Visualization and inference based on wavelet coefficients, SiZer and SiNos

CHEOLWOO PARK

Statistical and Applied Mathematical Sciences Institute,

19 T. W. Alexander Drive, P.O. Box 14006,

Research Triangle Park, NC 27709-4006

FRED GODTLIEBSEN

Department of Mathematics and Statistics,

University of Tromsø,

N-9037 Tromsø, Norway

MURAD TAQQU

Department of Mathematics and Statistics,

Boston University,

Boston, MA 02215

STILIAN STOEV

Department of Mathematics and Statistics,

Boston University,

Boston, MA 02215

J. S. MARRON

Department of Statistics and Operations Research,

University of North Carolina,

Chapel Hill, NC 27599-3260

March 5, 2004

<div align="center">SUMMARY</div>

SiZer (SIgnificant ZERo crossing of the derivatives) and SiNos (SIgnificant NOnStationarities) are scale-space based visualization tools for statistical inference. They are used to discover meaningful structure in data through exploratory analysis involving statistical smoothing techniques. Wavelet methods have been successfully used to analyze various types of time series. In this paper, we propose a new time series analysis approach, which combines the wavelet analysis with the visualization tools SiZer and SiNos. We use certain functions of wavelet coefficients at different scales as inputs, and then apply SiZer or SiNos to highlight potential non-stationarities. We show that this new methodology can reveal hidden local non-stationary behavior of time series, that are otherwise difficult to detect.

Key words: Internet Traffic, Long–range dependence, Nonstationarity, Scale-space method, SiNos, SiZer, Time Series, Wavelet coefficients.

1   Introduction

1.1   Background

In a complicated time series, it can be challenging to detect underlying structure. For example, Internet traffic data typically have long–range dependence properties, see Leland, Taqqu, Willinger, and Wilson (1994), and even nonstationary behavior, see Hernandez-Campos et al (2004). In this context and in general, it is difficult to distinguish between long–range dependence and nonstationarity. Some work in this direction includes Teverovsky and Taqqu (1997), Dang and Molnár (1999), Mikosch and Stărică (2000), Rondonotti and Marron (2001), Giraitis, Kokoszka and Leipus (2001) and Stoev, Taqqu, Park, Michailidis and Marron (2004).

In this paper, we propose a new method to detect hidden nonstationary behavior of time series which may possess long–range dependence. An important advantage of this method is that it can be used to identify stationary segments of the process as well as localized nonstationary fluctuations. By using this procedure, one may be able to better understand and possibly explain the observed behavior in nonstationary regions of the process. The methodology is potentially useful in many applications, including financial mathematics, climatology, speech recognition, imaging and Internet traffic data. In this paper, we illustrate our methods in the Internet traffic context.

Many tools have been developed for the analysis of the long–range dependence structure

<div align="center">2</div>

of a time series. Some of the most popular ones are, for example, the *aggregation of variance* method, see Beran (1994), or Teverovsky and Taqqu (1997), the *periodogram based Whittle* method, see Robinson (1995), or Taqqu and Teverovsky (1997), and the *wavelet based* method, see Abry and Veitch (1998), or Feldmann, Gilbert, Huang, and Willinger (1999), Doukhan, Oppenheim, and Taqqu (2003). Among these, the wavelet methods have been very popular since they allow us to visualize the scaling behavior of the data (e.g. Internet traffic) as well as to obtain estimates of the Hurst long–range dependence parameter see Stoev, Taqqu, Park, and Marron (2004). The Hurst parameter quantifies the degree of long–range dependence of a time series (see, for example, Berran (1994)).

SiZer (SIgnificant ZERo crossing of the derivatives), which was originally proposed by Chaudhuri and Marron (1999), combines the scale-space idea of simultaneously considering a family of smooths with the statistical inference that is needed for exploratory data analysis in the presence of noise. It brings an immediate insight into a central scientific issue in exploratory data analysis: *which features observed in a smooth of data are "really" there?*. *Dependent SiZer*, see Rondonotti and Marron (2001) and Park, Marron, and Rondonotti (2004) extends SiZer to dependent data and the latter paper provides a goodness of fit test of the underlying model. This can be done by adjusting the statistical inference using an autocovariance function of the supposed model. Dependent SiZer is particularly useful in the analysis of Internet traffic data. Indeed, it provides not only visual insights in the structure of these data but also involves statistical tests for significance of the visually observed features against those in an assumed noise model.

In addition to the scale-space methods mentioned above, we utilize the *SiNos* method, developed by Olsen and Godtliebsen (2003), to extract information from the wavelet coefficients. SiNos simultaneously looks for significant changes in the mean, variance, and the first lag autocorrelation of the observed time series under the null hypothesis that the process is stationary.

1.2 Motivating examples

Figure 1 about here.

Figure 1 (a) displays a time series of binned packet counts in Internet traffic coming into the University of North Carolina, Chapel Hill (UNC). They were measured at the link of UNC on April 11, Thursday, from 1 p.m. to 3 p.m., 2002 (Thu1300). This time series,

which is the number of packets arriving at the link every 10 milliseconds, can be written as $Y = \{Y(i),\ i = 1, 2, \ldots, 726000\}$. The plot shows several spikes shooting up and down. Figure 1 (b) displays what we call the wavelet spectrum of the Thu1300 time series. Here, the wavelet spectrum shows signal power obtained by wavelet coefficients (discussed in more detail in Section 2.1) at each scale. See Abry and Veitch (1998), or Stoev, Taqqu, Park, and Marron (2004) for more details on the wavelet spectrum. We used Daubechies wavelets, see for example, Daubechies (1992), with three zero moments for all wavelet spectra in this paper. Details of this plot are explained in Section 2.1. The wavelet spectrum of this time series shows a bump at the scale $j = 11$. Since the y–axis of the spectrum corresponds to the *energy* of the signal at a given scale, this bump suggests a unusual behavior at $j = 11$. This has a serious impact on the estimation of the Hurst parameter (see Stoev, Taqqu, Park, and Marron (2004) for more details). A serious weakness of the wavelet spectrum statistics is that they fail to provide information about the time location of the unusual behavior, because they involve an average through time.

Since Internet traffic data are known to exhibit the self-similar property as well as long–range dependence, see Leland, Taqqu, Willinger, and Wilson (1994), Fractional Gaussian Noise (FGN) has been a useful model of Internet traffic behavior. Figures 1 (c) and (d) display the dependent SiZer of the Thu1300 time series. The data are tested against a null hypothesis of FGN where the Hurst parameter, H, and the variance are 0.9 and 485.6 respectively (see Section 2.2 for the choice of the parameters). The goal is to see whether FGN can model the data at large time scales, and if not, to see how far the data are from FGN. Because FGN is a stationary series, significant features discovered in the dependent SiZer analysis can be considered as nonstationarities.

