirical evidence that measured WAN traffic is consistent with multifractal scaling behavior and answer the question "Why is WAN traffic multifractal?" by arguing " ... because networks appear to act as conservative cascades!" In particular, using measured WAN traffic at the level of individual TCP or port-to-port connections, we suggest that the conservative cascades underlying individual TCP connections give rise to a multiplicative structure that is recovered at the aggregate level and causes aggregate WAN traffic to exhibit multifractal scaling. While this leaves open the "big" question "Why are packets within individual TCP connections distributed in accordance with a conservative cascade construction?" it clearly identifies the TCP layer as the most promising place in the networking hierarchy for searching for the main cause for why modern data networks seem to act like conservative cascades. Clearly, progress on these problems will require a close collaboration with networking experts.

As far as their impact on network performance-related issues is concerned, multifractals and cascades suggest novel ways for dealing with networking problems and help in building intuition and physical understanding about the possible implications of multifractal scaling. For example, there are many attractive features that a wavelet-based time-scale analysis has for many aspects of analyzing and interpreting network traffic-related measurements and that have been inaccessible to date due to the limitations of existing techniques. In particular, there exists considerable potential for being able to relate network-, application-, or user-specific features with local irregularities observed in appropriate network measurements and quantified using possibly more refined wavelet techniques than has been presented in this paper. However, while this work puts in place a structure that provides for extensive and novel explorations of many areas of interest to the networking community, we have barely begun exploring its full potential.

Acknowledgments

We thank S. Alexander and S. Gao from AT&T Labs-Research for making the Florham Park traffic collection possible, and we acknowledge the help of many other colleagues at AT&T Labs-Research, especially of J. Friedmanns and A. G. Greenberg, with the overall data collection effort. We are also very grateful to P. Abry and D. Veitch for making available their programs to perform the wavelet-based global scaling analysis, to R. Riedi for helpful conversations and for providing his recent preprints, to E. Waymire for many enlightening discussions concerning cascades and multifractals, and to Ted Sweet and Sid Resnick for pointing out errors in an earlier version of this paper. Finally, we thank three anonymous reviewers for their constructive criticism and invaluable suggestions for improving and clarifying the presentation of the material.

Appendix

Wavelet bases fit naturally into the framework of multiresolution analysis which formalizes the notion of coarse and fine approximations and the increment in information needed to pass from one resolution to another. See [7] for a more thorough treatment of this subject. A *multiresolution analysis (MRA)* of $L^2(\mathbf{R})$ is a decomposition of the space into a chain of closed approximation subspaces

$$\cdots \subset V_l \subset \cdots \subset V_1 \subset V_0 \subset V_{-1} \subset \cdots \subset V_{-l} \cdots$$

such that

$$\overline{\bigcup_{j\in\mathbf{Z}} V_j} = L^2(\mathbf{R}) \quad \text{and} \quad \bigcap_{j\in\mathbf{Z}} V_j = \{0\}.$$

Let P_j denote the orthogonal projection operator onto V_j. We say that $P_j X$ is an approximation to X at scale j or at resolution level 2^j. We have the additional requirements that each subspace V_j is a rescaled version of the base space V_0:

$$X \in V_j \Longleftrightarrow X(2^j \cdot) \in V_0$$

and that the base space V_0 is invariant under integer translations:

$$X \in V_0 \Longrightarrow X(\cdot - n) \in V_0 \quad \text{for all } n \in \mathbf{Z}.$$

Finally, we require that there exists $\phi \in V_0$ (called the scaling function) so that ϕ and all of its integer-translates form an orthonormal basis of V_0. We can conclude that the set $\{\, \phi_{j,k} \,|\, k \in \mathbf{Z} \,\}$ is an orthonormal basis for each subspace V_j. Here $\phi_{j,k}$ denotes a translation and dilation of ϕ:

$$\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k).$$

As a consequence of the above properties, there is an orthonormal wavelet basis

$$\{\, \psi_{j,k} \,|\, j \in \mathbf{Z}, k \in \mathbf{Z} \,\}$$

of $L^2(\mathbf{R})$, $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$, such that for all X in $L^2(\mathbf{R})$

