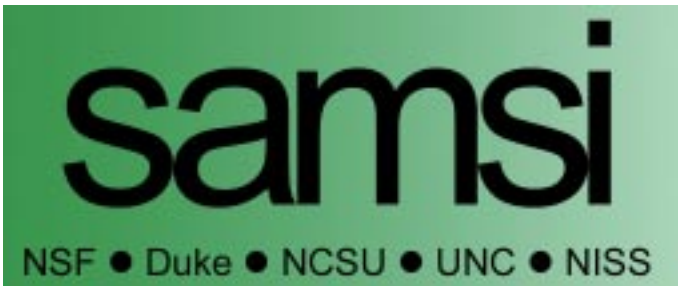

Strengths and Limitations of the Wavelet Spectrum Method in the Analysis of Internet Traffic

Stilian Stoev, Murad Taqqu, Cheolwoo Park and J.S. Marron

Technical Report #2004-8
March 26, 2004

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

Strengths and limitations of the wavelet spectrum method in the analysis of Internet traffic

Stilian Stoev, Boston University
Murad S. Taqqu, Boston University
Cheolwoo Park, SAMSI, Research Triangle Park
J. S. Marron, University of North Carolina at Chapel Hill

February 2, 2004

Abstract

The fluctuations of Internet traffic possess an intricate structure which cannot be simply explained by long-range dependence and self-similarity. In this work, we explore the use of the wavelet spectrum, whose slope is commonly used to estimate the Hurst parameter of long-range dependence. We show that much more than simple slope estimates are needed for detecting important traffic features. In particular, the multi-scale nature of the traffic does not admit simple description of the type attempted by the Hurst parameter. We also demonstrate some practical limitations of the wavelet spectrum. We explore the causes of these limitations using simulated data. This analysis leads us to a better understanding of a number of challenging phenomena observed in real network traffic.

1 Introduction

In the past decade the study of Internet traffic has attracted the interest of researchers in computer science, engineering, applied probability and statistics. A number of studies in the mid 1990's have shown that the classical models for telephone traffic do not apply to the traffic in modern computer tele-communication networks (see Leland, Taqqu, Willinger and Wilson (1993) and Paxson and Floyd (1995)). Statistical analyses indicate that the random fluctuations in the arrival rates, measured in packets or bytes per unit time, are strongly correlated over large time lags, i.e. are *long-range dependent*.

The long-range dependence property of the traffic fluctuations has important implications on the performance, design and dimensioning of the network. A simple, direct parameter characterizing the degree of long-range dependence is the Hurst parameter. This parameter controls the regularity and the magnitude of the fluctuations in the data on medium and large time scales (for a precise definition and more references on the long-range dependence phenomenon see Section 2.1, below).

A number of methods have been proposed to estimate the Hurst parameter. Some of the most popular include the *aggregated variance*, *local Whittle* and the *wavelet-based* methods. These and other methods are described and discussed in Taqqu, Teverovsky and Willinger (1995), Taqqu and Teverovsky (1998), Abry, Flandrin, Taqqu and Veitch (2003), Bardet, Lang, Oppenheim,

Philippe, Stoev and Taqqu (2003) and Hernández-Campos, Le, Marron, Park, Park, Pipiras, Smith, Smith, Trovero and Zhu (2004). This last study performs an extensive comparison of the most successful methods of estimation by using simulated, synthetic and real Internet traffic data sets. It reveals a number of important challenges which one faces when estimating the long-range dependence parameter in Internet data traffic traces. Our goal in this paper is to explore some of these challenges in more detail by using the wavelet spectrum method.

Wavelets capture both time and frequency features in the data and often provide a richer picture than the classical Fourier analysis. Since the seminal work of Abry and Veitch (1998), they have become a very popular tool for studying the long-range dependence properties of network traffic. The self-similar scaling which is often encountered in the presence of long-range dependence is naturally captured by the wavelet spectrum of the data. This scaling is used to define the wavelet estimator of the Hurst parameter (see Sections 2.1 and 3 below), in terms of the slope of the wavelet spectrum.

The Hurst parameter, however, characterizes the dependence of the traffic only on large scales. Very often, the wavelet spectrum contains additional useful information about the dependence in the data on medium and small time scales. It can also capture important traffic features such as periodic components, deterministic breaks in the mean traffic rate and other intricate non-stationary features in the data. This information is being neglected when one uses only the estimate of the Hurst long-range dependence parameter to characterize network traffic.

In this paper, we focus on the wavelet spectrum tool for the analysis of Internet traffic data. Our main goal here is to first demonstrate how this tool can be used not only to estimate the Hurst parameter reliably but also to detect important details in the data. We then explore the advantages and limitations of the wavelet estimators and of the wavelet spectrum.

The paper is structured as follows. In Section 2.1, we define the Hurst parameter and provide intuition on some basic notions related to the long-range dependence phenomenon. In Section 2.2, we discuss in detail several wavelet spectra for real network traffic traces. The first of these is "well-behaved" in the sense of showing features predicted by current prevalent models for traffic. In particular, the Hurst parameter captures most of the important dependence features in the data. We also demonstrate that the classical *fractional Gaussian noise* time series provides an effective model for this example of traffic, over a wide range of time scales (see Figures 1 and 2 below). But other examples in Section 2.2 show more challenging phenomena in the data (see Figures 3 – 6, below). These wavelet spectra exhibit unusual features which are inconsistent with the classical models of long-range dependence. Our goal in the rest of the paper is to understand the origins of such features and the limitations of wavelet spectrum analysis.

In Section 3, we start by briefly introducing the discrete wavelet transform of a *signal* and discuss how it can be used in practice to analyze discrete time series. We then present the classical wavelet estimator of the Hurst parameter of a second order stationary time series. In Section 3.3, we focus on the *wavelet spectrum* and discuss its connection to the classical Fourier spectrum of a time series.

Section 4 is devoted to exploring the wavelet spectrum using controlled simulated examples. We start by demonstrating that, in ideal situations, the wavelet estimator is comparable with some of the best available quasi-maximum likelihood estimators such as the local Whittle estimator. We also show that, unlike the local Whittle estimator, the wavelet estimator is rather robust with respect to smooth, non-stationary trends in the data (see Figures 7 and 8, below). Then,

in the analyses presented in Figures 11 and 12, we explore the effects of deterministic trends on the wavelet spectrum. We also show how some short-range dependent stochastic models affect the wavelet spectrum. An important observation is that each of these can induce structure in the wavelet spectrum that looks roughly the same as that of long-range dependence.

Using the insights developed in Section 4, we revisit in Section 5 the challenging examples of wavelet spectra of network traffic traces, shown in Section 2. We interpret some of the observed features and identify the limitations of the wavelet spectrum in practice. These limitations, suggest the need of further tools. We conclude by describing briefly two of them, which are presented in detail in Park, Godtliebsen, Taqqu, Stoev and Marron (2004) and Stoev, Taqqu, Park, Michailidis and Marron (2004).

Conclusion. The wavelet estimator of the Hurst parameter stands out as one of the most reliable estimators in practice. It is semi-parametric, in nature, and quite robust with respect to smooth polynomial trends in the data. This estimator is based on the wavelet spectrum, which in turn captures additional important characteristics of the dependence structure of the data as well as deterministic fluctuations. And yet, in practice, when used blindly both the wavelet Hurst parameter estimator and the wavelet spectrum can mislead the practitioner. A traffic trace with a number of deterministic shifts in the mean rate, for example, results in a *steep* wavelet spectrum which leads to overestimating the Hurst parameter. In addition, complex short-range dependent stochastic models can have an even more subtle effect on the wavelet spectrum and also lead to overestimating the Hurst parameter.

These limitations cannot be overcome if one focuses only on the wavelet spectrum of the data. One can gain an important insight, however, by making a deeper analysis of the features in the wavelet coefficients of the data. This can be done by using the scale-space type analysis in Park, Godtliebsen, Taqqu, Stoev and Marron (2004). Another possibility is to perform a *local* analysis of the long-range dependence so that one reduces the effect of intricate non-stationarity and in addition obtains a richer picture of the local dependence structure in the data. We explore this approach in Stoev, Taqqu, Park, Michailidis and Marron (2004).

2 Motivating examples

We first briefly review some basic notions related to the long-range dependence phenomenon and then present examples of Internet traces and their corresponding wavelet spectra. The wavelet spectrum of a stationary time series and the corresponding wavelet-based estimator of the Hurst long-range dependence parameter are briefly presented in Section 3.

2.1 Long-range dependence

A second order stationary time series $Y = \{Y(k)\}_{k\in\mathbb{Z}}$ is said to be long-range dependent (LRD), if the series of its covariances is non-summable:

$$\sum_{k=0}^{\infty} \mathrm{Cov}(Y(k), Y(0)) = \infty.$$

In general, LRD time series Y are also *asymptotically self-similar*. Namely, suppose that $\mathbb{E}Y(k) = 0,\ k \in \mathbb{Z}$ and let

$$Y_n(k) := Y^{(n)}(k) := Y(nk+1) + Y(nk+2) + \cdots + Y(nk+n), \quad k \in \mathbb{Z},$$

denote the aggregations of the time series Y over blocks of size $n \in \mathbb{N}$. The time series Y is called *asymptotically self-similar*, if, as $n \to \infty$, one has

$$\{Y_n^{(m)}(k),\ k \in \mathbb{Z}\} \sim_d \{m^H Y_n(k),\ k \in \mathbb{Z}\}, \tag{2.1}$$

where $\sim_d$ means asymptotic equivalence of the finite-dimensional distributions. The parameter H is called the Hurst parameter of the time series Y. It controls the asymptotic self-similarity scaling as well as the degree of long-range dependence.

The process Y_n can be interpreted as a "zoomed-out" version of the time series Y viewed on the time scale n. Thus Relation (2.1) means that on large time scales, the process Y looks (approximately) statistically the same, up to a multiplicative factor, which depends on the time scale. The time series Y_n is called *self-similar* when Relation (2.1) holds exactly.

The fractional Gaussian noise (FGN), $G_H = \{G_H(k)\}_{k\in\mathbb{Z}}$, is a fundamental model which exhibits both long-range dependence and self-similarity. It is a stationary Gaussian process with mean zero and covariances:

$$\mathrm{Cov}\Big(G_H(k), G_H(0)\Big) = \frac{\sigma^2}{2}\Big(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\Big),$$

where $\sigma^2 = \mathbb{E}G_H(0)^2 = \mathrm{Var}(G_H(0))$ and where the parameter H is restricted to the range

$$0 < H < 1.$$

One can show that the FGN is the only (up to rescaling) stationary self-similar Gaussian time series. When $H = 1/2$, the FGN time series G_H becomes uncorrelated white noise. However, it is long-range dependent, when $1/2 < H < 1$.

Many classical statistical results do not work in the presence of long-range dependence. For more details on the fractional Gaussian noise and its role in the theory of long-range dependence, see, for example, Taqqu (2003) and the references therein.

2.2 Examples of wavelet spectra

In Figure 1, we present one packet trace collected from the UNC main link in 2002 on April 13, Saturday from 19:30 to 21:30. We focus on the time series $Y = \{Y(k),\ k = 1, 2, \ldots, N\}$ of the number of packets arriving on the link every 1 millisecond. The wavelet spectrum of this time series is shown on the bottom-right plot of Figure 1. This plot displays the statistics S_j as a function of the scale j (see Section 3 below for more details). The statistic S_j quantifies the *energy* of the time series $\{Y(k)\}$ concentrated at a frequency range corresponding to scale j. The *scale* j can be related to a frequency in the spectral density of the time series Y. Large j correspond to low frequencies and small j to high frequencies. The energy of a white noise time series is uniformly distributed across all scales or frequencies and hence its wavelet spectrum is essentially constant or *flat*. The short-term dependence structure of a time series is related

to its high frequency behavior and therefore affects its wavelet spectrum only at small scales j. However, if the time series does not have strong dependence at large time lags, the wavelet spectrum at large scales j continues to be flat or close to a constant. As seen in Figure 1, this is not the case with the time series of Internet traffic. The linear, non-constant scaling of the wavelet spectrum at large scales $j = 10,\ 11,\ \ldots,\ 20$ indicates that the time series Y is long-range dependent and, in fact, approximately self-similar.

Figure 1 around here

The Hurst parameter H of the time series Y can be estimated by using the slope of the wavelet spectrum at large scales j (for more details, see Section 3 below). When the value of the parameter H is in the range $1/2 < H < 1$, the time series exhibits long-range dependence and it cannot be modeled by using classical telephony models (see, for example, Leland, Taqqu, Willinger and Wilson (1993) and Paxson and Floyd (1995)).

The bottom-left plot in Figure 1 displays the time series obtained by using block-wise aggregation over blocks of size 1000 observations of the original time series appearing on the top plot. It thus corresponds to the number of packets per 1 second time intervals of the traffic trace. The marginal distributions of the aggregated time series are essentially symmetric and approximately Gaussian (see also Figure 2).

In Figure 2, we compare the time series of aggregated traffic (at 1 second, that is, aggregated by a factor of 1000) shown on the bottom left plot of Figure 1 to a *fitted* FGN time series. Roughly, the wavelet spectrum of the aggregated time series ($1000 \approx 2^{10}$), corresponds to the large-scale components ($10 \leq j \leq 20$) of the wavelet spectrum of the original time series. As seen on the bottom-right plot of Figure 2 this is indeed the case, since the wavelet spectrum therein is essentially linear and has about the same slope.

The wavelet spectrum of the original 1-millisecond traffic trace at small time scales $1 \leq j \leq 5$ appears to be flat, which suggests that the traffic fluctuations are rather weakly correlated on small time scales. This feature is typical for many other traffic traces, in particular, in case of heavily multiplexed back-bone links and it has already been noted in the literature (see, for example, Cao, Cleveland, Lin and Sun (2002) and Zhang, Ribeiro, Moon and Diot (2003)).

At large scales, however, the long-range dependence of traffic is ubiquitous. Figure 2 shows that the short-term dependence structure has been washed-out after aggregation and the traffic appears to be statistically self-similar and long-range dependent.

Figure 2 around here

The analysis shown in Figure 2, indicates that the fractional Gaussian noise model captures very well the key features of the network traffic trace on large time scales, that is, its approximate self-similarity and long-range dependence behavior. The empirical distribution of the packet trace time series is essentially normal as seen from the QQ-plot in Figure 2, with the exception of the far tails which appear heavier than the normal ones. The downwards spike at time 3000 of the packet trace contributes to the outliers in the QQ-plot.

We also display several other wavelet spectra of Internet traffic traces. Whereas the wavelet spectra shown in Figures 1 and 2 appear consistent with the traditional fractional Gaussian

noise type models of network traffic, these other traces possess a number of unsettling features.