The thin blue curves in Figure 1 (c) display a family of smooths of the Thu1300 time series. These are kernel regression estimates of the underlying curves. The curve estimates are obtained from the observations, some of which are displayed as jittered green dots in Figure 1 (c). The blue curves correspond to different levels of smoothing, i.e. bandwidths. The goal of dependent SiZer is to determine which features of the blue curves are different from those features one would expect to get by the assumed FGN model in this example. Observe that a sharp valley appears around $x = 2500$ (seconds). Figure 1 (d) displays the SiZer map of the data. The horizontal locations in the SiZer map are the same as in the top panel, and the vertical locations correspond to the same logarithmically spaced bandwidths that were used for the family of smooths (blue curves) in the top panel. Blue (red) colors in the SiZer map indicate that the blue curves in Figure 1 (c) are significantly increasing

4

(decreasing, respectively), at the point indexed by the horizontal location and the bandwidth corresponding to that row, compared to FGN. Purple colors mean that the trend shows no increase or decrease that is significantly different from those typical of FGN at that particular location. We will revisit the details in Section 2.2. The SiZer map in Figure 1 (d) suggests that the valley around $x = 2500$ is different from FGN with the given parameters at fine scales. This indication is not strong in the sense that there are only a few bandwidths (scales) and locations in the SiZer map where the valley around $x = 2500$ is flagged as significant. Since dependent SiZer provides both location and scale information simultaneously, it is very useful to find nonstationary behavior when a stationary model for the noise is assumed. However, it sometimes fails to locate nonstationarities that have short durations. Furthermore, the model must be specified before analyzing the data. Hence, estimation of parameters in the autocovariance function of the specified FGN model for the noise is needed.

Nonstationary behavior in the data can be further explored by zooming in on the region of the nonstationarity. Because we have evidence of unusual behavior in the Thu1300 time series from Figure 1, one can split the full time series into several subtraces and apply the wavelet method and dependent SiZer to the subtraces.

Figure 2 about here.

Since the wavelet spectrum in Figure 1 (b) provides only scale information aggregated over the full time series, we use these subtraces of the data to find more local information about the nonstationarities. The full time series is divided into 12 subtraces, each of which is 600 seconds long. Figure 2 (b) shows the wavelet spectra of subtraces of the Thu1300 time series. Each spectrum is obtained from one of the subtraces. The spectrum with vertical gray lines corresponds to the full time series. The spectrum with the thick black line corresponds to the subtrace marked by red vertical lines in Figure 2 (a), and this subtrace includes the location where nonstationary behavior was flagged in the dependent SiZer plot in Figures 1 (c) and (d). The wavelet spectrum corresponding to the red window in Figure 2 (a) shows that the bump at the scale $j = 11$ of the full time series comes from this particular subtrace. Figures 2 (c) and (d) display the dependent SiZer plot for this subtrace. The data is tested on FGN with the same parameters as in Figures 1 (c) and (d). This strongly confirms that the valley is significantly different from FGN with the given parameters.

Figures 1 and 2 show that by combining a zoomed view with the wavelet spectrum and dependent SiZer, we can find locations where the time series has strong nonstationary be-

havior. However, there are serious practical limitations to this approach, caused by the lack of location information in the wavelet spectrum, by the need to choose the length of subtraces, by the need to specify an assumed model for the data, and by the need for parameter estimation in dependent SiZer.

Motivated by the idea of combining the wavelet spectrum and dependent SiZer, and by overcoming the difficulties that they cause, we propose a new visualization tool, entitled wavelet SiZer. This tool combines wavelet coefficients with the original SiZer. Wavelet SiZer uses functions of wavelet coefficients at different scales as inputs. Next, SiZer is used to determine where and how they are significantly different from what one would have expected, had the data been stationary. By doing so, wavelet SiZer considers not only the wavelet coefficients at all scales, but also adds location information at each scale obtained from SiZer. One advantage, compared to applying dependent SiZer on the raw time series, is that conventional SiZer can be used on the wavelet coefficients because they are essentially uncorrelated (see Section 2.1). This removes the burdens of model specification and parameter estimation.

<div style="border:1px solid; display:inline-block; padding:4px;">Figure 3 about here.</div>

Figure 3 shows the wavelet SiZer of the Thu1300 time series. The top panel displays the family of smooths of the full time series. When SiZer is applied directly to the wavelet coefficients, very few features are typically found. It turns out, however, that by applying SiZer to functions of the wavelet coefficients, interesting features in the underlying time series can be found. In particular, the squared wavelet coefficients are useful for finding features of interest. This is illustrated in the remaining rows of Figure 3, where a SiZer analysis is applied to the squared wavelet coefficients. Section 3 further explores this point. We used Daubechies wavelets with three zero moments for all wavelet SiZer plots in this paper. The second row displays the family of smooths and the SiZer map of the square of the wavelet coefficients at the scales $j = 1$ (the first and the second column) and $j = 2$ (the third and the fourth column). In the same way, the other rows display SiZer plots of the square of the wavelet coefficients from $j = 3$ to $j = 10$. The larger scales are not presented here because they do not provide further useful information. The SiZer plots in Figure 3 at different scales are mostly dominated by four major spikes, labelled as 1, 2, 3, and 4 in the family of smooths at scale $j = 7$. Spike 1 appears from $j = 1$ to $j = 5$, which are fine scales. Spike 2 comes up at $j = 5$ and lasts until $j = 10$, a relatively coarse scale. Note that this spike matches the location that dependent SiZer had found in Figures 1 and 2. Spikes 3 and 4 appear from

$j = 2$ to $j = 9$, which cover both fine and coarse scales. Note that spikes 1, 3, and 4 could not be observed in Figures 1 and 2, but the wavelet SiZer is able to reveal these hidden local nonstationarities.

Figure 4 about here.

However, the SiZer maps do not confirm the significance of all four spikes. For example, spike 2 is not significant for all $j = 5, \ldots, 10$ according to the SiZer maps although the smoothing families clearly show the sharp peak. Another scale-space visualization tool which can be combined with wavelet coefficients is SiNos. This methodology explores potential nonstationarities in a stochastic process, see Olsen and Godtliebsen (2003). We describe more details of this methodology in Section 2.3. Here, we simply show a plot of the significant changes detected in the variance of the wavelet coefficients, which is related to the second power wavelet SiZer, at scale $j = 9$ for the Thu1300 data set. In the upper panel of Figure 4, each curve represents an estimate for the variance of the time series. Different curves correspond to different degrees of smoothing. The lower panel of Figure 4 shows the feature map for the variance of the time series. This feature map is entitled a Variance SiNos plot. Black, white and gray colors means increasing, decreasing and no change in the variance respectively. Light gray shows areas where no inference can be done. This shows the added value of SiNos, because both peaks are flagged as significant, while they were not flagged by wavelet SiZer in the lower left part of Figure 3.

Figure 5 about here.

Having found these 4 regions of strong local nonstationarity, it is natural to investigate more deeply. To do so, we zoom into the original time series. The four regions highlighted in red in Figure 5 (a) correspond to the locations of the four spikes flagged as significant by wavelet SiZer in Figure 3. Figures 5 (b), (c), (d), and (e) show the zoomed time series corresponding to the four red windows in Figure 5 (a), respectively. Figure 5 (b) shows that there is a dropout (absence of signal) for a duration about 0.1 second perhaps caused by a short router pause, which is a clear type of nonstationarity. Since this dropout is quickly over with, the spike appears only at fine scales in wavelet SiZer. Figure 5 (f) displays the wavelet spectrum for the full time series with this subtrace (shown in Figure 5 (b)) removed and the remaining two parts concatenated. This does not look different from the original spectrum in Figure 1 (b), suggesting that this feature did not cause the $j = 11$ bump in the spectrum.