$$P_{j-1}X = P_j X + \sum_{k\in\mathbf{Z}} \langle X, \psi_{j,k} \rangle \, \psi_{j,k},$$

where $\psi(\cdot)$ is a linear combination of translates of $\phi(2\cdot)$. If we define W_j to be the orthogonal complement of V_j in V_{j-1}, then

$$V_{j-1} = V_j \oplus W_j.$$

We have, for each fixed j, an orthonormal basis $\{\, \psi_{j,k} | k \in \mathbf{Z} \,\}$ for W_j. Finally, we may decompose $L^2(\mathbf{R})$ into a direct sum

$$L^2(\mathbf{R}) = V_0 \bigoplus_{j \leq 0} W_j = \bigoplus_{j \in \mathbf{Z}} W_j.$$

The operator Q_j is the orthogonal projection operator onto the space W_j.

The key feature of an MRA is that we can write an approximation of a signal X at scale j (with resolution 2^j) as the sum of a coarser approximation at scale $j+1$ (with resolution 2^{j+1}) and the difference between these two approximations. We may iterate this procedure, writing the approximation at scale $j+1$ as a sum of a coarser approximation and the difference. This procedure can be implemented using a pyramidal filter-bank algorithm which has computational cost $O(N)$ for data of length N. We write the *wavelet decomposition* of a signal X as

$$X = \sum_{k \in \mathbf{Z}} \langle X, \phi_{0,k} \rangle \, \phi_{0,k} + \sum_{j \leq 0} \sum_{k \in \mathbf{Z}} \langle X, \psi_{j,k} \rangle \, \psi_{j,k} = \sum_{j \in \mathbf{Z}} \sum_{k \in \mathbf{Z}} \langle X, \psi_{j,k} \rangle \, \psi_{j,k}.$$

We call the inner products $\langle X, \psi_{j,k} \rangle$ of X with the rescaled and translated copies of the wavelet ψ the *wavelet coefficients* $d_{j,k}$ of X. The set of all wavelet coefficients is generally referred to as the *discrete wavelet transform (DWT)* of the signal X. The coefficient $|d_{j,k}|^2$ measures the amount of energy in a signal X about the time $t_0 = 2^j k$ and about the frequency $2^{-j}\lambda_0$, where λ_0 is a reference frequency which depends on the wavelet ψ.

We can also define the wavelet transform of a measure μ in similar and straightforward manner. The wavelet coefficients $d_{l,k}$ of the measure are the integrals of the wavelets $\psi_{l,k}$ with respect to μ:

$$d_{l,k} = \int \psi_{l,k}(t) \, d\mu(t).$$

The Haar wavelet ψ is a step-function which takes on the value 1 on the first half of the unit interval $I = [0, 1]$ and the value -1 on the second half of I. Suppose that the indices $j_1, \ldots, j_l$ form the dyadic expansion of the point $2^{-l}(n)$, then the dilated version $\psi_{-l,n}(t) = 2^{l/2}\psi(2^l t - n)$ of the basic wavelet is supported on the interval $I(j_1, \dots, j_l)$. Also, the wavelet $\psi_{-l,n}$ is (up to the normalization by a factor of $2^{l/2}$) the characteristic function of the left subinterval $I(j_1, \dots, j_l, 0)$ minus the characteristic function of the right subinterval $I(j_1, \dots, j_l, 1)$.

References

[1] P. Abry, P. Gonçalvès, and P. Flandrin. Wavelets, spectrum analysis and $1/f$ processes. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and statistics*, *Lecture Notes in Statistics*, vol. 105, pages 15–30. Springer-Verlag, 1995.

[2] P. Abry and D. Veitch. Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory*, 44:2–15, 1998.

[3] A. Arneodo. *Wavelets: Theory and Application*, Wavelet analysis of fractals: From the mathematical concept to experimental reality, pages 349–502. Oxford University Press, New York, 1996.

[4] A. Arneodo, E. Bacry, and J. F. Muzy. Random Cascades on Wavelet Dyadic Trees. *Journal of Math. Physics*, 39(8):4124–4164, 1998.