Firstly, the estimated Hurst exponent H, based on the wavelet spectra, is sometimes greater than 1, that is, outside the theoretically admissible range $0 < H < 1$, as in Figures 3 and 6. The wavelet spectrum of these traces, however, still exhibits linear, long-range dependence-type scaling at large scales. This shows that it is inadvisable to blindly estimate Hurst parameters to understand the long range dependence properties of internet traffic data, and that the classical long range dependence model, parametrized by a single Hurst parameter, is insufficient to characterize some of the phenomena present in such data.

Secondly, as seen in Figures 3 – 6, the wavelet spectrum contains (at all scales) important information about the second order properties of the traffic traces, which are not captured by the Hurst parameter. Namely, note the *spike* at scales $j = 11$ and 12 of the wavelet spectrum in Figure 3 and also the *unusually steep* slope in the spectrum at scales $5 \leq j \leq 7$ in Figure 5.

Figure 3 around here

Figure 4 around here

Figure 5 around here

Figure 6 around here

We provide an explanation of these strange phenomena in Section 5. The explanation is based on understanding how such features can be generated, using a number of simulated examples, explored in Section 4. First a quick wavelet overview, together with a precise definition of the wavelet spectrum is given in Section 3.

The Internet traffic data sets, displayed on Figures 1–6 are freely available from `http://www-dirt.cs.unc.edu/ts/` (add `unc02_ts`, `unc03_ts` or `Abilene-I_ts`). The file names corresponding to the data sets in Figures 1–6 are:
`2002_Apr_13_Sat_1930.7260.sk1.1ms.B_P.ts.gz`,
`2002_Apr_13_Sat_1300.7260.sk1.1ms.B_P.ts.gz`,
`2002_Apr_11_Thu_1300.7260.sk1.1ms.B_P.ts.gz`,
`2003_Apr_10_Thu_1500.3660.em1.10ms.B_P.ts`,
and `IPLS-KSCY-1.ts.gz`, respectively.

3 Wavelet Spectrum

We introduce here the discrete wavelet transform of a deterministic or a random signal, discuss some of its classical applications and its computation in practice. We then define the wavelet spectrum and recall the *classical* wavelet estimator of the Hurst parameter, proposed by Abry and Veitch (1998).

3.1 The discrete wavelet transform

We start with the discrete wavelet transform in the space of square integrable functions $L^2(dt) := \{f : \mathbb{R} \mapsto \mathbb{R},\ \int_{\mathbb{R}} f^2(t)dt < \infty\}$. More details on the discrete wavelet transform,

and its applications can be found, for example, in the book of Mallat (1998).

Let $\psi(t) \in L^2(dt)$ be a square integrable function with $M \in \mathbb{N}$ zero moments, $M \geq 1$, that is

$$\int_{\mathbb{R}} t^m \psi(t) dt = 0, \quad \text{for all } m = 0, 1, \ldots, M-1. \tag{3.1}$$

Consider the functions

$$\psi_{j,k}(t) := 2^{-j/2} \psi(2^{-j}t - k), \ j, k \in \mathbb{Z},$$

obtained by dyadic *dilations* and integer *translations* of ψ.

The function ψ is called an orthogonal *mother wavelet*, if the set $\{\psi_{j,k}(t)\}_{j,k\in\mathbb{Z}}$ is an *orthonormal basis* of the Hilbert space $L^2(dt)$. That is, if for any signal $g(t) \in L^2(dt)$, we have

$$g(t) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} d_{j,k}(g) \psi_{j,k}(t), \tag{3.2}$$

where the coefficients $d_{j,k}(g)$ are given by

$$d_{j,k}(g) := \int_{\mathbb{R}} g(t) \psi_{j,k}(t) dt = \int_{\mathbb{R}} g(t) 2^{-j/2} \psi(2^{-j}t - k) dt, \ j, k \in \mathbb{Z}. \tag{3.3}$$

and where the series in (3.2) converges in the $L^2(dt)$ sense. The basis $\{\psi_{j,k}\}_{j,k}$ is called an orthogonal wavelet basis of $L^2(dt)$ and the coefficients $d_{j,k}(g)$ are called the *discrete wavelet transform* coefficients of the signal g.

The indices j and k of the wavelet coefficients $d_{j,k}(g)$ are called *scale* and *location*, respectively. When the scale j is large, the effective support of the function $\psi_{j,k}$ in (3.3) is wide and then the coefficient $d_{j,k}(g)$ captures *low frequency* or *coarse scale* behavior of the signal $g(t)$. Conversely, at small scales j, the $d_{j,k}(g)$s quantify the *high-frequency* or *fine scale* details of the signal $g(t)$. The index k controls the time-location of the function $\psi_{j,k}(t) = \psi_{j,0}(t - 2^j k)$ via translation.

Thus the set of DWT coefficients $d_{j,k}(g), \ j, k \in \mathbb{Z}$ captures both time and frequency behavior of the signal g. This is a powerful feature which distinguishes the wavelet analysis from the classical Fourier analysis of signals (see, for example, Ch. 1 in Daubechies (1992) and Flandrin (1999)).

The mere existence of functions ψ, which generate orthogonal wavelet bases is not trivial. The Haar function $\psi(t) = 1_{[0,1/2)}(t) - 1_{[1/2,1)}(t), \ t \in \mathbb{R}$ is the simplest example of an orthogonal wavelet. In the past two decades many families of orthogonal wavelets have been constructed by using elegant ideas from the theory of *multiresolution analysis* (MRA) (see Mallat (1989, 1998)). Wavelets have interesting approximation theoretical properties as well as many important applications in numerical analysis, signal processing and statistics (see, for example, Cohen (2003), Vetterli and Kovacevic (1995), Ogden (1996) and Vidakovic (1999)).

In applications, the wavelet function ψ is chosen to be well-localized in the time and in the frequency domains. In particular, the Daubechies wavelets are very useful in practice, since they have *compact* support in the time domain and since their support in the frequency domain is *well-localized*. These wavelets can be chosen to be smooth and with several zero moments M (for more details, see Ch. 6 in Daubechies (1992)). For simplicity, in the sequel we suppose that ψ is a Daubechies wavelet with compact support and $M \geq 1$ zero moments.

The multiresolution analysis involves, in addition, a *father wavelet* function $\phi(t) \in L^2(dt)$, which is orthogonal to the *mother wavelet* ψ. The integer translates $\phi_k(t) := \phi(t-k),\ k \in \mathbb{Z}$ of the father wavelet are required to be orthonormal. Furthermore, the function ϕ is chosen so that $\{\phi_k\}_{k\in\mathbb{Z}}$ becomes an equivalent basis to the mother wavelet basis, $\{\psi_{j,k},\ j = 1, 2, \ldots,\ k \in \mathbb{Z}\}$. The expansion (3.2) can be equivalently expressed by using the father wavelet basis, as follows:

$$g(t) = \sum_{k\in\mathbb{Z}} a_k(g)\phi_k(t) + \sum_{j\le 0}\sum_{k\in\mathbb{Z}} d_{j,k}(g)\psi_{j,k}(t), \tag{3.4}$$

where the coefficients $a_k(g)$ are:

$$a_k(g) = \int_{\mathbb{R}} g(t)\phi_k(t)dt = \int_{\mathbb{R}} g(t)\phi(t-k)dt,\ k \in \mathbb{Z}.$$

The first term in (3.4) is equal to the contribution of the coarse scale wavelets, that is,

$$g_0(t) = \sum_{k\in\mathbb{Z}} a_k(g)\phi_k(t) = \sum_{j=1}^{\infty}\sum_{k\in\mathbb{Z}} d_{j,k}(g)\psi_{j,k}(t). \tag{3.5}$$

Therefore the set of functions $\{\phi_k(t),\ k \in \mathbb{Z}\} \cup \{\psi_{j,k}(t),\ j \le 0,\ k \in \mathbb{Z}\}$ is also an orthonormal basis of $L^2(dt)$.

The father wavelet associated with the Haar mother wavelet $\psi(t) = 1_{[0,1/2)}(t) - 1_{[1/2,1)}(t)$ is the box-car function $\phi(t) := 1_{[0,1)}(t),\ t \in \mathbb{R}$. Observe that in this case the functions $\{\phi_k(t)\}_k$ are orthogonal and they are indeed orthogonal to all finer scale wavelets $\psi_{j,k}(t) = 2^{-j/2}(1_{[2^jk,2^j(k+1/2))}(t) - 1_{[2^j(k+1/2),2^j(k+1))}(t))$, where the index $j \le 0$ is a *negative* integer or zero.

The first term in (3.4) can be viewed as a coarse scale approximation to the signal $g(t)$. The resolution of this approximation is determined by the time scale of the father wavelet ϕ. The second term in (3.4), involves the successive fine scale or higher frequency details of the signal $g(t)$, which are captured by the wavelet coefficients $d_{j,k}(g),\ j \le 0,\ k \in \mathbb{Z}$. Thus often the coefficients $a_k(g)$ are called *approximation* coefficients and the $d_{j,k}(g)$s are called *detail* wavelet coefficients.

3.2 Applications and practical issues

Given a finite discrete-time sample $g(k),\ k = 1, \ldots, N$ of a signal $g(t) \in L^2(dt)$, one cannot obtain exactly the wavelet coefficients $d_{j,k}(g)$, since they involve integrals. Furthermore, the highest frequency details of the signal are limited by the sampling rate. Thus, in practice, one can only compute approximations to the coarse scale wavelet coefficients $d_{j,k}(g),\ j \ge 1$ appearing in (3.5). This can be done, efficiently, by using Mallat's *fast discrete wavelet transform* algorithm.

Namely, set

$$a_k := g(k), \quad k = 1, \ldots, N, \tag{3.6}$$

and consider the function

$$\widetilde{g}_0(t) := \sum_{k=1}^{N} a_k\phi(t-k)$$

to be an *approximation* to the projection $g_0(t)$ of the signal $g(t)$ given in Relation (3.5) above. Using a computationally efficient algorithm, involving discrete convolutions and sub-sampling, one can compute the coarser scale wavelet coefficients involved in the sum on the right-hand side of (3.5). Since we work with compactly supported wavelets, only a finite number of these coefficients will be non-zero. One typically obtains a "triangular" array of about N approximate wavelet coefficients:

$$\widetilde{d}_{j,k}(g),\ k = 1, \ldots, N_j,$$

where

$$N_j \approx N/2^j, \quad \text{and} \quad j = 1, \ldots, J,\ J \leq [\log_2(N)].$$

That is, one has about $J \approx [\log_2(N)]$ dyadic scales and the number, N_j, of wavelet coefficients on scale j is about $N/2^j$. For more details on Mallat's algorithm and its applications, see, for example, Ch. 7.3.1 in Mallat (1998).

In practice, the error made on the *initialization step* (3.6) of Mallat's algorithm becomes essentially negligible on large scales j. The resulting approximate wavelet coefficients $\widetilde{d}_{j,k}(g)$ often represent very well the important features of the theoretical wavelet coefficients $d_{j,k}(g)$. In applications, when additional information on the signal g is available, one can reduce the error due to initialization, by setting a_k in (3.6) to be equal to some pre-filtered version of the sampled signal $g(k)$ (see the Remarks below).

Mallat's fast discrete wavelet transform algorithm can be inverted and the original set of approximation coefficients $a(k),\ k = 1, \ldots, N$ can be recovered. Furthermore, for compactly supported wavelets, Mallat's algorithm and its inverse have time and memory complexity of order $\mathcal{O}(N)$. The efficiency of these algorithms and the useful time/frequency localization property of wavelets make them a very useful tool in engineering and signal processing, which sometimes outperforms the classical Fourier transform tools (see, for example, Rioul and Vetterli (1991)).

The wavelet expansions of large classes of deterministic signals are *sparse*. That is, there are very few coefficients of large magnitude which capture most features of the signal. In particular, far fewer wavelet coefficients are required to represent well functions with discontinuities or cusps, as opposed to Fourier coefficients, for example. These features lead to the development efficient algorithms for *image compression* and numerous statistical techniques for smoothing and denoising data (see, for example, Ch. 11 and 9 in Mallat (1998)).

3.3 Wavelet spectrum and the estimation of the Hurst parameter

Consider now a second order stationary stochastic signal $Y = \{Y(t)\}$ and let

$$Y(t) \quad \propto \quad \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{j,k}(Y) \psi_{j,k}(t) \tag{3.7}$$

be its *formal* wavelet series, where the *wavelet coefficients* $d_{j,k}(Y)$ are defined as in (3.3) with g replaced by Y. Since the wavelet ψ has compact support, the integrals in (3.3) converge in the L^2-sense and the random variables $d_{j,k}(Y)$ are well defined. (Observe, that $Y(t)$ need not belong to $L^2(dt)$ and therefore in (3.7), we may not have an equality.)

The stationarity of Y implies the stationarity of the wavelet coefficients $\{d_{j,k}(Y)\}_{k \in \mathbb{Z}}$, for all scales $j \in \mathbb{Z}$. Let $\mathcal{E}_j$ denote the mean *energy* of the wavelet coefficients at scale j, that is,

$$\mathcal{E}_j := \mathbb{E} d_{j,k}^2(Y), \quad j \in \mathbb{Z}.$$

The wavelet coefficients $d_{j,k}(Y)$ on scale j capture frequency details of Y in a frequency range about 2^j. Thus, one expects that the *wavelet energy spectrum* $\mathcal{E}_j,\ j \in \mathbb{Z}$ can be related to the classical spectral density of the process Y.

Indeed, suppose that Y has a spectral density $f_Y(\xi),\ \xi \in \mathbb{R}$, that is

$$f_Y(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi t} r_Y(t) dt,$$

where $r_Y(t) := \mathrm{Cov}(Y(t+s), Y(s)),\ t \in \mathbb{R}$ denotes the auto-covariance function of Y. By using the Parseval identity and a change of variables one can show that

$$\begin{aligned} \mathbb{E}d_{j,k}^2(Y) &= \int_{\mathbb{R}} \psi_{j,k}(t) \int_{\mathbb{R}} \psi_{j,k}(s) r_Y(t-s) ds dt = \int_{\mathbb{R}} |\widehat{\psi_{j,k}}(\xi)|^2 f_Y(\xi) d\xi, \\ &= 2^j \int_{\mathbb{R}} |\widehat{\psi}(2^j \xi)|^2 f_Y(\xi) d\xi = \int_{\mathbb{R}} |\widehat{\psi}(\eta)|^2 f_Y(\eta/2^j) d\eta. \end{aligned} \tag{3.8}$$

The last expression in Relation (3.8) relates the mean energy $\mathcal{E}_j$ of the wavelet coefficients $d_{j,k}(Y)$ to the spectral density of the stationary signal $Y(t)$. For large scales j, the function $f_Y(\eta/2^j),\ \eta \in \mathbb{R}$ can be viewed as an *expanded* or *zoomed* version of the spectral density $f_Y(\eta)$ around the zero frequencies. Therefore, the integral in the right-hand side of (3.8) picks out the spectral behavior of Y at low frequencies. Conversely, the energy $\mathcal{E}_j$ on small scales j corresponds to the high-frequencies in the the spectral density f_Y.