Figure 5 (c) shows a long dropout for about 8 seconds, again a strong local nonstationarity. Since this dropout is much longer, the spike appears at much coarser scales in wavelet SiZer in Figure 3. The spike appears at a number of scales because the signal power of a jump discontinuity is spread across a range of wavelet scales. Figure 5 (g) displays the wavelet spectrum of the full time series with the subtrace (shown in Figure 5 (c)) removed as for Figure 5 (f), and the bump at $j = 11$ in Figure 1 (b) disappears. This suggests that this feature plays an important role in causing the $j = 11$ bump. Figures 5 (d) and (e) show two unusual drops in both subtraces. Because the sizes of drops are about half of the size of the signal, this could be caused by one of two routers pause. Since some of them are short and some are long, and since they are located relatively close to each other, the wavelet coefficients feel the effect over many scales from fine to coarse in Figure 3. Although Figures 5 (h) and (i) display the wavelet spectra of the full time series with the subtraces (shown in Figure 5 (d) and (e) respectively) removed as above, they are not different from the original spectrum. It is interesting to see that all spikes in wavelet SiZer come from valleys in the full time series, rather than upward spikes. We conclude that the long dropout in Figure 5 (c) causes the bump at $j = 11$ in the original wavelet spectrum in Figure 1 (b) by comparing the original spectrum with the one in 5 (g) because the bump disappears. Also, additional hidden nonstationarities were found by wavelet coefficient based visualization tools.

This example suggests that the idea of combining wavelet coefficients and scale-space methods is a powerful technique for detecting hidden local nonstationary behavior of time series. In fact, the new tools enable us to easily find underlying structures in the original time series that would have been hard to find by simply examining wavelet coefficients. Although we have focused on Internet traffic data in this paper, the described methodology is also applicable and useful for other types of time series with potential local nonstationarities.

Section 2 describes the details of the wavelet spectrum and scale-space methods. Details and issues about wavelet coefficient based scale-space methods are provided in Section 3. Additional real data analysis and the proposal of a new graphical device summarizing scale information are presented in Section 4. This new graphical device greatly reduces the number of plots that must be studied to find significant features, which is very convenient for exploring multiple time series. Section 5 provides some concluding remarks.

2 Wavelets and scale-space inference

This section describes the wavelet method proposed by Abry and Veitch (1998), and the SiZer method proposed by Chaudhuri and Marron (1999) and Park, Marron, and Rondonotti (2004). In addition, a description of SiNos, proposed by Olsen and Godtliebsen (2003), is given.

2.1 Wavelet spectrum

Here, we briefly review the notion of wavelet spectrum of a time series. We discuss its interpretation and sketch its use for the estimation of the Hurst long–range dependence parameter. More details and deeper insights can be found in the seminal works of Abry and Veitch (1998, 1999) (see also Abry, Flandrin, Taqqu, and Veitch (2003)).

Consider a discrete time series $Y = \{Y(i), \ i = 1, \ldots, N\}$. Using Mallat's fast discrete wavelet transform algorithm one obtains the set of transformation coefficients of Y:

$$\{d_{j,k}, \quad k = 1, \ldots, N_j\}, \ j = 1, \ldots, J,$$

where $N_j \approx N/2^j$ and $J \approx \log_2 N$. These coefficients are computed efficiently in $\mathcal{O}(N)$ operations. The coefficients $d_{j,k}$ can be represented as:

$$d_{j,k} = \int_{\mathbb{R}} \widetilde{Y}(t)\psi_{j,k}(t)dt, \tag{2.1}$$

where $\psi_{j,k}(t) := 2^{-j/2}\psi(2^{-j}t - k)$, $j, k \in \mathbb{Z}$ and where $\widetilde{Y}(t)$, $t \in \mathbb{R}$ is a suitable continuous-time *approximation* of the time series Y.

The function ψ involved in (2.1) is called an orthonormal mother wavelet. It is chosen so that the set of dyadic dilations and integer translations $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$, $j, k \in \mathbb{Z}$ of ψ becomes an orthonormal basis of the space $L^2(dt)$. The class of Daubechies wavelets ψ is particularly useful in practice. These wavelets have compact support and a number of other important properties. For more details on the discrete wavelet transform and Mallat's algorithm see, for example, Ch. 6 in Daubechies (1992).

In view of (2.1), the coefficient $d_{j,k}$ captures features of the signal $\widetilde{Y}(t)$, which match the time-location ($\approx 2^j k$) and the time-scale ($\approx 2^j$) of the basis function $\psi_{j,k}$. Therefore, the indices j and k of the $d_{j,k}$'s are typically called *scale* and *location*, respectively. For large j, the support of $\psi_{j,k}$ is wide and consequently the $d_{j,k}$'s extract coarse scale or low frequency

9

features of $\widetilde{Y}(t)$. Conversely, the wavelet coefficients at small scales j contain fine scale or high frequency details of the signal.

Suppose now that $Y = \{Y(i)\}_{i \in \mathbb{Z}}$ is a second order stationary random time series. The wavelet coefficients $d_{j,k}$ of Y reflect naturally its self–similarity and long–range dependence properties. Indeed, for all j, the time series $d_{j,k}$, $k \in \mathbb{Z}$ is stationary and, as $j \to \infty$, one has that

$$\log_2 \left(\mathbb{E}d_{j,k}^2 \right) \sim j(2H - 1) + C, \tag{2.2}$$

where C does not depend on j and where H denotes the Hurst long–range dependence exponent of the time series Y. Furthermore, even though the $Y(i)$'s can be strongly dependent, in practice, for each fixed scale j the wavelet coefficients $d_{j,k}$ are essentially uncorrelated in k (see Abry and Veitch (1998) and, for example, Stoev, Taqqu, Park, and Marron (2004)). Using this fact one can estimate well the mean energy $\mathbb{E}d_{j,k}^2$ of the wavelet coefficients on scale j by using sample statistics. Namely, let

$$S_j := \log_2 \left(\frac{1}{N_j} \sum_{k=1}^{N_j} d_{j,k}^2 \right) - g_{N_j}(j),$$

where $g_{N_j}(j) \approx 1/(\ln(2)N_j)$ is a suitable first order bias correction term (for more details see, Abry and Veitch (1998)). The statistics S_j are asymptotically unbiased estimators of the quantities $\log_2(\mathbb{E}d_{j,k}^2)$.

The set of statistics S_j, $j = 1, \ldots, J$ is called the *wavelet spectrum* of the time series $Y(i)$, $i = 1, \ldots, N$. The wavelet spectrum can be related to the classical Fourier spectrum of the time series (see Abry and Veitch (1998) and, for example, Stoev, Taqqu, Park and Marron (2004)). Large scales j correspond to low frequency features in the spectral density of Y and therefore the statistics S_j represent the long–range dependence properties of the data (see (2.2), above). The statistics S_j for small scales j, however, capture high-frequency features pertinent to the short term dependence structure of the time series.