[5] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani. On the multifractal nature of fully developed turbulence and chaotic systems. *J. Phys. A*, 17:3521–3531, 1984.

[6] J. Beran. *Statistics for Long-Memory Processes.* Chapman and Hall, New York, 1994.

[7] I. Daubechies. *Ten lectures on wavelets.* SIAM, Philadelphia, 1992.

[8] C. J. G. Evertsz and B. B. Mandelbrot. *Chaos and Fractals: New Frontiers in Science*, Multifractal measures. Springer-Verlag, 1992.

[9] A. Feldmann, A. C. Gilbert, and W. Willinger. Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic. In *Proc. of the ACM/SIGCOMM'98*, pages 25–38, Vancouver, B.C., 1998.

[10] A. Feldmann, A. C. Gilbert, W. Willinger, and T. G. Kurtz. The changing nature of network traffic: Scaling phenomena. *Computer Communication Review*, 28(2):5–29, 1998.

[11] U. Frisch and G. Parisi. *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, Fully developed turbulence and intermittancy. North-Holland, Amsterdam, 1985.

[12] A. C. Gilbert, W. Willinger, and A. Feldmann. Visualizing multifractal scaling behavior: A simple coloring heuristic. In *Proc. of the 32nd Asilomar conference on signal, systems, and computers*, Pacific Grove, CA, Nov. 1998. to appear.

[13] V. K. Gupta and E. C. Waymire. A statistical analysis of mesoscale rainfall as a random cascade. *Journal of Applied Meteorology*, 32:251–267, 1993.

[14] R. Holley and E. C. Waymire. Multifractal dimensions and scaling exponents for strongly bounded random cascades. *Annals of Applied Probability*, 2:819–845, 1992.

[15] J. P. Kahane and J. Peyriere. Sur certaines martingales de Benoit Mandelbrot. *Adv. in Math.*, 22:131–145, 1976.

[16] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2:1–15, 1994.

[17] J. Levy-Vehel and R. Riedi. *Fractals in Engineering*, Fractional Brownian motion and data traffic modeling: The other end of the spectrum, pages 185–202. Springer-Verlag, Berlin, 1997.

[18] S. Lovejoy and D. Schertzer. Multifractals, universality classes and satellite and radar measurements of cloud and rain fields. *J. Geophys. Res.*, 95:2021–2034, 1990.

[19] B. B. Mandelbrot. Intermittant turbulence in self-similar cascades: Divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, 62:331–358, 1974.

[20] B. B. Mandelbrot. Limit Lognormal Multifractal Measures. In Gotsman, Ne'eman, and Voronel, editors, *The Landau Memorial Conference*, pages 309–340, Tel Aviv, 1990.

[21] P. Mannersalo and I. Norros. Multifractal analysis of real ATM traffic: A first look. Technical Report COST257TD, VTT Information Technology, 1997.

[22] C. Meneveau and K. R. Sreenivasan. Simple multifractal cascade model for fully developed turbulence. *Phys. Rev. Lett.*, 59:1424–1427, 1987.

[23] F. Ben Nasr. Mesures aléatoires de Mandelbrot associées à des substitutions. *C. R. Acad. Sci. Paris Sér. I*, 304(10):255–258, 1987.

[24] R. Riedi. An Improved Multifractal Formalism and Self-similar Measures. *J. Math. Anal. Appl.*, 189:462–490, 1995.

[25] R. Riedi. Introduction to multifractals. Preprint, 1997.

[26] R. Riedi, M. Crouse, V. Ribeiro, and R. Baraniuk. A Multifractal Wavelet Model with Application to TCP Network Traffic. Preprint, 1998.

[27] R. Riedi and B. B. Mandelbrot. Exceptions to the Multifractal Formalism for Discontinuous Measures. *Math. Proc. Cambr. Phil. Soc.*, 123:133–157, 1998.

[28] E. C. Waymire and S. C. Williams. Multiplicative cascades: Dimension spectra and dependence. *The Journal of Fourier Analysis and Applications*, pages 589–609, 1995.