Now let Y be a long-range dependent process, that is, with a spectral density which is unbounded at the origin. Suppose that, as $t \to 0$,

$$f_Y(t) \sim c_f \frac{1}{|t|^\alpha}, \quad \text{with } \ 0 < \alpha < 1, \tag{3.9}$$

where $c_f > 0$ and where $\sim$ means that the ratio of the left and the right-hand sides converges to 1. Relations (3.8) and (3.9) imply that, as $j \to \infty$,

$$\mathbb{E}d_{j,k}^2(Y) \sim c_f \int_{\mathbb{R}} |\widehat{\psi}(\eta)|^2 |\eta/2^j|^{-\alpha} d\eta = c_f C 2^{j\alpha}, \tag{3.10}$$

where $C = C(\psi, \alpha) = \int_{\mathbb{R}} |\widehat{\psi}(\eta)|^2 |\eta|^{-\alpha} d\eta$.

Relation (3.10) suggests that the long-range dependence parameter α, and correspondingly the related Hurst parameter

$$H = (1+\alpha)/2$$

can be estimated by using the fact that

$$\log_2(\mathcal{E}_j) = \log_2(\mathbb{E}d_{j,k}^2(Y)) \sim (2H-1)j + \text{const}\,, \quad \text{as } j \to \infty. \tag{3.11}$$

As discussed above, given a finite sample $Y(k),\ k = 1, 2, \ldots, N$ of the signal Y, by using Mallat's algorithm, one can obtain a triangular array of approximate wavelet coefficients $d_{j,k}(Y),\ k = 1, \ldots, N_j,\ j = 1, \ldots, J$. Thus, one can estimate $\log_2(\mathcal{E}_j)$ by using the sample energy of these coefficients:

$$S_j := \log_2 \Big(\frac{1}{N_j} \sum_{k=1}^{N_j} d_{j,k}^2(Y) \Big) \approx \log_2(\mathcal{E}_j). \tag{3.12}$$

For large N_j we expect the statistics S_j to approximate well the quantity $\log_2(\mathcal{E}_j)$. The set of statistics $S_j,\ j = 1, \ldots, J$ is often called the *log-scale wavelet spectrum* of the time series $\{Y(k),\ k = 1, \ldots, N\}$ or the *wavelet spectrum*, in short. In view of Relation (3.11), the wavelet spectrum of long-range dependent random signals scales linearly on large scales with slope $\alpha = 2H - 1$. One can thus define the following estimator of the Hurst parameter H:

$$\widehat{H}_{[j_1,j_2]} = \frac{1}{2} \sum_{j=j_1}^{j_2} w_j S_j + \frac{1}{2}, \tag{3.13}$$

where $1 \leq j_1 < j_2 \leq J$ and where $w_j,\ j = j_1, \ldots, j_2$, are such that

$$\sum_j w_j = 0 \quad \text{and} \quad \sum_j j w_j = 1.$$

That is, H is estimated from the slope $2\widehat{H} - 1$ of a linear regression fit of the wavelet spectrum S_j against j, over the range of scales $j_1, \ldots, j_2$ (see Figure 7 in Section 4, below).

The wavelet estimator $\widehat{H}_{[j_1,j_2]}$ in (3.13) performs well when the data are not far from a FGN. Empirical studies indicate that it is very robust with respect to smooth deterministic trends and changes in the short-range dependence structure of the time series $Y(k),\ k = 1, \ldots, N$ (see, for example, Figure 7 below, Abry, Flandrin, Taqqu and Veitch (2003) and Bardet, Lang, Oppenheim, Philippe, Stoev and Taqqu (2003)).

Theoretically, the linear scaling of the wavelet spectrum $\{S_j\}$ occurs at large scales j (see (3.11)). Therefore, when estimating H in practice, one should choose a range of relatively large scales $[j_1, j_2]$. Nevertheless, in many cases the left-hand side in Relation (3.11) scales almost exactly as the right-hand side therein. For more details on the choice of scales j_1 and j_2, see the Remarks below.

The asymptotic statistical properties of the wavelet estimator $\widehat{H}_{[j_1,j_2]}$ have been established when Y is a Gaussian process, under general conditions (see Bardet, Lang, Moulines and Soulier (2000)). These authors have shown that the estimator $\widehat{H}_{[j_1,j_2]}$ is consistent, as

$$j_1 \ \text{ and } \ N/2^{j_2} \longrightarrow \infty.$$

Under additional conditions on the rate of growth of the scales j_1 and j_2, as a function of the sample size N, one also obtains the asymptotic normality of the estimator $\widehat{H}$ (for more details, see Bardet *et al.* (2000)).

These results follow from a key statistical property of the wavelet coefficients $d_{j,k}(Y)$. Namely, that $d_{j,k}(Y),\ k \in \mathbb{Z}$ are short-range dependent. More precisely, under general conditions on Y, it follows that for sufficiently large fixed j,

$$\mathrm{Cov}(d_{j,k}(Y), d_{j,m}(Y)) = \mathbb{E} d_{j,k}(Y) d_{j,m}(Y) = \mathcal{O}(|k - m|^{2H-1-2M}), \tag{3.14}$$

where M denotes the number of zero moments of the wavelet ψ (see Bardet *et al.* (2000) and the related results in Kim and Tewfik (1992) and Bardet (2002)). Thus, when ψ has sufficiently large number of zero moments M, the wavelet coefficients $d_{j,k}(Y),\ k \in \mathbb{Z}$ can be considered essentially uncorrelated and the sample approximation of the mean energy $\mathcal{E}_j$ in (3.12) becomes very good in practice.

Wavelet estimators for the Hurst and/or self-similarity parameter can be also defined for stable infinite variance processes, such as the *fractional autoregressive integrated moving average* (FARIMA) time series with stable innovations and the *linear fractional stable motion* (see Stoev and Taqqu (2003), Pipiras, Taqqu and Abry (2001) and Stoev, Pipiras and Taqqu (2002)).

In Section 4, we discuss the positive features of the wavelet estimator $\widehat{H}$ and study extensively its limitations through simulated examples. In the following remarks we list some of the key issues encountered when using the estimator $\widehat{H}$ in practice (see also Veitch and Abry (1999)).

Remarks

1. *(Initialization of Mallat's algorithm)* Veitch, Taqqu and Abry (2000) propose an *initialization* procedure for Mallat's algorithm, which is well-suited to the case of fractional Gaussian noise processes Y.

2. *(The choice of wavelet ψ)* Usually one uses Daubechies wavelets ψ with compact support and at least $M = 2$ zero moments. These wavelets are well-localized in the time and frequency domains and they yield essentially *decorrelated* wavelet coefficients $d_{j,k}(Y),\ k \in \mathbb{Z}$. Numerical experiments have shown that choosing a very large number of zero moments M (eg $M > 5$) does not improve the performance of the estimators of H. Wavelets with infinite support, such as the Mexican Hat or the Meyer wavelets can be also used. One expects, however, the estimates of H to have higher sample variance than in the case when wavelets with compact support are used.

3. *(Bias correction)* Suppose that $\mathbb{E}d^2_{j,k}(Y) = \text{const}\, 2^{j(2H-1)}$ and that for all j, the wavelet coefficients $d_{j,k}(Y),\ k \in \mathbb{Z}$ are independent and Gaussian. Then, the estimator $\widehat{H}$ in (3.13) is asymptotically unbiased, as $N_j \to \infty,\ j = j_1, \ldots, j_2$. For finite, N_j, however, $\widehat{H}$ is biased, due to the non-linearity of the function $\log_2$, that is, since $\mathbb{E}S_j \neq \log_2(\mathbb{E}d^2_{j,k}(Y))$. Under the above idealized assumptions, Veitch and Abry (1999) have corrected the bias of the statistics S_j:

$$\widetilde{S}_j := \log_2\Big(\frac{1}{N_j}\sum_{k=1}^{N_j} d^2_{j,k}(Y)\Big) + g(N_j, j). \tag{3.15}$$

   The bias correction term $g(N_j, j)$ is asymptotic to $-1/(\ln(2)N_j)$, as $N_j \to \infty$ and can be essentially ignored for large values of N_j. It has only a small effect on the largest scales j (see Section II.B in Abry and Veitch (1998), for more details).

4. *(The choice of weights w_j)* Given a finite sample $Y(1), \ldots, Y(N)$ of the process Y, one can extract only about $N_j \approx N/2^j$ wavelet coefficients at scale j, $j = 1, \ldots, J \approx [\log_2(N)]$. Thus, the statistics $N_j^{-1}\sum_{k=1}^{N_j} d^2_{j,k}(Y)$, involved in (3.13), have larger variance at large scales j, rather than at small scales. One can compensate for this by choosing appropriately the weights w_j in (3.13), so that the estimator $\widehat{H}$ has a lower sample variance (see Veitch and Abry (1999)).

5. *(The choice of the range of scales $[j_1,\ j_2]$)* As argued in the previous point, given a time series $Y(k),\ k = 1, \ldots, N$ of length N, one has at most $J \approx [\log_2(N)]$ available scales $j = 1, \ldots, J$. Theoretically, the long-range dependence scaling occurs at large scales j,

which correspond to low frequencies in the spectrum of the process Y. In practice, one can have different and often non-linear scaling of $\mathbb{E}d^2_{j,k}(Y)$ at the small scales j. The wavelet coefficients at the small scales j capture the short-term dependence structure of the process Y, which in principle, is not related to the Hurst parameter H.

Therefore, in practice one chooses all sufficiently large available scales $[j_1, \ldots, j_2]$, where the log-energy $\log_2 \mathbb{E}d^2_{j,k}(Y)$ appears to scale linearly. In the case of time series $Y(k)$ which exhibit exact self-similarity such as the FGN, one can choose essentially all available scales.

The choice of scales is a very difficult problem in practice and is analogous to the bias-variance trade-off when choosing the bandwidth in kernel density estimation. Veitch, Taqqu and Abry (2003) have proposed a method for automatic selection of the range of scales $[j_1, j_2]$ based on the mean square error loss function.

6. *(Implementation)* Veitch and Abry have implemented in Matlab the wavelet estimator discussed above. Their code has become standard among practitioners and is freely available from `http://www.cubinlab.ee.mu.oz.au/~darryl/secondorder_code.html`. This implementation, uses Daubechies wavelets with various zero moments, it involves *bias correction*, options for *initialization of Mallat's algorithm* and automatic choice of the scales j_1 and j_2, as discussed in the references, above.

4 Benchmark wavelet spectra

In this section, we explore the wavelet spectrum tool by using several simulated examples. Our goal is to understand its practical limitations and to provide a guideline for its use in the estimation of the Hurst parameter and in explaining characteristic features in data.

4.1 The effect of deterministic trends

We start by briefly illustrating the advantages of the wavelet spectrum for the analysis of long-range dependent data.

- *Fractional Gaussian Noise*

Consider first an ideal, benchmark situation when the data $Y(k) = B_H(k+1) - B_H(k),\ k = 1, \ldots, N$ is a sample of fractional Gaussian noise with self-similarity (Hurst) parameter H. In this case the wavelet estimator performs very well and is essentially as good as one of the best available quasi-maximum likelihood estimators – the local Whittle estimator.

In the top-right plot of Figure 7, we display the *log-scale wavelet spectrum* of one path of FGN. This plot shows the statistics $\widetilde{S}_j(Y)$, defined in (3.15) above, as a function of the scale j, where $d_{j,k}(Y),\ k = 1, \ldots, N_j,\ j = 1, \ldots, J$ are all available wavelet coefficients, computed from the data $Y(k),\ k = 1, \ldots, N$.

The vertical segments on the wavelet spectrum plot indicate the variability of the statistics $\widetilde{S}_j(Y)$ above. The length of these intervals grows as j increases since there are fewer wavelet coefficients $N_j \approx N/2^j$ at the large scales j.

Observe that the wavelet spectrum is approximately linear. The slope of this line corresponds to $2H - 1$ which yields an estimate of H (see (3.11) and (3.12)). Usually, due to the problem of

initialization or due to the presence of non-trivial short-range dependence, one needs to ignore the first few smallest scales. On the bottom-left plot of Figure 7, we give estimates of H obtained by using different starting scales j_1 and all available larger scales. Note that, for all choices of j_1, but $j_1 = 1$ and $j_1 = 2$ the sample confidence intervals cover the true value of H.

The bottom-right plot in Figure 7 displays confidence intervals for the local Whittle estimator with different values of the parameter m, which controls the range of frequencies in the periodogram of the data that are used in the estimation. Roughly, the parameters j_1 and m can be related by $j_1 = \log_2(N/m)$. That is, large scales j correspond to low values for the frequency cut-off parameter m and vice versa. Therefore, to be able to compare the wavelet and local Whittle estimators, we chose values of $m \approx N/2^{j_1}$, $j_1 = 1, 2, \ldots, 7$ so that small values of j_1 on the bottom-left plot correspond to large values of m on the bottom-right plot. For more details on the local Whittle estimator and its applications, see Robinson (1995), Taqqu and Teverovsky (1997) and, for example, Hernández-Campos, Le, Marron, Park, Park, Pipiras, Smith, Smith, Trovero and Zhu (2004).

As seen in Figure 7, the wavelet estimator is essentially as good as the local Whittle estimator. Since the local Whittle estimator is an approximate quasi-maximum likelihood estimator for this model it is essentially one of the best possible estimators. Observe that the sample standard deviations of the wavelet estimator are always slightly larger than the corresponding ones for the local Whittle estimator.

Figure 7 around here

• *FGN plus a smooth trend*

One major advantage of the wavelet estimator for the Hurst parameter is that it is immune to smooth polynomial trends in the data. Indeed, if

$$\widetilde{Y}(t) = Y(t) + P_n(t),\ k = 1, \ldots, N,$$

where $P_n(t) = a_0 t^n + \cdots + a_{n-1}t + a_n,\ t \in \mathbb{R}$ is a polynomial of degree $n \in \mathbb{N}$. When $n < M$, where M is the number of zero moments of the wavelet ψ, Relation (3.1), implies that

$$d_{j,k}(\widetilde{Y}) = \int_{\mathbb{R}} (Y(t) + P_n(t))\psi_{j,k}(t)dt = \int_{\mathbb{R}} Y(t)\psi_{j,k}(t)dt = d_{j,k}(Y).$$

Hence, theoretically, the estimators of H, based on the wavelet coefficients of the perturbed process $\widetilde{Y}$ will be identical to those based on the process Y. This, as shown in Figure 8, yields practical benefits.

Observe that the local Whittle estimator is greatly affected by the trend and is essentially unusable, whereas the wavelet estimator is unaffected by the trend.

Figure 8 around here

The relative robustness of the wavelet estimator with respect to smooth trends and the clear visual interpretation of the wavelet spectrum are very important in practice. These features make wavelets a preferable tool in the analysis of long-range dependent network traffic data, which often exhibit non-trivial trends.