In view of (2.2) one can use the wavelet spectrum to estimate the Hurst parameter H. Indeed, let $1 \le j_1 < j_2 \le J$ and set

$$\widehat{H} := \left(\sum_{j=j_1}^{j_2} w_j S_j + 1 \right)/2,$$

where $\sum_{j=j_1}^{j_2} w_j = 0$ and $\sum_{j=j_1}^{j_2} j w_j = 1$. Using such weights w_j, the estimator $\widehat{H}$ is obtained from the slope $2\widehat{H} - 1$ of a weighted linear regression fit of S_j against j over the range of

scales $j_1, j_1 + 1, \ldots, j_2$. Since (2.2) holds for large j, when estimating the Hurst parameter H, one focuses on the largest scales of the wavelet spectrum. These scales, however, involve fewer wavelet coefficients and hence the statistics S_j have greater variability. The choice of the range of scales $[j_1, j_2]$ is a subtle problem in the estimation of H. It is partly addressed in Veitch, Abry and Taqqu (2003). More details on the asymptotic statistical properties of $\widehat{H}$ and other related wavelet-based estimators of the Hurst parameter can be found in Bardet, Lang, Moulines, and Soulier (2000), see also Pipiras, Taqqu and Abry (2001) and Bardet, Lang, Oppenheim, Philippe, Stoev, and Taqqu (2002) and the references therein.

While the Hurst parameter is important, it is not the only parameter of interest. The whole range of the wavelet spectrum can carry useful information about the data. In Stoev, Taqqu, Park and Marron (2004), the strengths and the limitations of the wavelet spectrum in Internet traffic context are explored. As indicated in Section 1.2, one major limitation of the statistics S_j is that they can average out important time location information contained in the wavelet coefficients $d_{j,k}$. To obtain a richer picture, which captures interesting local nonstationarity features in of time series, one should analyze in detail the time series of wavelet coefficients $d_{j,k}$. This is done in the following section using natural scale-space smoothing tools such as SiZer and SiNos.

2.2   SiZer and dependent SiZer

SiZer analysis is a visualization method which enables statistical inference for discovery of meaningful structure within the data, while doing exploratory analysis using statistical smoothing methods. In particular, SiZer addresses the question of "which features observed in a smooth are really there?", meaning representing important underlying structure, not artifacts of the sampling noise.

SiZer is based on scale-space ideas from computer vision, see Lindeberg (1994). Scale-space is a family of kernel smooths indexed by the scale, which is the smoothing parameter or bandwidth h. SiZer considers a wide range of bandwidths which avoids the classical problem of bandwidth selection. Furthermore, the target of a SiZer analysis is shifted from finding features in the "true underlying curve" to inferences about the "smoothed version of the underlying curve", i.e. the "curve at the given level of resolution". The idea is that this approach contains all the information that is available in the data at each given scale.

SiZer visually displays the significance of features over both location x and scale h, using a color map. It is based on confidence intervals for the derivatives of the underlying curve

11

and it uses multiple comparison level adjustment. Each pixel shows a color that gives the result of a hypothesis test for the slope of the smoothed curve, at the point indexed by the horizontal location x, and by the bandwidth corresponding to the row h. At each (x, h), if the confidence interval is above (below) 0, which means that the curve is significantly increasing (decreasing), then that particular map location is colored blue (red, respectively). On the other hand, if the confidence interval contains 0, which means that the curve is not significantly increasing or decreasing, then that map location is given the intermediate color of purple. Finally, if there are not enough data points to carry out the test, then no decision can be made and the location is colored gray.

Let us consider a regression problem with a fixed design setting. Given the data $(x_i, Y(i))$ where $x_i = i/N$, for $i = 1, \ldots, N$, a regression problem is described as

$$Y(i) = g(x_i) + \sigma \epsilon_i, \tag{2.3}$$

where g is a regression function, $\sigma > 0$, and the ϵ_i's are identically and independently distributed with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = 1$ for all i. A time series setting can be viewed as a regression setting in (2.3) with $x_i = i$. The other difference is that the ϵ_i's are not generally independent.

SiZer applies the local linear fitting method, see e.g. Fan and Gijbels (1996), for obtaining a family of kernel estimates and derivatives of a regression function. Precisely, at a particular point x_0, they are obtained by minimizing

$$\sum_{i=1}^{N} \{Y(i) - (\beta_0 + \beta_1(x_0 - x_i))\}^2 K_h(x_0 - x_i) \tag{2.4}$$

over $\beta = (\beta_0, \beta_1)'$, where $K_h(\cdot) = K(\cdot/h)/h$. K is called a kernel function which is usually a symmetric probability density function. By Taylor expansion, it is easy to show that $\beta_0 \approx g(x_0)$, and $\beta_1 \approx g'(x_0)$, thus the solution of (2.4) provides estimates of a regression function and its first derivatives for different bandwidths. From this solution, we can construct the family of smooths and the confidence intervals that underlie the SiZer analysis. The details can be found in Chaudhuri and Marron (1999). Hannig and Marron (2003) have recently suggested a procedure, using advanced distribution theory, to improve the multiple comparison tests. We use the original version of SiZer in this paper because the main lessons are the same for both procedures.

SiZer is a useful tool to find meaningful structures in the given data, but its usefulness can be diminished in the case of dependent data because it assumes independent errors

12

and compares against a white noise null hypothesis. In cases of dependent data, significant features in SiZer are often generated as dependence artifacts. Examples for this issue were presented in Park, Marron, and Rondonotti (2004). This motivates the need for distinguishing trends due to dependence. Rondonotti and Marron (2001) extended SiZer to time series. That method finds features of the underlying trend function, while taking into account the dependence structure.

Dependent SiZer has a slightly different goal from SiZer for time series. The dependent SiZer, proposed by Park, Marron, and Rondonotti (2004), uses a true autocovariance function of an assumed model instead of estimating it from the observed data. By doing so, a goodness of fit test can be conducted and we can see how different the behavior of the data is from that of the assumed model. The only difference between original SiZer and dependent SiZer is that the latter compares the data with the specified model rather than with white noise.

Implementation of dependent SiZer requires estimation of the parameters involved in the autocovariance function. The autocovariance function of FGN, $\gamma(\cdot)$ is given by

$$\gamma(l) = \tau^2((l+1)^{2H} + (l-1)^{2H} - 2l^{2H})/2, \quad l = 0, 1, 2, \ldots,$$

where H is the Hurst parameter and τ^2 is the variance parameter. Thus, one needs to estimate H and τ^2 to conduct a goodness of fit test using dependent SiZer. This parameter estimation is important and detailed discussion of this issue can be found in Park, Marron, and Rondonotti (2004), and Hernandez-Campos et al (2004). In these papers, they used richer data sets to obtain good estimates of the two parameters.

2.3 SiNos

SiNos (SIgnificant NOnStationarities) is a scale-space method based on the idea that we estimate and check for local nonstationarity in terms of the mean (μ), variance (σ^2), and lag one autocorrelation (ρ) of a time series. This methodology is designed to handle both dependent and independent data (see Olsen and Godtliebsen (2003)), using model assumptions of the type used by dependent SiZer.