• *FGN plus high-frequency oscillating trends*

Although the wavelet estimator is robust with respect to a large class of smooth low-frequency trends, it can be quite sensitive to high-frequency deterministic oscillations. Consider for example

$$\widetilde{Y}(k) = G_H(k) + h_\nu(k),\ k = 1, \ldots, N, \quad \text{where} \quad h_\nu(t) = \sin(2\pi\nu t/N),\ \nu > 0, \tag{4.1}$$

and where $G_H(k) = B_H(k+1) - B_H(k),\ k = 1, \ldots, N$ is a FGN time series. Here ν corresponds to the number of oscillations of h_ν in the interval $[0, N]$. If $\nu << M$, where M is the number of zero moments of ψ, then the function $h_\nu(k)$ can be essentially interpolated by a polynomial of degree $n < M$, and hence the wavelet estimator of H remains unaffected.

However, as shown in Figure 9, if $\nu > M$, then the oscillations of h_ν start to affect significantly the wavelet spectrum of $\widetilde{Y}$. This perturbation results in a *bump* of the spectrum at scales j of the order of $\log_2(N/\nu)$. Indeed, consider, for example, the wavelet coefficient $d_{j,0}(h_\nu)$ of the function h_ν. In view of (3.3), by making a change of variables, we get

$$d_{j,0}(h_\nu) = 2^{j/2} \int_{\mathbb{R}} \sin(2\pi\nu 2^j \tau/N)\psi(\tau)d\tau = \sqrt{2\pi}\, 2^{j/2} \mathcal{I}m\Big(\widehat{\psi}(2\pi\nu 2^j/N)\Big), \tag{4.2}$$

where $\widehat{\psi}(\xi) := (2\pi)^{-1/2} \int_{\mathbb{R}} e^{i\xi t}\psi(t)dt$ denotes the Fourier transform of ψ and where $\mathcal{I}m(z)$ denotes the imaginary part of the complex number z.

Since the mother wavelet ψ is well-localized in the frequency domain and since $\widehat{\psi}(0) = (2\pi)^{-1/2} \int_{\mathbb{R}} \psi(t)dt = 0$, the right-hand side of (4.2) vanishes rapidly as $\nu 2^j/N$, approaches ∞ or 0. Thus $|d_{j,0}(h_\nu)|$ is non-negligible when

$$\frac{\nu 2^j}{N} \approx 1 \quad \text{or, equivalently,} \quad j \approx \log_2\Big(\frac{N}{\nu}\Big).$$

Figure 9 around here

Figure 10 displays the time series of wavelet coefficients of $\widetilde{Y}(k)$, of the deterministic function $h_\nu(k),\ \nu = 100$ and of the original FGN time series $G_H(k)$. Observe, that the wavelet coefficients of the function h_ν dominate those of the FGN on scales 10–11. These are the scales where the "bump" appears in the top-right plot in Figure 9.

Figure 10 around here

• *FGN plus breaks*

The wavelet spectrum of a time series $Y(k)$ is influenced in a different, and in fact more subtle, to way by *breaks* or *shifts in the mean*. Let

$$Y(k) = G_H(k) + h(k),\ k = 1, \ldots, N,$$

where $G_H(k) = B_H(k+1) - B_H(k)$ is a FGN time series and where the function $h(t)$ is a linear combination of indicator functions. In Figure 11 below, we show what happens for one specific choice of the function h.

Unlike the case of a high-frequency smooth perturbation (see Figures 9 and 10), in the case of breaks, the spectrum of the deterministic perturbation is essentially linear (on the scales, we observe). This results in a subtle change of the scaling exponent of the data at large scales j rather than in a bump of the spectrum of the data. The location of the "knee" can be controlled by changing the magnitude of the function h.

Observe that, visually, based on the top-left plot in Figure 11, it may be difficult to determine that we have a FGN time series plus a simple deterministic trend. Thus, in practice, one may be misled to model the data with a long-range dependent time series with large Hurst exponent $H \approx 0.99$ rather than to account for the breaks. Such situations are often encountered when dealing with network traffic (see Section 5, below).

Figure 12 displays the wavelet coefficients of the data, the deterministic function and the original FGN time series on several scales. It gives some insight into the differences in the wavelet spectra on the top-right plot of Figure 11.

Figure 11 around here

Figure 12 around here

4.2 The effect of stochastic models

Here, we investigate the wavelet spectra of stochastic models, which are essentially different from the classical fractional Gaussian noise model. We do so in order to understand the possible causes of the features observed in the wavelet spectra of network traffic, in Section 2.2.

• *Non-homogeneous Poisson processes*

Let $Y(k),\ k = 1, \ldots, N$ be a time series of independent Poisson random variables with parameters $\Lambda(k) = \mathbb{E}Y(k) > 0$, driven by a deterministic function $\Lambda(t),\ t \in \mathbb{R}$. The process $\{Y(k)\}$ can be viewed as a discretized (binned) version of a non-homogeneous Poisson point process N with a step-wise intensity function $\Lambda(t) = \Lambda(k),\ t \in [k, k+1)$. That is, $Y(k)$ equals the number of arrivals of the point process N in the interval $[k, k+1)$.

In Figure 13, we display the wavelet spectrum of the process $Y(k),\ k = 1, \ldots, N$, where the function Λ is "periodic":

$$\Lambda_\nu(k) := b \times \Big(\sin(2\pi\nu k/N) + a \Big), \tag{4.3}$$

with $a = 10$, $b = 2$. We let $\nu = 10,\ 100$ and 1000.

The wavelet spectrum of a discretized homogeneous Poisson process is essentially flat. Indeed, the Poisson process has finite variance, stationary and independent increments therefore it is asymptotically self-similar with Hurst parameter $H = 1/2$. On large time scales it can be viewed as an approximation to the classical Brownian motion plus drift.

The non-homogeneous Poisson process with bounded intensity should also have approximately flat wavelet spectrum at large scales. In practice, however, due to limited sample size one may observe significant deviations from a constant wavelet spectrum at large scales (see the top-right plot in Figure 13, below).

Figure 13, shows how the frequency ν of the periodic intensity function Λ_ν affects the wavelet spectrum of the Poisson process. As in Relation (4.2), above, one can argue that this effect, is limited to a range of scales about $j \approx \log_2(N/\nu)$ (see Figure 9).

Figure 13 around here

• *Cox processes*

Now let $N = \{\epsilon_j\}_{j\in\mathbb{Z}}$ be a homogeneous Poisson point process on $\mathbb{R}$, with intensity λ. Consider the following *kernel smoothed* version of this point process:

$$\Lambda_{\lambda,h}(t) := \mu \sum_{j\in\mathbb{Z}} \Phi_h(t-\epsilon_j), \quad t \in \mathbb{R},$$

where $\Phi_h(\cdot) = h^{-1}\Phi(\cdot/h)$ and where $\Phi(x) = (2\pi)^{-1/2}\exp\{-x^2/2\},\ x \in \mathbb{R}$ denotes the standard Gaussian kernel.

The random process $\Lambda_{\lambda,h} = \{\Lambda_{\lambda,h}(t)\}_{t\in\mathbb{R}}$ can be viewed as a (two-sided) *shot-noise* process. Given the function $\Lambda_{\lambda,h}(t)$, consider a non-homogeneous Poisson point process $Q = \{Q_j\}_{j\in\mathbb{Z}}$ with intensity $\Lambda_{\lambda,h}(t) > 0$. The point process Q is a particular case of a Cox process or doubly stochastic Poisson point process.

As before, let $Y(k),\ k \in \mathbb{Z}$ be the discretized or binned version of the point process Q. Thus, the random variables $Y(k),\ k \in \mathbb{Z}$ are conditionally independent and Poisson with parameters $\int_k^{k+1} \Lambda_{\lambda,h}(t)dt,\ k \in \mathbb{Z}$.

Figure 14 shows one simulated path of the process $Y(k),\ k = 1, \ldots, N$, where $h = 0.1,\ \lambda = 1000,\ \mu = 0.01$ and $N = 50\,000$. The wavelet spectrum of this path and the wavelet spectrum of the underlying intensity $\Lambda_{\lambda,h}$ are given on the bottom-right plot.

Observe that, although, theoretically, this Cox process is short-range dependent, the wavelet spectrum in Figure 14 suggests that this is not the case. The doubly stochastic nature of this process is responsible for non-trivial short-range dependence in the form of apparent non-stationarity. This, in turn, can mislead the wavelet spectrum and, in fact, any other methodology (that we are aware of) for the analysis of long-range dependence.

Figure 14 around here

5 Examples revisited

We now comment briefly on the examples of wavelet spectra of network traffic traces shown in Figures 3–6, using insights developed in Section 4.

Firstly, the spike at scale $j = 11$ in the wavelet spectrum in Figure 3 may be due to a high-frequency deterministic component in the network traffic trace (see Figure 9). This is also confirmed by the periodogram of the traffic trace shown in Figure 15. This feature of the Internet traffic trace in Figure 3 is further analyzed by using scale-space smoothing tools such as *SiZer* and *SiNos* in Park, Godtliebsen, Taqqu, Stoev and Marron (2004). That analysis reveals an extremely strong periodic component (at the indicated frequency) that is localized within the peak visible at time 3269–3636 (sec), perhaps caused by an intense IP port scan.

Secondly, although the Abilene trace shown in Figure 6 has an essentially linear wavelet spectrum at large scales $5 \le j \le 16$, the corresponding slope is quite steep. In fact, the estimated Hurst parameter is greater than the theoretically admissible value of 1. This behavior is consistent with non-stationarity in the form of *breaks* or abrupt changes in the mean and

is analogous to the situation studied in Figures 11 and 12. Our experience with a number of other traces indicates that the Hurst parameter is typically overestimated when the traffic traces possess such non-stationary features.

The wavelet spectra of the traces shown in Figures 4 and 5 are quite interesting. The high variability of the wavelet spectrum in Figure 4 indicates that traffic is strongly dependent, but that it also lies outside the scope of the traditional models for long-range dependence such as the fractional Gaussian noise. The bumps in the spectrum at large scales cannot be reliably connected to any of the situations studied in Section 4. The following two approaches can be used in addition to the wavelet spectrum to understand better the structure network traffic.

• The statistics $S_j,\ j = 1, \ldots, J$ of the wavelet spectrum involve averaging over time of the wavelet coefficients of the data (see (3.12)). Consequently, the wavelet spectrum is essentially blind to subtle non-stationarity effects of the underlying time series. The tools introduced in Park, Godtliebsen, Taqqu, Stoev and Marron (2004) provide a way of addressing this problem by detecting hidden non-stationarity. This is done by applying scale-space smoothing methods such as SiZer and SiNos to the time series of the wavelet coefficients of the traffic trace. For example, as illustrated in that paper, by using this tool to analyze the trace in Figure 4, one notices significant changes in the network traffic.

Indeed, the lower-right plot in Figure 4 shows that the wavelet spectrum has a "bump" at scale $j = 14$. This suggests that there is unusual scaling behavior corresponding to this particular scale, that is, $2^{14} = 16\,384$ (ms) ≈ 16 (sec). From this plot, it is not easy to determine which part of the trace causes the "bump". However, the lower left plot, which displays the aggregate number of packets in 1 second intervals, shows a sharp and deep drop in the packet arrival rate at time about 2 500 seconds. Park, Godtliebsen, Taqqu, Stoev and Marron (2004) used the scale-space smoothing methods to determine that this "bump" is due to the sharp drop in the lower-left plot. In fact, this drop turns out to be an 8 second long network outage.

• In Stoev, Taqqu, Park, Michailidis and Marron (2004), we propose a tool for analyzing the local dependence structure of network traffic traces. We develop statistical and visualization techniques for the *local wavelet spectrum* and the corresponding *local Hurst parameter* of the data. Such analysis can be very useful when dealing with large Internet traffic data sets. It bypasses many problems due to non-stationarity and provides a more detailed picture of the local dependence structure of the traffic. We also propose and implement a robustified version of the wavelet spectrum which involves taking sample medians rather than sample averages in (3.12). This yields more reliable estimates of both the local and global Hurst parameters.

Finally, the wavelet spectrum displayed in Figure 5 is strikingly different from most of the wavelet spectra of other traffic traces we have encountered. Note the steep slope of the spectrum at scales $5 \leq j \leq 7$. A detailed analysis of this traffic trace based on IP packet level information which separates different protocols and communication ports is presented in Park, Rolls, Hernandez-Campos, Smith and Marron (2004). This strong feature turns out to be driven by a single UDP (User Datagram Protocol) based application, with a strongly quasi-periodic component. This confirms the fact that the wavelet spectrum can incorporate subtle details of network behavior.

6 Conclusion

The wavelet spectrum is a useful tool for analyzing the dependence structure of network traffic. It can be used to obtain relatively reliable estimators of the critical Hurst parameter for stationary traffic traces. The use of the wavelet spectrum must not be limited to the estimation of the Hurst parameter. Very often non-stationarity and important periodic components of the traffic (perhaps localized in time) can be detected by using the small and medium scales of the spectrum. On the other hand, severe bursts and non-stationarity in the form of mean shifts and, in fact, lower frequency oscillations affect the wavelet spectrum in a subtle way. In such cases, the analyst may be mislead and overestimate the Hurst long-range dependence parameter, in particular when the wavelet estimator is used blindly without inspecting the whole spectrum.

We concluded Section 5 by referring to two new, complementary tools for a deeper analysis of the dependence in network traffic. The first tool is used to detect significant deviations from stationarity by using graphical scale-space smoothing techniques applied to the wavelet coefficients of the traffic. It provides useful time and frequency information about the nature of the non-stationary effects. The second tool can be used to obtain more reliable estimates for the Hurst parameter. It focuses on estimation of the local the Hurst parameter and also includes a robust median-based version of the wavelet spectrum. More details on these two approaches can be found in Park, Godtliebsen, Taqqu, Stoev and Marron (2004) and Stoev, Taqqu, Park, Michailidis and Marron (2004), respectively.

Figure 15 around here

Acknowledgments

All datasets we present in this paper were extracted from logs of IP packets by members of the DIRT group at UNC, Department of Computer Science (`http://www.cs.unc.edu/Research/dirt/`). We are particularly indebted to Felix Hernandez-Campos. We also thank Vladas Pipiras for fruitful discussions. The Matlab code used to compute the local Whittle estimator was kindly given to us by Patrice Abry. This research was done while Murad S. Taqqu and Stilian Stoev were visiting the University of North Carolina at Chapel Hill and the Statistical and Applied Mathematical Sciences Institute (SAMSI) at the Research Triangle Park. They would like to thank them for providing an excellent environment and inspiring working atmosphere. This research was also partially supported by the NSF Grant ANI–9805623 at Boston University.