Different window widths M are used to explore changes on different scales. The basic idea is : First, we compare two consecutive independent estimates $\widehat{est}_r$ and $\widehat{est}_l$, where r and l denote right and left and $\widehat{est}$ is an estimate of μ, σ^2 or ρ. Second, we evaluate whether they are significantly different. The result (for each of the three parameters) is reported through

a family plot and a feature map as is done by SiZer. With each window (relabelling the data inside as $Y(1), \ldots, Y(M)$), the estimators for the three parameters are :

$$
\begin{aligned}
\hat{\mu} &= \overline{Y} = \frac{1}{M} \sum_{i=1}^{M} Y(i) \\
\hat{\sigma}^2 &= \frac{1}{M} \sum_{i=1}^{M} (Y(i) - \overline{Y})^2 \\
\hat{\rho} &= \frac{\sum_{i=1}^{M-1} (Y(i+1) - \overline{Y})(Y(i) - \overline{Y})}{\sum_{i=1}^{M-1} (Y(i) - \overline{Y})^2}
\end{aligned}
$$

These estimators are generally asymptotically normally distributed for time series defined in terms of linear combinations of an underlying sequence of white noise variates. See Shumway and Stoffer (2000) for a detailed set of conditions under which this holds. A "rule of thumb" for the number of observations needed to get a mean value approximately normally distributed is to use at least 30 observations. Hence, smaller window widths than 30 are never used.

The time series is divided into windows of M observations and μ, σ^2 and ρ are estimated for each window width using a moving window. If we compare two consecutive segments of observations, the two compared estimates will be correlated. To simplify the calculations we choose the distance between $\widehat{est}_r$ and $\widehat{est}_l$ large enough to assume that the two estimates are independent. A simple choice for this distance can be found by assuming that the autocorrelation of the process follows an AR1 model. Since, the lag m autocorrelation of an AR1 process is given by $\text{Corr}(Y(l+m), Y(l)) = \rho^m$, the distance m_0 between two compared segments is large enough for them to be considered "independent" when $\rho^m \leq 0.05$, i.e. when $m_0 \geq \ln(0.05)/\ln(\hat{\rho})$.

Some experimentation revealed that the variance part of SiNos is frequently the most powerful for detecting features in the wavelet coefficients. Hence, we focus here on how the variance of the estimated $\hat{\sigma}^2$ is found. For a stationary process,

$$
\text{Var}(\hat{\sigma}^2) = \frac{2}{M} \left[\sigma^2 + 2 \sum_{m=1}^{M} \left(1 - \frac{|m|}{M} \right) R(m)^2 \right] \tag{2.5}
$$

where $R(m)$ is the autocovariance of the process at lag m. The autocovariances are in practice estimated by

$$
\widehat{R(m)} = \frac{1}{M} \sum_{i=1}^{M-m} (Y(i+m) - \overline{Y})(Y(i) - \overline{Y})
$$

14

By plugging this into (2.5), an estimate of $\text{Var}(\hat{\sigma}^2)$ is found. Estimated variances of $\overline{Y}$ and $\hat{\rho}$ are found in an analogous way.

To test for changes in the estimated parameters we use the fact that the estimates are asymptotically normally distributed. The test is given by

$$\left| \frac{\widehat{est}_r - \widehat{est}_l}{\sqrt{2\text{Var}(\widehat{est})}} \right| > u_{\alpha'/2} \tag{2.6}$$

where $u_{\alpha'/2}$ is a suitable Gaussian quantile. The significance level, α', for each test is chosen by the number of independent blocks approach developed in SiZer by Chaudhuri and Marron (1999). More details about the choice of α' is given in Olsen and Godtliebsen (2003).

In this paper we apply SiNos, not to the original data which can have long–range dependencies, but to the wavelet coefficients which, under common assumptions, are essentially uncorrelated.

3 Wavelet coefficient based scale-space methods

As pointed out in Section 1, using both the wavelet spectrum and dependent SiZer can provide information about nonstationary behavior. However, the wavelet spectrum lacks location information and dependent SiZer needs specification of a model and estimation of the parameters of the autocovariance function in the assumed model. Wavelet coefficient based scale-space methods combine wavelet decompositions and scale-space tools to provide deeper insights. First, the time series is decomposed in the wavelet domain and then SiZer or SiNos are applied to each wavelet scale.

The results of statistical inference are indexed by a location parameter k, which is related to a location in the original time series, and by two scale parameters which are j (frequency) and h (bandwidth in SiZer) or M (window width in SiNos).

> Figure 6 about here.

In wavelet SiZer, any function f, for example $f(x) = |x|^p$, $p > 0$, of the wavelet coefficients $\{d_{j,k}\}$ in (2.1) of the data can be used as an input for a SiZer analysis. Because wavelet coefficients are short–range dependent even when the original time series is long–range dependent, as indicated in Section 2.1, it is possible to apply the conventional SiZer to $f(\{d_{j,k}\})$. Figure

15

6 shows the wavelet SiZer of FGN with $H = 0.9$ and sample size $N = 720,000$. In this case, the square of the wavelet coefficients (second power), i.e. $\{d_{j,k}^2\}$, are displayed. In principle, wavelet coefficients are available at scales $j = 1, \ldots, [\log_2(N)]$ where $[\log_2(N)]$ means the greatest integer that does not exceed $\log_2(N)$. However, only 10 scales are presented in this paper because coarser scales usually do not provide further information and the very coarse scales contain only a few data points. The shape of the family of smooths looks stationary at all scales, and almost all SiZer maps show purple, which confirms that the second power wavelet coefficients of FGN are not different from white noise. This is a great advantage compared to applying dependent SiZer because it does not require the specification of a model to the original time series and the estimation of parameters even if the original series has long–range dependence properties. Thus, we can use the regression model in (2.3) by slightly modifying it:

$$f(d_{j,k}) = g(k) + \sigma_j \epsilon_{j,k}, \quad k \in Z, \tag{3.1}$$

where σ_j is the standard deviation of $f(d_{j,k})$ for all $j = 1, \ldots, [\log_2(N)]$.

We typically use $f(x) = x^2$ in this paper because it captured the hidden nonstationary behavior very well in the examples we considered. This choice is also comparable to the wavelet spectrum which also involves the squared wavelet coefficients. Sometimes the fourth power of the wavelet coefficients can be useful because it captures very fine scaling behavior of the time series. An example will be presented in Figure 11 in Section 4.

In SiNos, the raw wavelet coefficients, $\{d_{j,k}\}$, are used for Variance SiNos, which is related to the second power wavelet SiZer. SiNos was also applied to the wavelet coefficients of a FGN series. To save space, the plots are not presented in this paper, but almost no features were found. This indicates that SiNos and SiZer methods flag no features beyond those naturally generated by a FGN process.

Figure 7 about here.

However, we do not recommend using the raw wavelet coefficients (first power) for wavelet SiZer, or Mean SiNos. Figure 7 shows the first power wavelet SiZer of the Thu1300 time series. Although we observed strong local nonstationary behavior for this time series in Section 1, this plot does not show any indication of nonstationarities. To save space, the plots for Mean SiNos are not presented in this paper, but the results are similar, that is they do not show significant features.

16

The last issue in wavelet coefficient based scale-space methods is the choice of wavelets. Daubechies wavelets are usually used because they have compact support and good regularity properties. We use three zero moments in this paper. Different wavelets with different zero moments result in slight differences in the wavelet SiZer and Variance SiNos analyses, but they do not change the main conclusions we draw from the plots. To save space, we do not report the results in this paper, but they are posted at http://www-dirt.cs.unc.edu/net_lrd/wavesizer.html with many different packet counts traces.