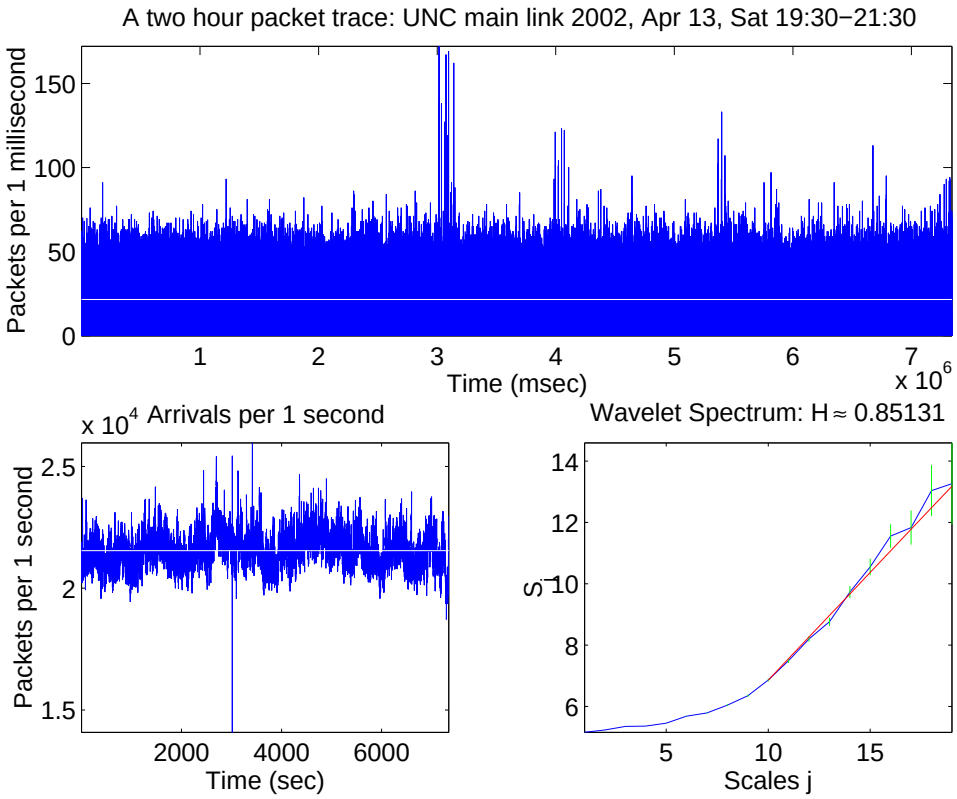

Figure 1: The top plot shows the time series of the number of packets arriving on a link in 1 millisecond time intervals. Observe that this time scale is rather fine and hence the distribution of the time series appears to be skewed and non-Gaussian. The mean of the time series is over-plotted in white. The lower-left plot displays the aggregate number of packets (for the same trace) in 1 second intervals. In white, we show the mean of this time series. On the lower-right plot we show the log-scale wavelet spectrum of the time series displayed in the top plot. Observe that the wavelet spectrum is essentially linear at large scales ($10 \leq j$). The Hurst parameter, estimated by fitting a line to the wavelet spectrum on scales $j_1 = 10, 11, \ldots, 20$ is $\widehat{H} \approx 0.8513$ (for more details, see Section 3 below). This trace is effectively modeled by standard FGN.

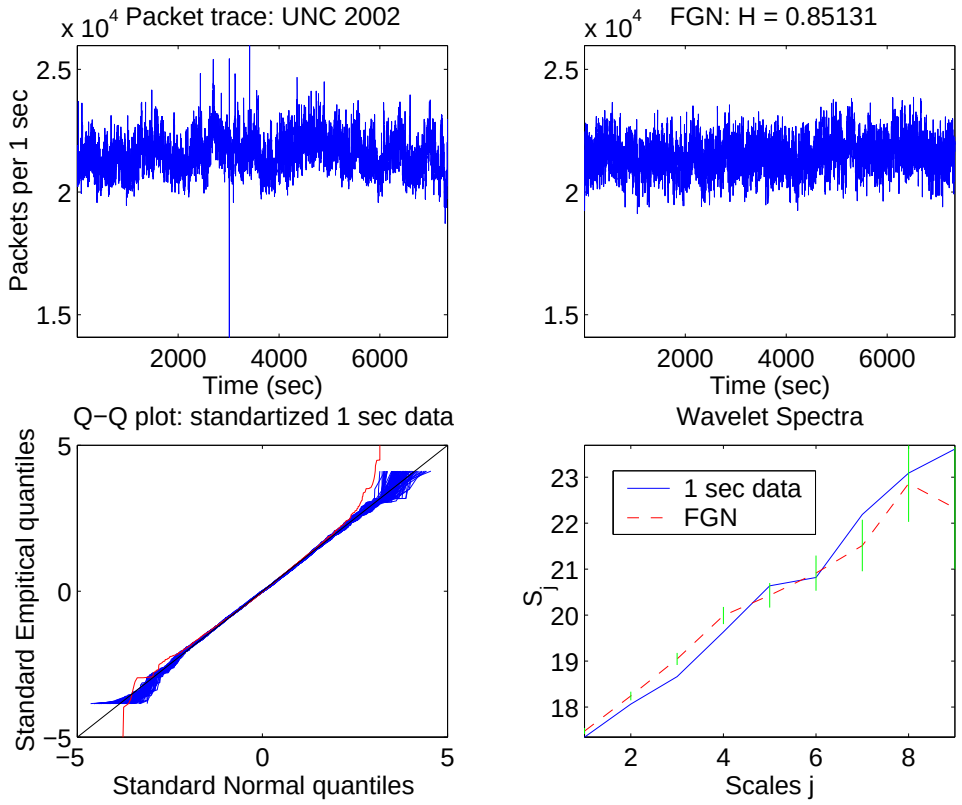

Figure 2: The top-left plot displays the time series of packet arrivals per 1 second obtained from the two-hour trace shown in Figure 1. The plot on the right displays a time series of simulated fractional Gaussian noise with sample mean, sample variance and Hurst parameter equal to the estimated mean, variance and Hurst parameter from the packet trace in Figure 1. The bottom-left plot shows the empirical quantiles of the standardized packet trace above versus the quantiles of standard normal distribution (in red). To indicate the sampling variability, we also add (in blue) 100 independent QQ-plots based on samples from the standard normal distribution. The bottom-right plot displays the wavelet spectrum of the packet trace in the top left and of the "simulated" FGN time series in the top right. Observe that the two spectra are very similar. (The vertical segments on the plot indicate 95% confidence intervals for the statistics S_j corresponding to FGN.)

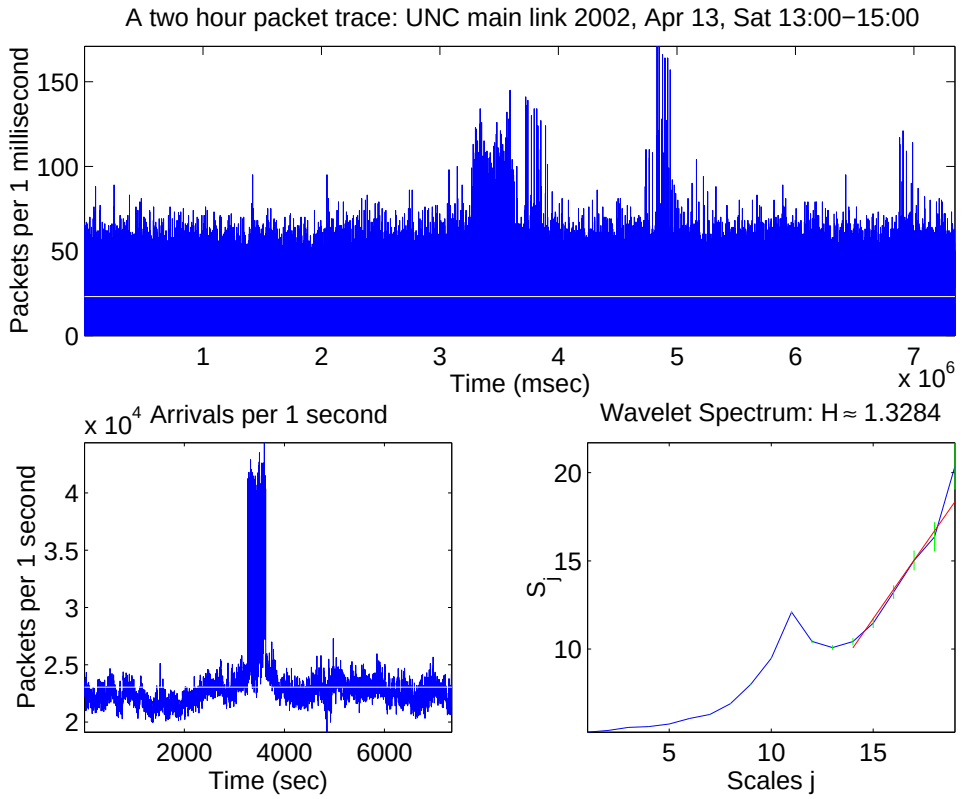

Figure 3: The top plot shows the time series of the number of packets arriving on a link in 1 millisecond time intervals. The mean of the time series is over-plotted in white. The lower-left plot displays the aggregate number of packets (for the same trace) in 1 second intervals. In white, we show the mean of this time series. On the lower-right plot we show the log-scale wavelet spectrum of the time series given in the top plot. Observe the sharp spike in the wavelet spectrum around scales $j = 11$ and 12. The spectrum appears linear on large scales $14 \leq j \leq 20$. The estimated Hurst parameter over this range of large scales ($j_1 = 14, 15, \ldots, 20$) is about $\widehat{H} \approx 1.328$. The spike is uncharacteristic of FGN data, and is explained in Section 5.

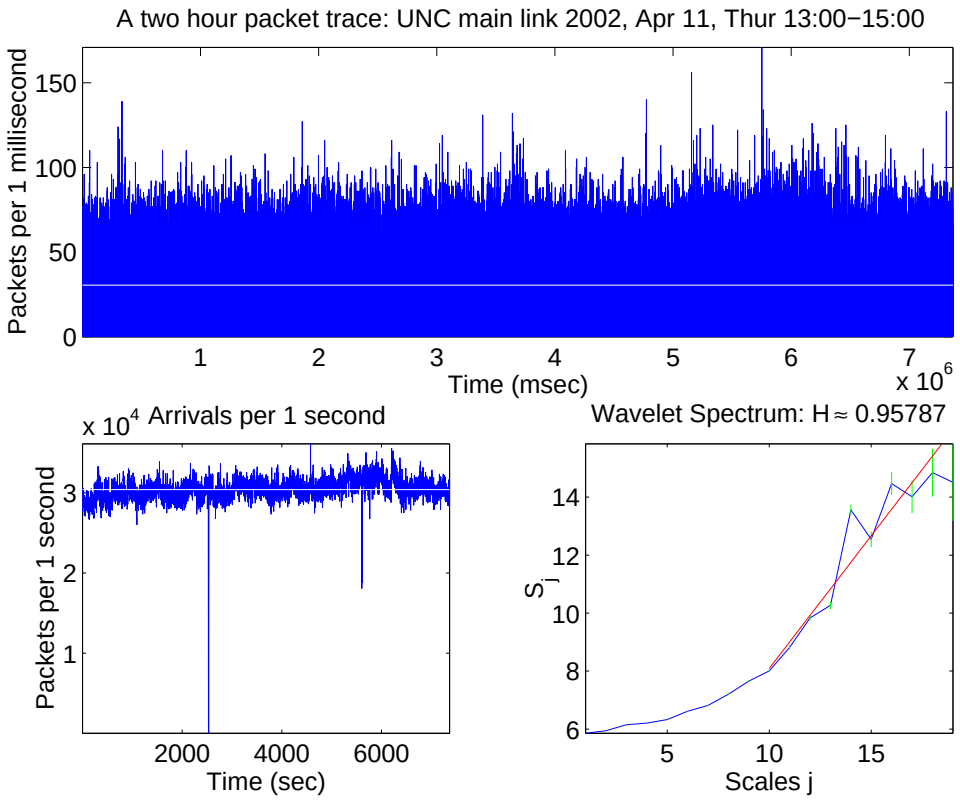

Figure 4: The top plot shows the time series of the number of packets arriving on a link in 1 millisecond time intervals. The mean of the time series is over-plotted in white. The lower-left plot displays the aggregate number of packets (for the same trace) in 1 second intervals. In white, we show the mean of this time series. On the lower-right plot we show the log-scale wavelet spectrum of the time series given in the top plot. Observe that the spectrum exhibits high variability on large scales. The estimated Hurst parameter over the range of scales $10 \leq j \leq 20$ is about $\widehat{H} \approx 0.958$. This variability in the spectrum is explained in Section 5.

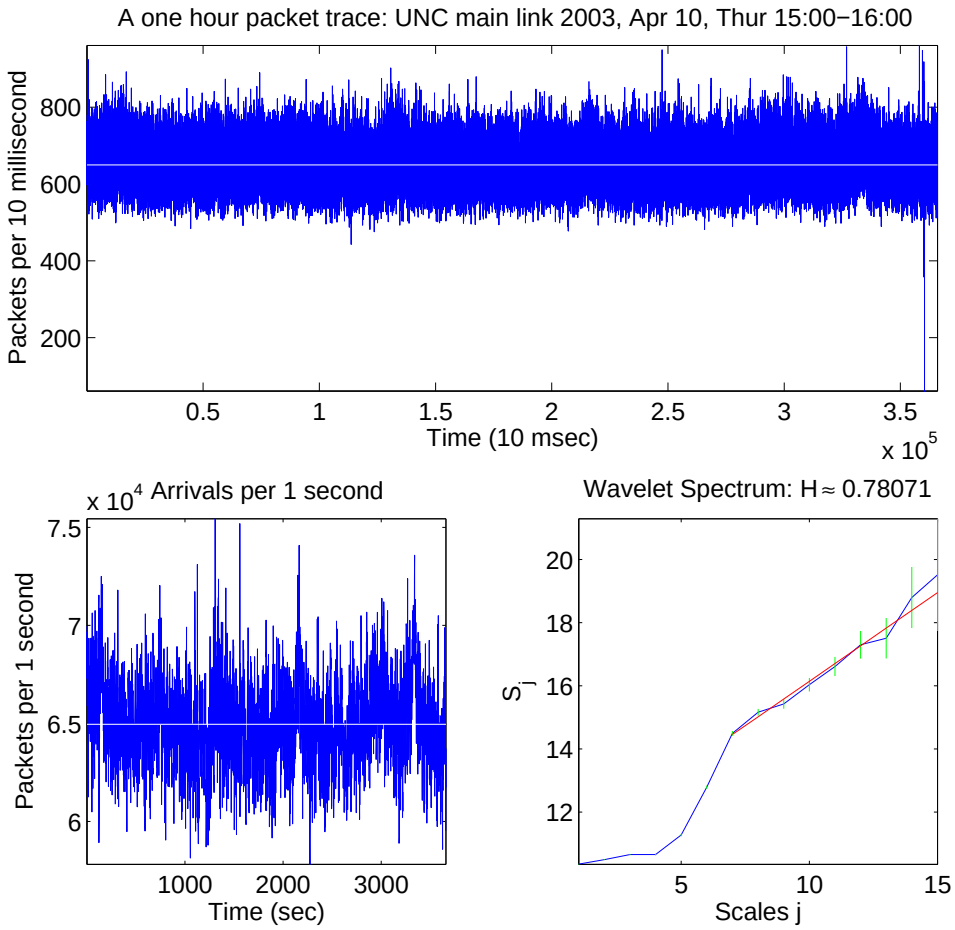

Figure 5: The top plot shows the time series of the number of packets arriving on a link in 10 millisecond time intervals. Observe that this time scale is coarser than the time scale of the top plots in Figures 1 and 3. The mean of the time series is over-plotted in white. The lower-left plot displays the aggregate number of packets (for the same trace) in 1 second intervals. In white, we show the mean of this time series. On the lower-right plot we show the log-scale wavelet spectrum of the time series given in the top plot. Note that the wavelet spectrum appears to be linear on large scales: $7 \leq j \leq 15$. The estimated Hurst parameter over this range of scales is $\widehat{H} \approx 0.781$. Observe, however, the striking difference between the shape of the wavelet spectrum in this case and in the cases shown in Figures 1 and 3, above, and Figure 6, below. This unusual wavelet spectrum shape is explained in Section 5 below.

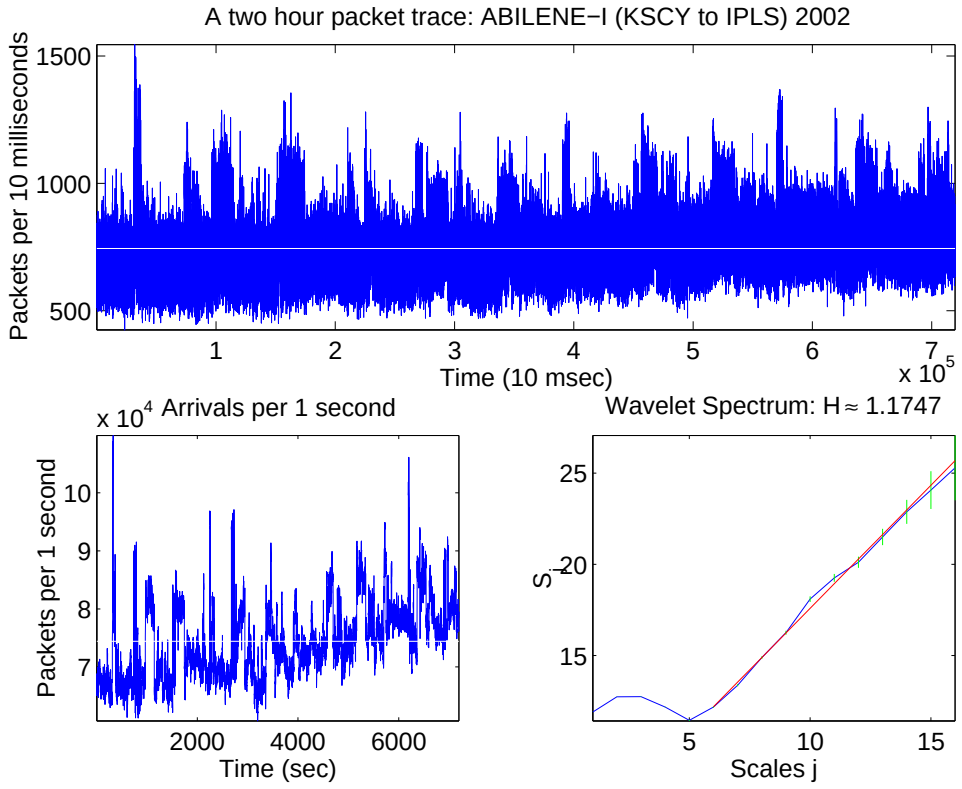

Figure 6: The top plot shows the time series of the number of packets arriving on a link in 10 millisecond time intervals. The data was collected on August 14, 2002 between 9:00 am and 11:00 am from the Internet-2 (Abilene) backbone link connecting Kansas City to Indianapolis. (for more details, see `http://pma.nlanr.net/Traces/long/ipls1.html`). Observe that this time scale is coarser than the time scale of the top plots in Figures 1 and 3. The mean of the time series is over-plotted in white. The lower-left plot displays the aggregate number of packets (for the same trace) in 1 second intervals and appears non-stationary. In white, we show the mean of this time series. On the lower-right plot we show the log-scale wavelet spectrum of the time series given in the top plot. Note that the wavelet spectrum appears to be essentially linear on a wide range of relatively large scales: $5 \leq j \leq 16$. The estimated Hurst parameter over this range of scales is $\widehat{H} \approx 1.044$. This wavelet spectrum shape will be explained in Section 5.

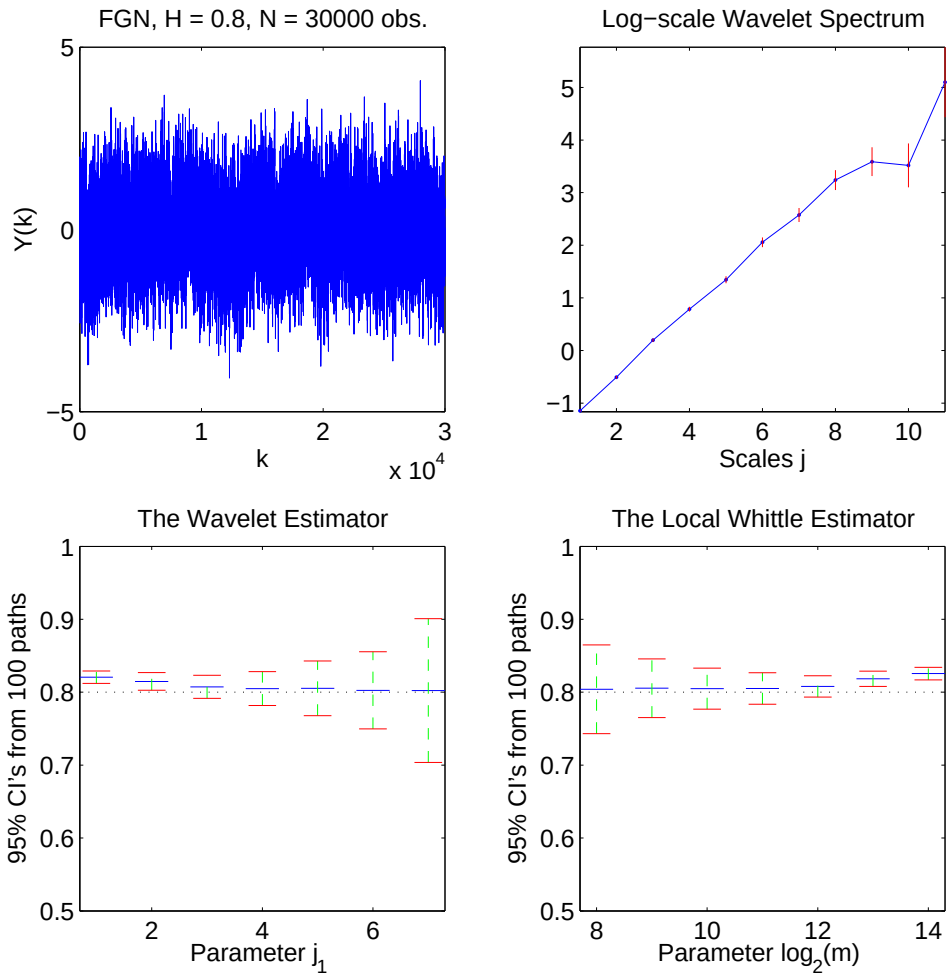

Figure 7: *Comparison between the wavelet and the local Whittle estimators for fractional Gaussian noise (FGN).* The top plot on the left displays one simulated path of FGN with Hurst parameter $H = 0.8$. The plot on the right shows the wavelet log-scale spectrum of this path. (We used the Daubechies wavelets with 3 zero moments.) The vertical segments on this plot are the estimated 95% confidence intervals of the log-mean-energy statistics of the wavelet spectrum. The bottom-left plot displays wavelet estimates of the Hurst parameter H obtained from 100 independently simulated paths of FGN by using different choices of the parameter j_1. The variability of these estimates is summarized through 95% sample confidence intervals, based on a normal approximation. There is a small bias at small j_1 (no initialization was done) and the variance increases as j_1 increases. The bottom plot on the right contains estimates of H obtained via the local Whittle estimator. They were computed over the same set of 100 independent paths as the wavelet estimators, with different choices of the frequency cut-off parameter m. As for the wavelet estimators, we display 95% sample confidence intervals. Note the resemblence between the two plots, with $m \approx N/2^{j_1}$.

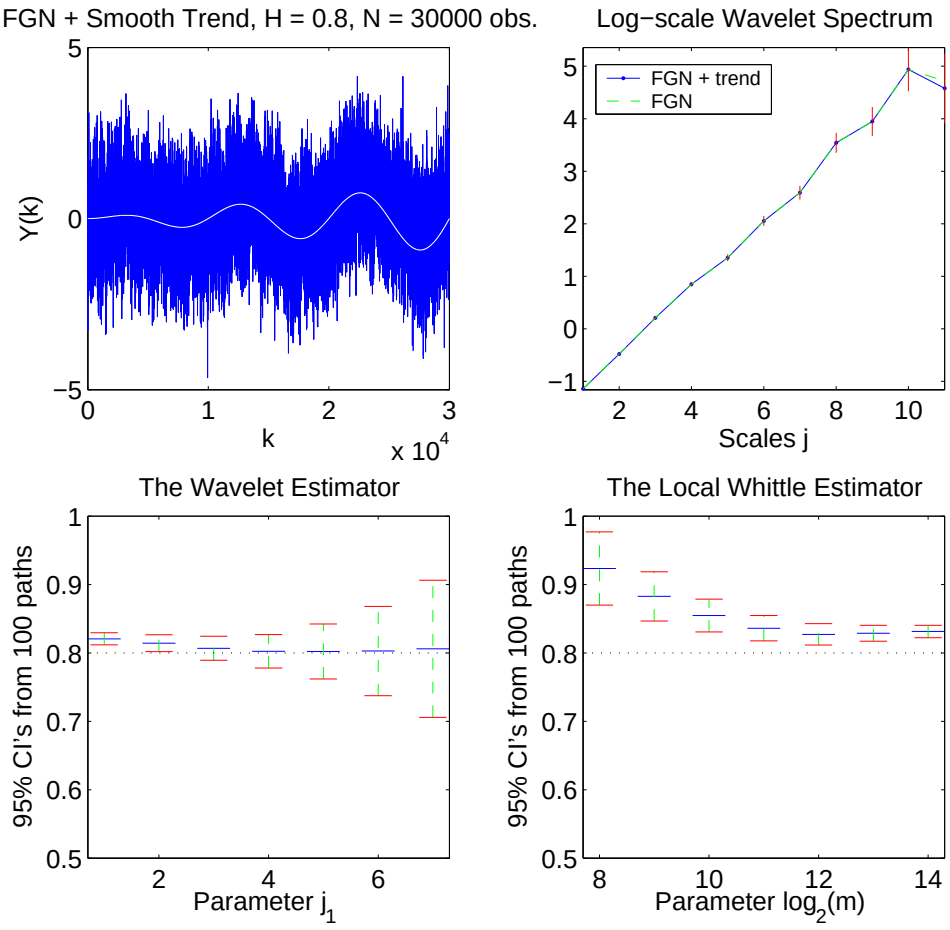

Figure 8: The top plot on the left displays one simulated path of fractional Gaussian noise (FGN) with $H = 0.8$ plus a "smooth" additive trend, over-plotted in white. The plot on the right shows the wavelet log-scale spectrum of this path and the spectrum of the corresponding path of "pure" FGN (see Figure 7). Observe that these two spectra essentially coincide. The bottom-left plot displays wavelet estimates of the Hurst parameter H obtained from 100 independently simulated paths of FGN plus the same additive trend. As in Figure 7, we show 95% sample confidence intervals for various choices of the scale j_1. The bottom plot on the right contains estimates of H obtained via the local Whittle estimator, computed over the same set of 100 independent paths as the wavelet estimators. As for the wavelet estimators, we display 95% sample confidence intervals. Note that, in contrast to the wavelet estimator, the local Whittle estimator is greatly affected by the trend.

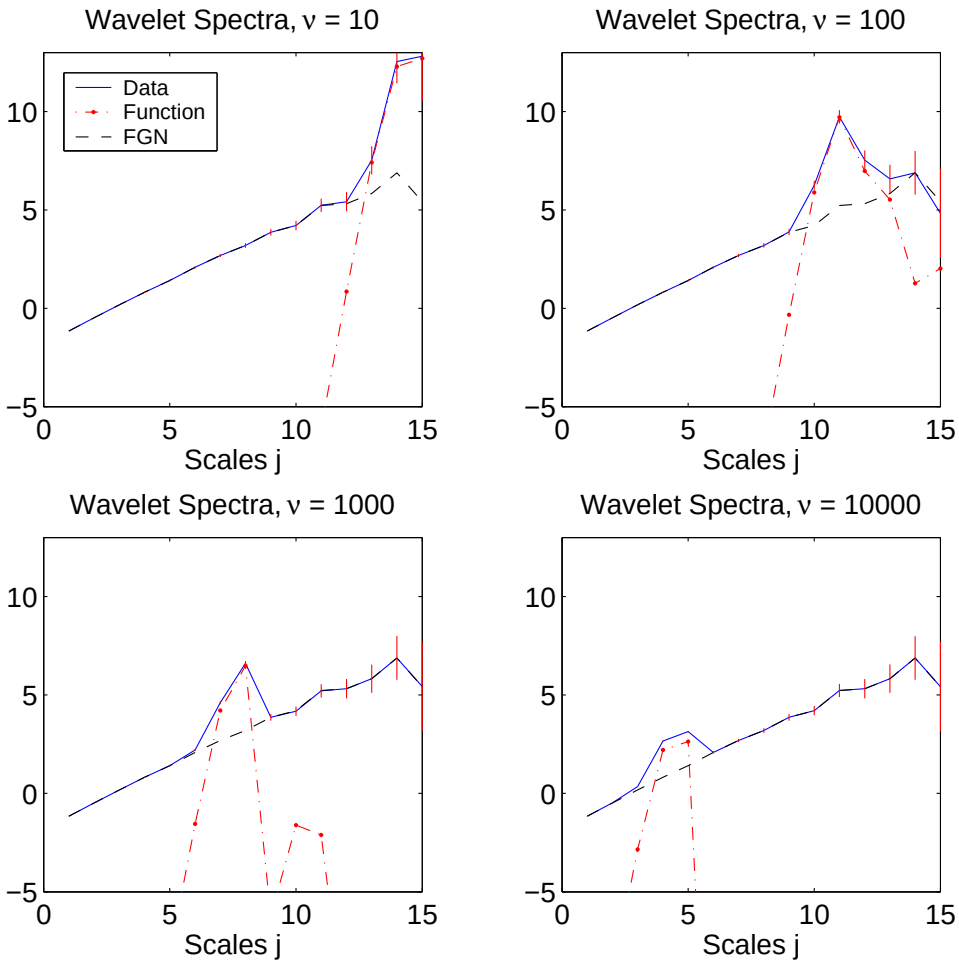

Figure 9: These four plots display wavelet spectra of a FGN, perturbed by adding a high-frequency trend as in (4.1). The length of the time series is $N = 300\,000$. We show what happens for four different values of the frequency $\nu = 10,\ 100,\ 1\,000$ and $10\,000$. Observe that the impact of this perturbation on the spectra is well-localized around scales $j \approx \log_2(N/\nu)$. The rest of the spectra essentially coincide with that of the FGN time series. If the number of scales j were significantly increased, we expect the top-left plot to look like the bottom-right plot. Here we used Daubechies wavelets with $M = 3$ zero moments.