4   Real data analysis

4.1   Wavelet SiZer and Variance SiNos results

The Thu1300 time series was used to illustrate the effectiveness of wavelet SiZer and Variance SiNos in Section 1. In this subsection, we analyze additional traces.

Figure 8 about here.

The top left plot in Figure 8 displays a time series of packet counts in Internet traffic coming into the UNC, which were measured every 10 millisecond at the link of UNC on April 12, Saturday, 2003, from 3 p.m. to 4 p.m. (Sat1500). This time series can be written as $Y = \{Y(i),\ i = 1, 2, \ldots, 366000\}$. The series looks rather flat and has less spikes than the Thu1300 time series in Figure 1 (a). The rest of the plot shows the second power wavelet SiZer of the Sat1500 time series. It shows increasing trends which correspond to blue colors from $j = 1$ to $j = 5$, and a dominant spike extending across scales $j = 5$ to $j = 10$.

Figure 9 about here.

To explore the possible generators of the phenomena discovered by wavelet SiZer in Figure 8, we generated a simulated example, which is displayed in Figure 9. The top left plot in Figure 9 shows the simulated time series. The underlying trend is a sine curve with amplitude increasing in time with two short lower frequency, which is overlayed in white. The time series is this trend plus FGN with $H = 0.8$. The trend function is given by

$$g(x_i) = (1 + 0.1x_i)\sin(0.2\pi N x_i) + \sin(0.02\pi N x_i)\,I(0.5 < x_i \le 0.507)$$
$$+ \sin(0.1\pi N x_i)\,I(0.927 < x_i \le 0.933),$$

17

where $x_i = i/N$ for $i = 1, \ldots, N$ and $N = 300000$ is the length of the simulated series. In other words, the simulated data has a sinusoidal component with increasing amplitude and two coarse scale bursts around $x_i = 0.5$ and $x_i = 0.93$. The increasing amplitude creates increasing trends in the wavelet sizer, which are present at the scale $j = 2$ and 3. Since these variations have high frequency, their effect appears at fine scales. The effect of the two bursts in the trend appears as two spikes at the scales $j = 4$ and 6 in the wavelet SiZer. Since the first burst has lower frequency, it appears at the coarser scale $j = 6$, and the second burst has higher frequency, it appears at the finer scale $j = 4$. Thus, the increasing fine scale trends in Figure 8 could be explained by this type of increasing magnitude periodic component and the spikes in the wavelet SiZer could be caused by local bursts with particular frequencies that match the scale locations.

Figure 10 about here.

Going back to Figure 8, note that the SiZer map does not flag some large spikes as significant, as observed in Section 1. Variance SiNos is better than SiZer at detecting apparent peaks at large scales for this data set as well. In particular, as illustrated in Figure 10, Variance SiNos shows that the peak around $k = 9500$ for scale $j = 5$ is significant. A similar situation (not shown here) occurs for scale $j = 9$ where again Variance SiNos detects the peak while SiZer does not.

Figure 11 about here.

In Figures 8 and 10, the spike around $x = 3000$ (seconds) has a similar pattern as the Thu1300 time series in Figure 5. The first red window in Figure 11 (a) corresponds to the spike in Figure 8 lasting from $j = 5$ to $j = 9$. Similar to Figure 5, two drops are observed in the window and these are displayed in Figure 11 (b). Thus, we confirm that this kind of drop causes a spike in the second power wavelet SiZer.

Sometimes, other choices of the function f in (3.1) can provide additional insight. Figure 11 (c) shows the SiZer plot of the scale $j = 1$ with the fourth power, i.e. $f(x) = x^4$, in wavelet SiZer. Compared to the second power wavelet SiZer in Figure 8, it clearly shows that there is a spike at the end of the series at $j = 1$. The second red window in Figure 11 (a) corresponds to this spike and Figure 11 (d) displays the zoomed series corresponding to that window. An apparent burst behavior is observed in this window. It is of very short duration and is now flagged as significant because the size of its variation is magnified by taking the

18

fourth power.

We analyzed many different packet count traces collected from the UNC link in 2002 and 2003, but we do not report the other results here to save space. However, all results are displayed and summarized at http://www-dirt.cs.unc.edu/net_lrd/wavesizer.html.

4.2 Summary graphic

Figure 12 about here.

Finally, we propose a new visualization tool for combining inference into a single plot. We have seen from the presented figures that a major problem with this methodology is that the user has to analyze and interpret numerous plots for each time series under study. This can, to some extent, be mitigated by the following approach. In Figure 12, we present a compressed plot of the information in Figure 3. The upper panel is, as for SiZer, a family plot of the smooths. In the lower panel we have, however, tried to summarize the important information of the feature maps in Figure 3. The horizontal axis shows time while the vertical axis represents the ten scales $j = 1$ to $j = 10$. For one particular scale and time, e.g. $j = 1$ and a specific location on the horizontal axis, we plot the number of times a significant feature appears for this time location in the feature map for $j = 1$ in Figure 3. If there are many (few) significant features for this scale and time in Figure 3, a dark (bright) point is plotted. By examining the plot in Figure 12, we see that it tells much of the same story as we obtain from all the maps in Figure 3. Of course, it does not give a detailed description of what is going on, but it certainly indicates that there are local nonstationarities for many scales and time points. This plot reveals that it is worthwhile to create and view the detailed plots shown in Figure 3 to learn more about the underlying structure in this data set. We have also applied this idea to the FGN data set used in Figure 6. The compressed plot (not shown here to save space) was in this case essentially white, indicating that it is not necessary to create and view all the feature maps presented in Figure 6. The idea of this summary plot approach is simply to disregard automatically those data sets with no statistically significant behavior. This is an important issue in situations where many data sets are to be analyzed. In addition, it gives an informative snap shot of the times and scales where potential nonstationarities are present.

5   Discussion

Our main aim in this paper has been to develop tools that, in an objective way, use the wavelet spectrum and wavelet coefficients to detect important features in a time series.

We have seen that the wavelet spectrum can give helpful information about the scales where unusual behavior appears. Combined with dependent SiZer, it also shows, to some extent, where in the time series the nonstationarities take place.

We have adapted two existing scale-space methods to the problem of objectively detecting significant features in (functions of) the wavelet coefficients. An important lesson is that most of the features found show up in the squared wavelet coefficients. For FGN both SiZer and SiNos are seen to do a good job since none of the methods detect any features from the wavelet coefficients. In the Sat1500 data set, the two methods give comparable results with the exception that SiNos detects some peaks that SiZer does not detect. The same phenomenon is observed for some peaks in the Thu1300 data set. We believe that this is due to the way the present version of SiZer estimates the variance of the test statistic. Preliminary results (see Hannig and Lee (2003)), using a robust estimator for the variance of this test statistic, seem promising in the sense that peaks become easier to detect. In our future research, we plan to develop a robust SiZer version and test how it performs on the data sets used in this paper.

An advantage of SiZer is that it is computationally faster than SiNos. For long time series, this is an important issue. We view this as additional motivation for developing a robust version of SiZer in future research.