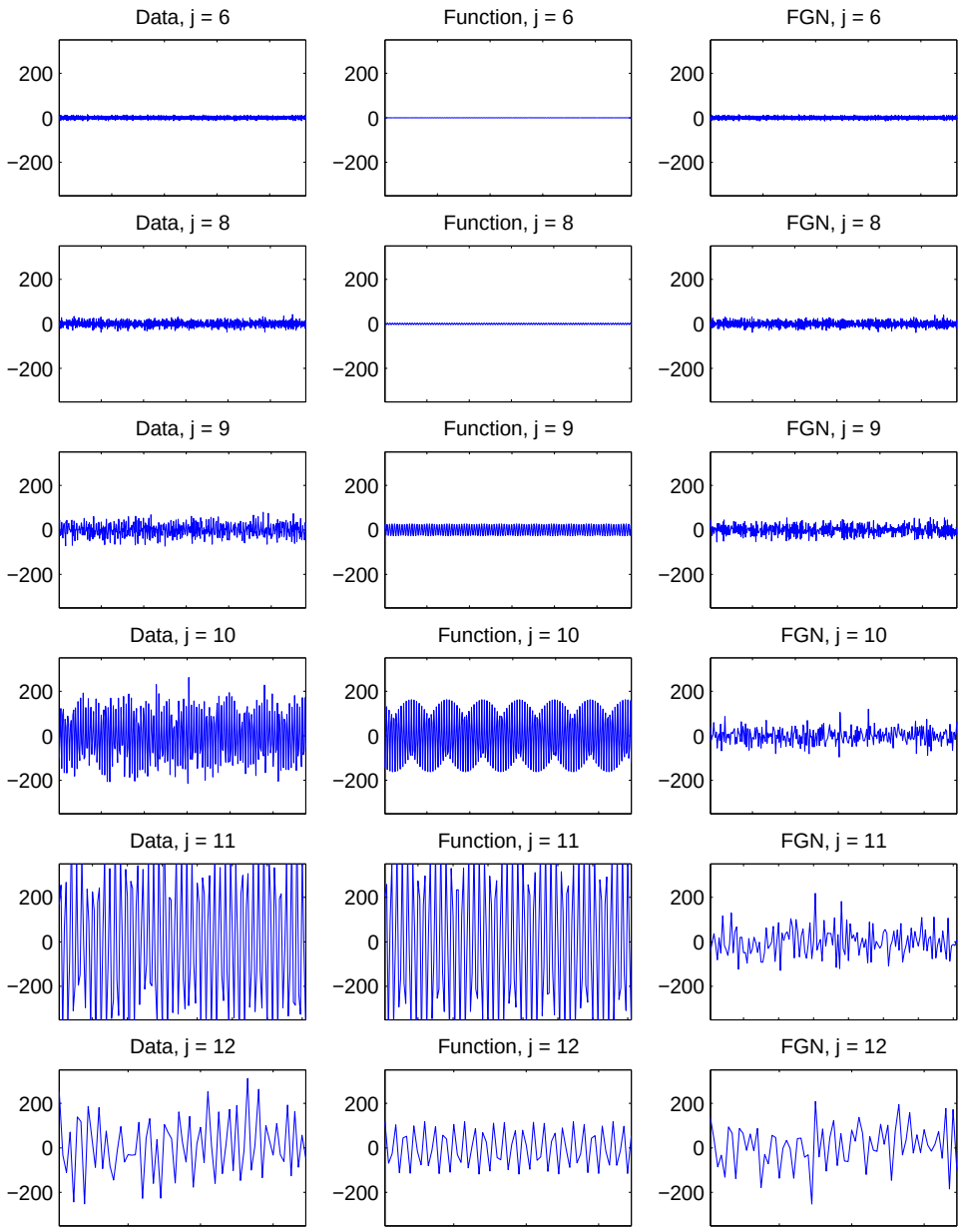

Figure 10: Let $\widetilde{Y}(k) = G_H(k) + h_\nu(k)$, where G_H is FGN and $h_\nu(k) = \sin(2\pi\nu k/N)$ with $\nu = 100$. The plots in the first column show the time series of wavelet coefficients (on several different scales j) of the time series $\widetilde{Y}(k)$. The second column of plots displays the wavelet coefficients of the function $h_\nu(k)$ and the third column - those of the FGN $G_H(k)$. The parameter ν equals 100. We used Daubechies wavelets with $M = 3$ zero moments. One clearly sees that the wavelet coefficients of the function h_ν dominate on scales 10 and 11, giving rise to the bump in the wavelet spectrum of Figure 9.

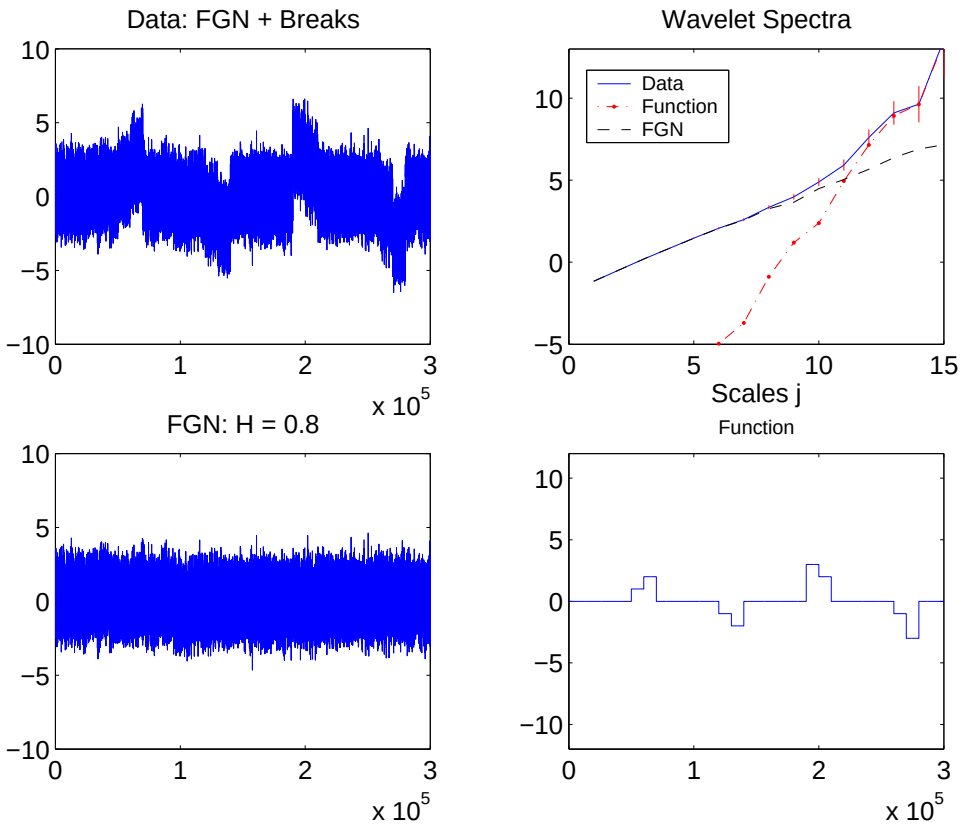

Figure 11: The top-left plot displays a FGN time series perturbed by adding deterministic breaks displayed in the bottom-right plot. The corresponding pure FGN time series is shown on the bottom-left plot. The plot on the top-right shows the wavelet spectra of the three time series: the *data* (perturbed FGN), the *function* (deterministic breaks) and the *FGN* (the original fractional Gaussian noise time series). Observe that the spectrum of the FGN is essentially linear, so is the spectrum of the function, however the slopes of the two lines are quite different. At large scales j, the spectrum of the function dominates that of the FGN and hence it determines the behavior of the spectrum of the data. Consequently, the wavelet estimator of the Hurst parameter is essentially misleading. A linear fit starting from scale $j_1 = 9$ and using all larger scales yields $\widehat{H}(data) = 1.1419$ whereas $\widehat{H}(FGN) = 0.8281$.

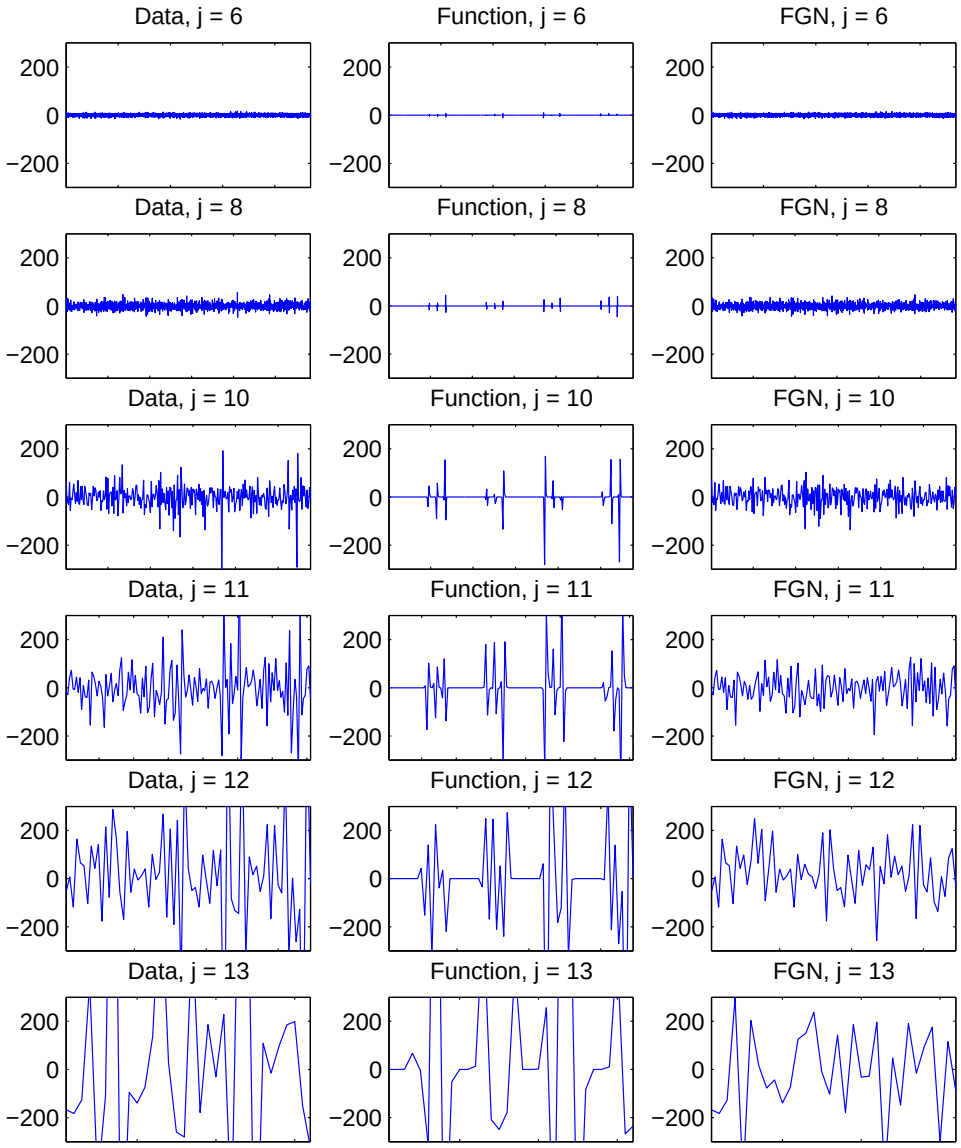

Figure 12: The first column of plots show time series of wavelet coefficients of the time series $Y(k)$, $k = 1, \ldots, N$, $N = 300\,000$, displayed in Figure 11. The second and third columns of plots contain the wavelet coefficients of the deterministic function and the FGN time series from Figure 11, respectively. Observe that at scales $j = 10$ to 13 the wavelet coefficients of the function become larger in magnitude than those of the FGN, and this affects the corresponding wavelet spectra (as shown in the top-right plot in Figure 11).

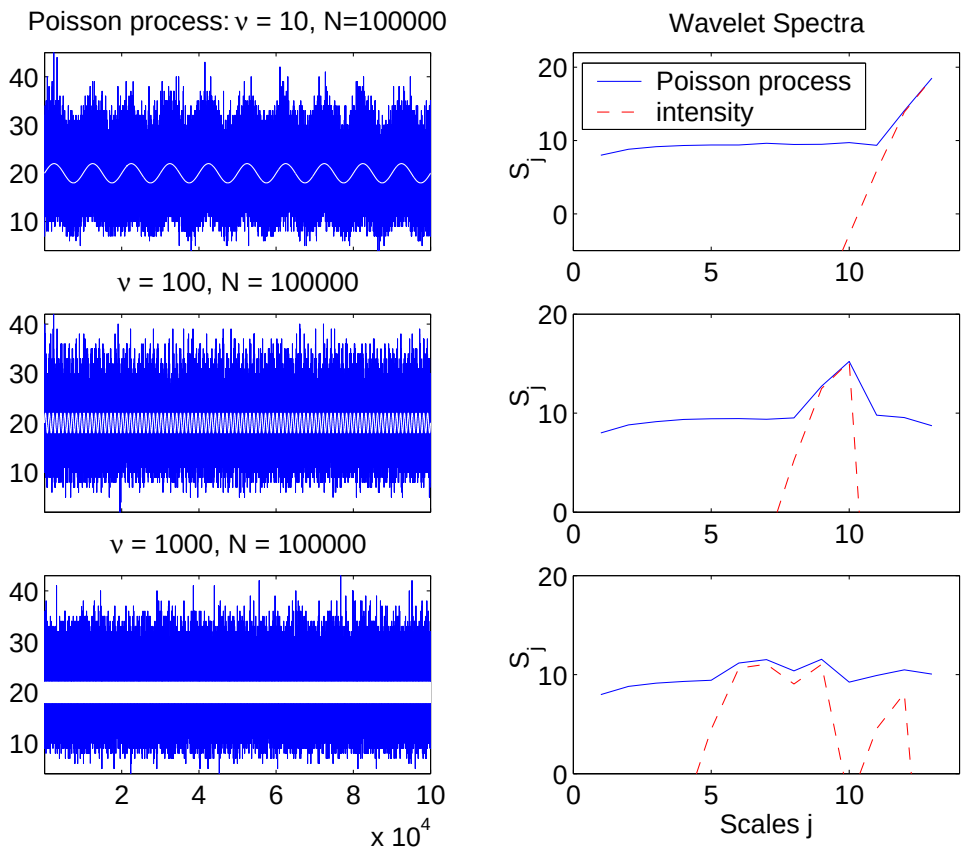

Figure 13: The first column of plots show realizations of *binned* paths of Poisson point processes with non-constant intensities Λ_ν, defined in (4.3), with $\nu = 10,\ 100,\ 1\,000$. The corresponding intensities are over-plotted in white (the white band in the bottom plot is due to the very rapid oscillations). The plots on the right show the wavelet spectrum of the corresponding binned path of the Poisson process and the wavelet spectrum of the corresponding intensity function Λ_ν. Note that the value of the frequency ν controls the location of the spike on the wavelet spectrum of the Poisson process as in Figures 9. The spike appears roughly around $j = 13,\ 10$ and 7, respectively.

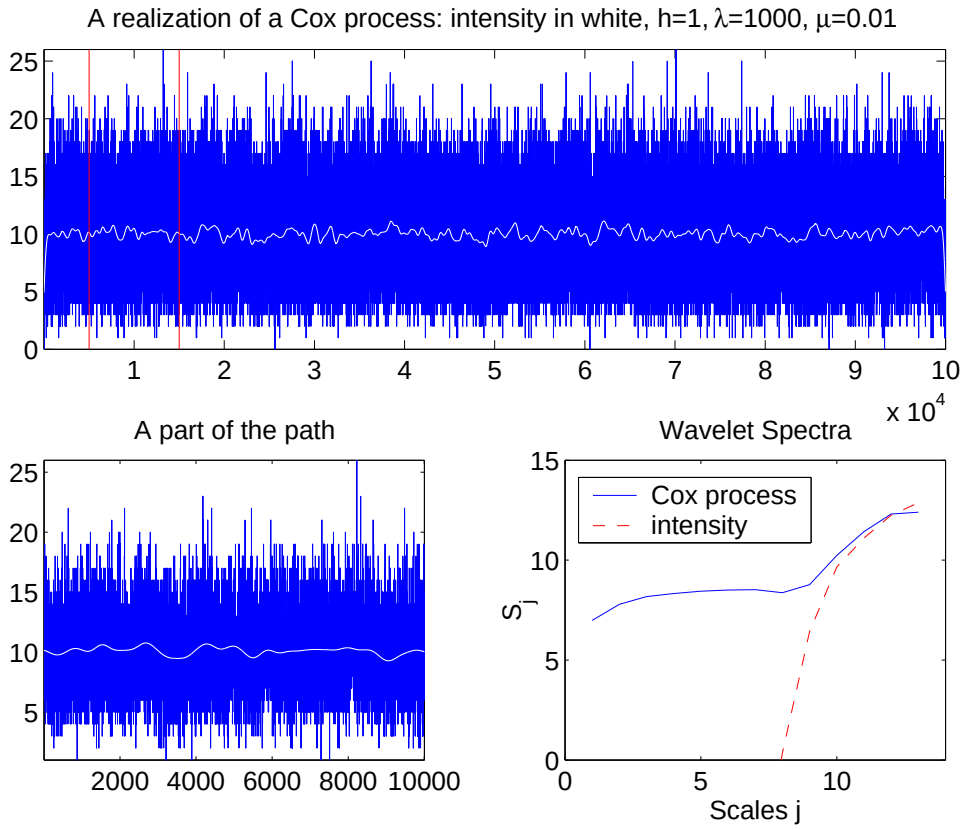

Figure 14: The top plot shows one realization if a *binned* path of a Cox process, which is a non-homogeneous Poisson process with random intensity. The underlying intensity is over-plotted in white. The lower-left plot displays a more detailed portion of the path above. The lower-right plot shows the wavelet spectrum of the path of the Cox process and the wavelet spectrum of its intensity. Observe that the two spectra essentially coincide on large scales. At small time scales, as expected, the wavelet spectrum of the Cox process is flat, which is consistent with the spectrum of a homogeneous Poisson process or a white noise process.

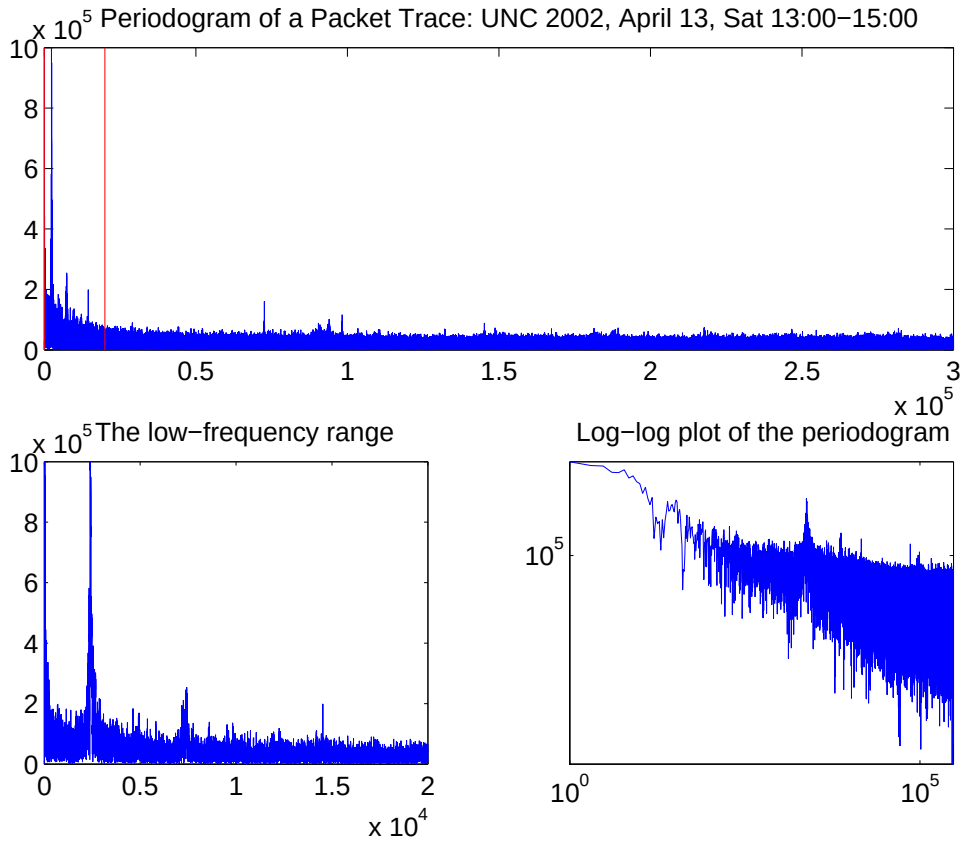

Figure 15: The top plot shows the periodogram of a two-hour packet trace. It was computed from the time series of packet arrivals per 10 millisecond time intervals. The bottom-left plot shows the low-frequency part of the periodogram (to the left of the vertical line in the top plot). The spike located at frequency $\nu \approx 0.2 \times 10^4 = 2000$ may be related to the bump in the wavelet spectrum of this traffic trace in Figure 3. Indeed, using $j \approx \log_2(N/\nu)$, we get $j \approx \log_2(300\,000/2000) \approx 7$. This corresponds to scale $j \approx 7 + \log_2(10) \approx 10$ in the wavelet spectrum displayed in Figure 3 which involves the time series of packet arrivals per 1 millisecond rather than per 10 milliseconds. The bottom-right plot indicates the $1/f$ or long-range dependence behavior of the time series of packet arrivals.

References

Abry, P. & Veitch, D. (1998), 'Wavelet analysis of long range dependent traffic', *IEEE Transactions on Information Theory* **44(1)**, 2–15.

Abry, P., Flandrin, P., Taqqu, M. S. & Veitch, D. (2003), Self-similarity and long-range dependence through the wavelet lens, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, '*Theory and Applications of Long-range Dependence*', Birkhäuser, pp. 527–556.

Bardet, J.-M. (2002), 'Statistical study of the wavelet analysis of fractional Brownian motion', *IEEE Transactions on Information Theory* **48**, 991–999.

Bardet, J.-M., Lang, G., Moulines, E. & Soulier, P. (2000), 'Wavelet estimator of long-range dependent processes', *Statistical Inference for Stochastic Processes* **3**, 85–99.

Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A., Stoev, S. & Taqqu, M. (2003), Semiparametric estimation of the long-range dependence parameter : A survey, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, '*Theory and Applications of Long-range Dependence*', Birkhäuser, pp. 579–623.

Cao, J., Cleveland, W., Lin, D. & Sun, D. (2002), Internet traffic tends toward poisson and independent as the load increases, *in* C. Holmes, D. Denison, M. Hansen, B. Yu & B. Mallick, eds, 'Nonlinear Estimation and Classification', Springer, New York.

Cohen, A. (2003), *Numerical Analysis of Wavelet Methods*, Vol. 32, North-Holland.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM Philadelphia. CBMS-NSF series, Volume 61.

Flandrin, P. (1999), *Time-Frequency/Time-scale Analysis*, 1st edn, Academic Press.

Hernandez-Campos, F., Le, L., Marron, J.-S., Park, C., Park, J., Pipiras, V., Smith, F. D., Smith, R. L., Trovero, M. & Zhu, Z. (2004), Long range dependence analysis of Internet traffic, In preparation.

Kim, M. & Tewfik, A. H. (1992), 'Correlation structure of the discrete wavelet coefficients of fractional Brownian motion', *IEEE Transactions on Information Theory* **38**(2), 904–909.

Leland, W. E., Taqqu, M. S., Willinger, W. & Wilson, D. V. (1993), 'On the self-similar nature of Ethernet traffic', *Computer Communications Review* **23**, 183–193. Proceedings of the ACM/SIGCOMM'93, San Francisco, September 1993. Reprinted in *Trends in Networking – Internet*, the conference book of the Spring 1995 Conference of the National Unix User Group of the Netherlands (NLUUG). Also reprinted *Computer Communication Review*, **25**, Nb. 1 (1995), 202-212, a special anniversary issue devoted to "Highlights from 25 years of the Computer Communications Review".

Mallat, S. (1998), *A Wavelet Tour of Signal Processing*, Academic Press, Boston.

Mallat, S. G. (1989), 'A theory for multiresolution signal decomposition: the wavelet representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **11**(7), 674–693.

Ogden, T. (1996), *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser, Boston.

Park, C., Godtliebsen, F., Taqqu, M., Stoev, S. & Marron, J. S. (2004), Wavelet coefficient-based visualization and inference, Preprint.

Park, C., Rolls, D., Hernandez-Campos, F., Smith, F. D. & Marron, J. S. (2004), An in depth analysis of changes in IP traffic, In preparation.

Paxson, V. & Floyd, S. (1995), 'Wide area traffic: The failure of Poisson modeling', *IEEE/ACM Transactions on Networking* **3**, 226–244.

Pipiras, V., Taqqu, M. S. & Abry, P. (2001), Asymptotic normality for wavelet-based estimators of fractional stable motion, Preprint.

Rioul, O. & Vetterli, M. (1991), 'Wavelets and signal processing', *IEEE Signal Processing Magazine*.

Robinson, P. M. (1995), 'Gaussian semiparametric estimation of long range dependence', *The Annals of Statistics* **23**, 1630–1661.

Stoev, S. & Taqqu, M. S. (2003), 'Asymptotic self-similarity and wavelet estimation for long-range dependent FARIMA time series with stable innovations', *Journal of Time Series Analysis*. To appear (Preprint 2002).

Stoev, S., Pipiras, V. & Taqqu, M. S. (2002), 'Estimation of the self-similarity parameter in linear fractional stable motion', *Signal Processing* **82**, 1873–1901.

Stoev, S., Taqqu, M., Park, C., Michailidis, G. & Marron, J. S. (2004), LASS: a tool for the local analysis of self-similarity, Preprint.

Taqqu, M. S. (2003), Fractional Brownian motion and long-range dependence, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, '*Theory and Applications of Long-range Dependence*', Birkhäuser, pp. 5–38.

Taqqu, M. S. & Teverovsky, V. (1997), 'Robustness of Whittle-type estimates for time series with long-range dependence', *Stochastic Models* **13**, 723–757.

Taqqu, M. S. & Teverovsky, V. (1998), On estimating the intensity of long-range dependence in finite and infinite variance series, *in* R. Adler, R. Feldman & M. S. Taqqu, eds, '*A Practical Guide to Heavy Tails: Statistical Techniques and Applications*', Birkhäuser, Boston, pp. 177–217.

Taqqu, M. S., Teverovsky, V. & Willinger, W. (1995), 'Estimators for long-range dependence: an empirical study', *Fractals* **3**(4), 785–798. Reprinted in *Fractal Geometry and Analysis*, C.J.G. Evertsz, H-O Peitgen and R.F. Voss, editors. World Scientific Publishing Co., Singapore, 1996.

Veitch, D. & Abry, P. (1999), 'A wavelet-based joint estimator of the parameters of long-range dependence', *IEEE Transactions on Information Theory* **45(3)**, 878–897.

Veitch, D., Taqqu, M. S. & Abry, P. (2000), 'Meaningful MRA initialisation for discrete time series', *Signal Processing* **80**, 1971–1983.

Veitch, D., Taqqu, M. S. & Abry, P. (2003), 'On the automatic selection of the onset of scaling', *Fractals* **11**(4), 377–390.

Vetterli, M. & Kovacevic, J. (1995), *Wavelets and Subband Coding*, Prentice Hall.

Vidakovic, B. (1999), *Statistical Modeling by Wavelets*, Wiley.

Zhang, Z.-L., Ribeiro, V., Moon, S. & Diot, C. (2003), Small-time scaling behaviors of internet backbone traffic: An empirical study, *in* 'IEEE Infocom', San Francisco.

Stilian Stoev
Department of Mathematics and Statistics,
Boston University,
Boston, MA 02215
sstoev@bu.edu

Murad Taqqu
Department of Mathematics and Statistics,
Boston University,
Boston, MA 02215
murad@bu.edu

Cheolwoo Park
Statistical and Applied Mathematical Sciences Institute,
19 T. W. Alexander Drive, P.O. Box 14006,
Research Triangle Park, NC 27709-4006
cwpark@samsi.info

J. S. Marron
Department of Statistics,
University of North Carolina,
Chapel Hill, NC 27599-3260
marron@email.unc.edu