References

ABRY, P. AND VEITCH, D (1998). Wavelet analysis of long–range dependent traffic. *IEEE Transactions on Information Theory* **44**, 2–15.

ABRY, P., FLANDRIN, P., TAQQU, M. S. AND 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., LANG, G., MOULINES, E. AND 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. AND TAQQU, M. (2002). Semi-parametric 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. 527–556.

BERAN, J. (1994). *Statistics for long-memory processes.* Chapman & Hall, London.

CHAUDHURI, P. AND MARRON, J. S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association* **94**, 807–823.

DANG, T.D. AND MOLNÁR, S. (1999). On the effects of non-stationarity in long–range dependence tests. *Peridica Polytechnica Series on Electrical Engineering*, **43(4)**, 227–250

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets.* SIAM,

DOUKHAN, P., OPPENHEIM, G., AND TAQQU M. S. (2003). *Theory and Applications of Long–range Dependence.* Birkhäuser.

FAN, J., AND GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall, London.

FELDMANN, A., GILBERT, A. C., HUANG, P. AND WILLINGER, W. (1999). Dynamics of IP traffic: A study of the role of variability and the impact of control. *Proceedings of the ACM/SIGCOMM '99,* Boston, MA, 301–313.

GIRAITIS, L., KOKOSZKA, P. AND LEIPUS, R. (2001) Testing for long memory in the presence of a general trend. *Journal of Applied Probability,* **38(4)**, 1033–1054.

HANNIG, J. AND LEE, T. (2003). Improved SiZer for robust regression structure exploration and outlier identification. Unpublished manuscript.

HANNIG, J. AND MARRON, J. S. (2003). Advanced Distribution Theory for SiZer. *Submitted.*

HERNANDEZ CAMPOS, F., LE, L., MARRON, J. S., PARK, C., PARK, J. ,PIPIRAS, V., SMITH F. D., SMITH, R. L., TROVERO, M., AND ZHU, Z. (2004). Long–range dependence analysis of Internet traffic. In preparation.

LELAND, W., TAQQU, M., WILLINGER, W., AND WILSON, D. (1994). On the Self-Similar Nature of Ethernet Traffic. *IEEE/ACM Transactions on Networking* **2**, 1–15.

LINDEBERG, T. (1994). *Scale Space Theory in Computer Vision.* Kluwer, Boston.

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

MANDELBROT, B. B. (1988). *The Fractal Geometry of Nature.* W. H. Freeman and Co., New York, New York.

MANDELBROT, B. AND VAN NESS, J. (1968). Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, **10(4)**, 422–437.

MIKOSCH, T. AND SĂRICĂ, C. (2000). Change of structure in financial time series, long–range dependence and the GARCH model. *Center for Analytical Finance, University of Aarhus, Denmark*. Working paper series No 58.

OLSEN, L. R., AND GODTLIEBSEN, F. (2003). A scale-space approach for detecting changes in statistical behavior of dependent data. Unpublished manuscript.

PARK, C., MARRON, J. S. AND RONDONOTTI, V. (2004). Dependent SiZer: goodness of fit tests for time series models. Unpublished manuscript.

PIPIRAS, V., TAQQU, M. S. & ABRY, P. (2001). Asymptotic normality for wavelet-based estimators of fractional stable motion. Unpublished manuscript.

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

RONDONOTTI, V. AND MARRON, J. S. (2001). SiZer for time series: a new approach to the analysis of trends. Unpublished manuscript.

SHUMWAY, R. H., AND STOFFER, D. S. (2000). *Time Series Analysis and Its Applications.* Springer Text in Statistics, New York.

STOEV, S., TAQQU, M., PARK, C. AND MARRON, J. S. (2004). Strengths and limitations of the wavelet spectrum of the wavelet spectrum method in the analysis of Internet traffic. *Preprint.*

STOEV, S., TAQQU, M., PARK, C., MICHAILIDIS, G. AND MARRON, J. S. (2004). LASS: a tool for the local analysis of self-similarity. *Preprint.*

TAQQU, M. AND TEVEROVSKY, V. (1997). Robustness of Whittle-type estimates for time series with long–range dependence. *Stochastic Models.* **13**, 723–757.

TEVEROVSKY, V. AND TAQQU, M. (1997). Testing for long–range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *Journal of Time Series Analysis.* **18**, 279–304.

VEITCH, D. AND 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., AND ABRY, P. (2000). Meaningful MRA initialisation for discrete time series. *Signal Processing.* **80**, 1971–1983.

VEITCH, D., ABRY, P. AND TAQQU, M. (2003). On the automatic selection of the onset of scaling. *Fractals.* **11**, 373–390
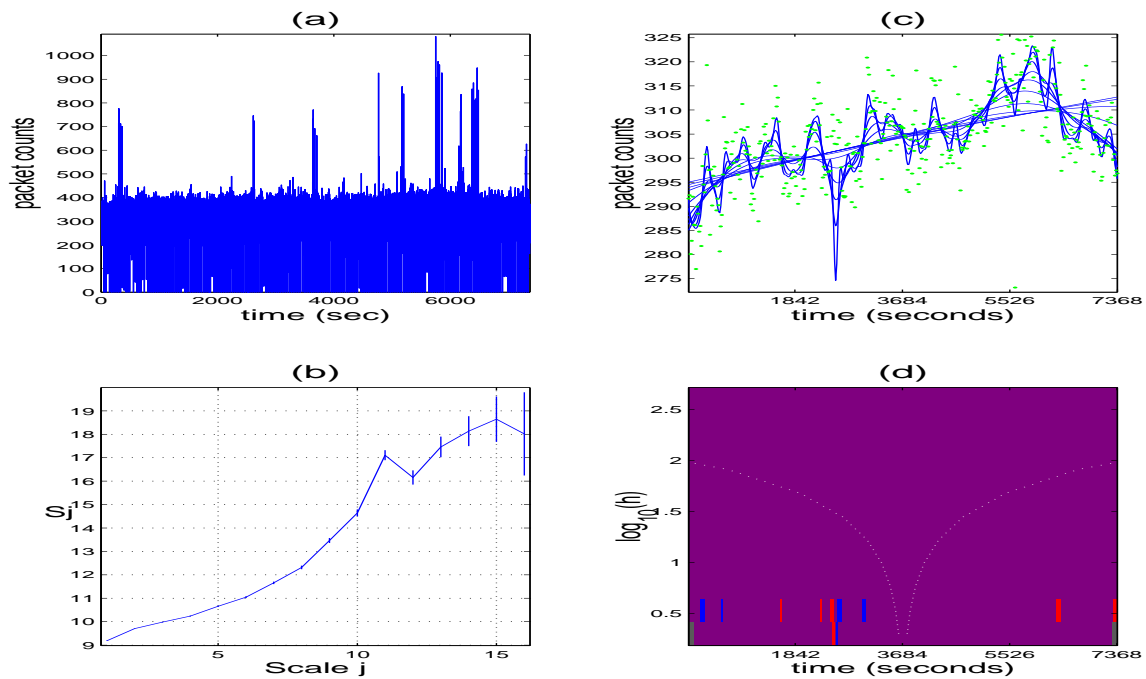
Figure 1: (a) Time series plot of packet counts measured every 10 milliseconds at the link of University of North Carolina, Chapel Hill (UNC) on April 11, Thursday, from 1 p.m. to 3 p.m., 2002 (Thu1300). Several spikes are shooting up and down. (b) The wavelet spectrum of the Thu1300 time series. There is a bump at the scale $j = 11$. (c) and (d) display the dependent SiZer of the Thu1300 time series.
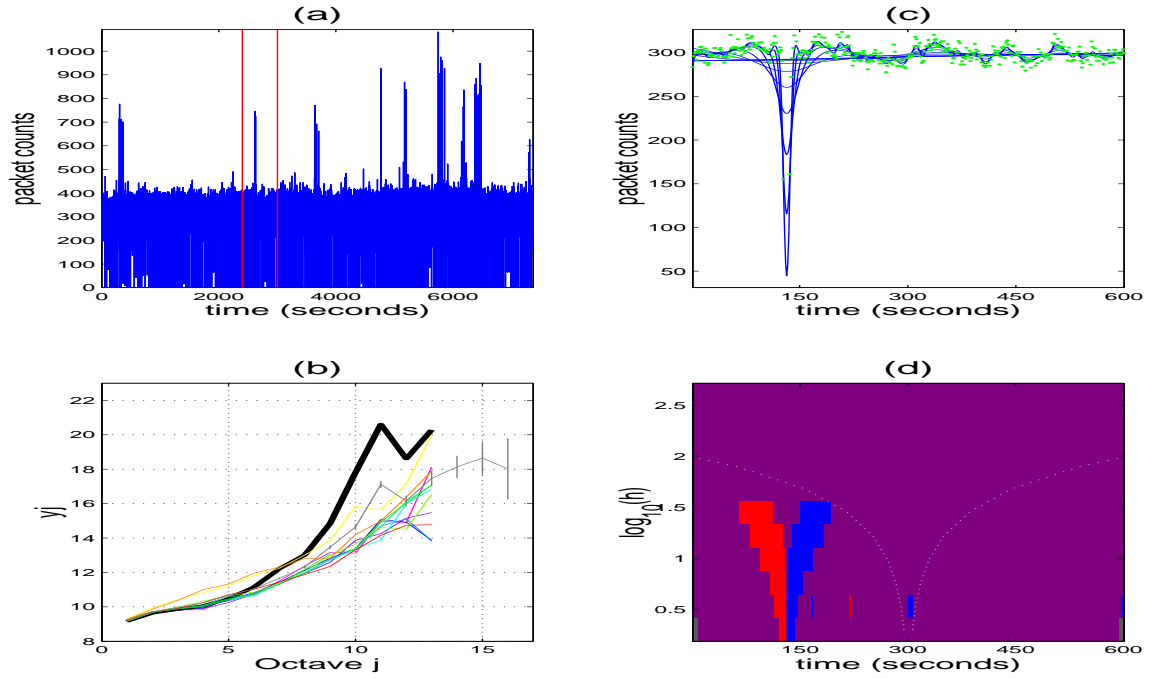
23

Figure 2: (a) Time series plot of the Thu1300 time series. The red vertical lines are the boundaries of the 600 second window between $x = 2400$ and $x = 3000$ (seconds). (b) Overlays of the wavelet spectra for each 600 seconds window. The spectrum with vertical lines (gray) corresponds to the full time series, which is the same as (b) in Figure 1. (c) and (d) display the dependent SiZer plot of the 600 second window between the red lines in (a). The subtrace is tested against FGN with $H = 0.9$ and the variance 485.6.
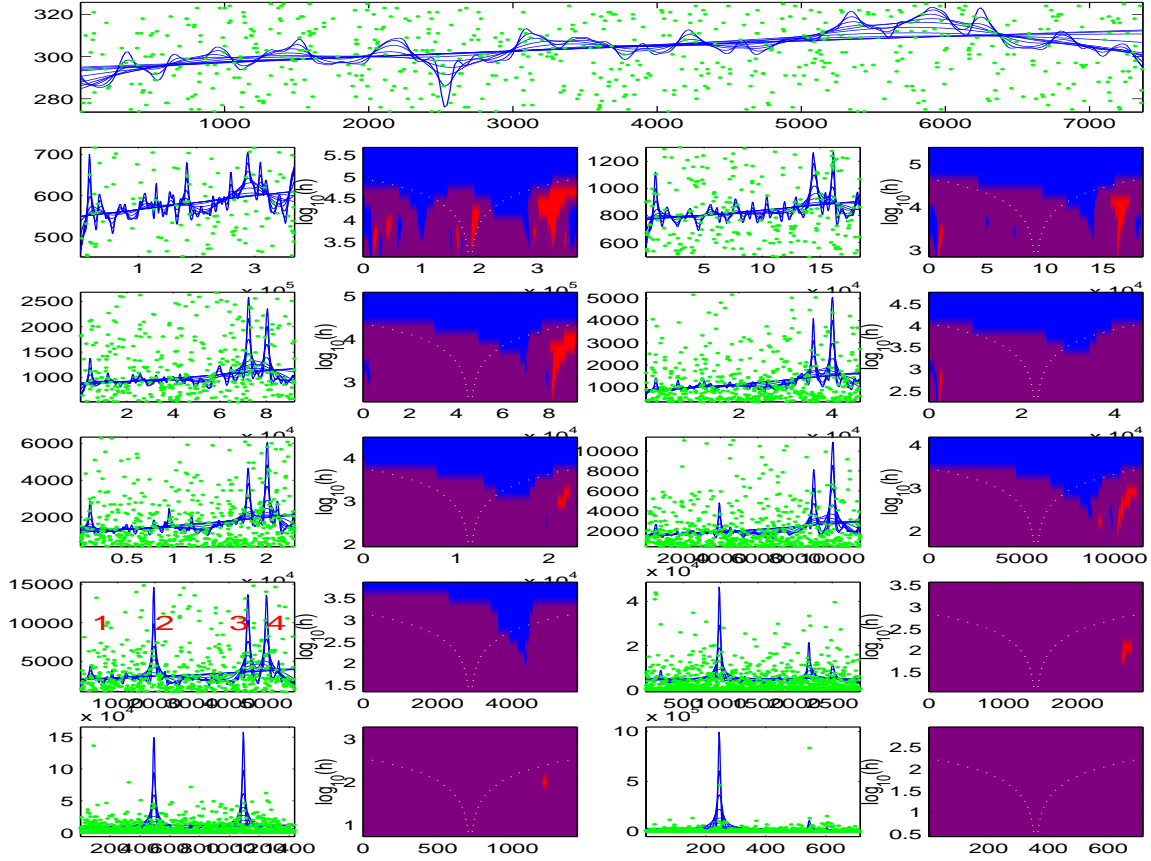
Figure 3: Wavelet SiZer plots of the Thu1300 time series. The top panel shows the family of smooths of the time series. The second row displays the family of smooths and the SiZer map of the second power wavelet coefficients at the scales $j = 1$ and $j = 2$. In the same way, the other rows display SiZer plots of the second power wavelet coefficients from $j = 3$ to $j = 10$. The SiZer plots at different scales are mostly dominated by four spikes, which are numbered at the $j = 7$ family of smooths.
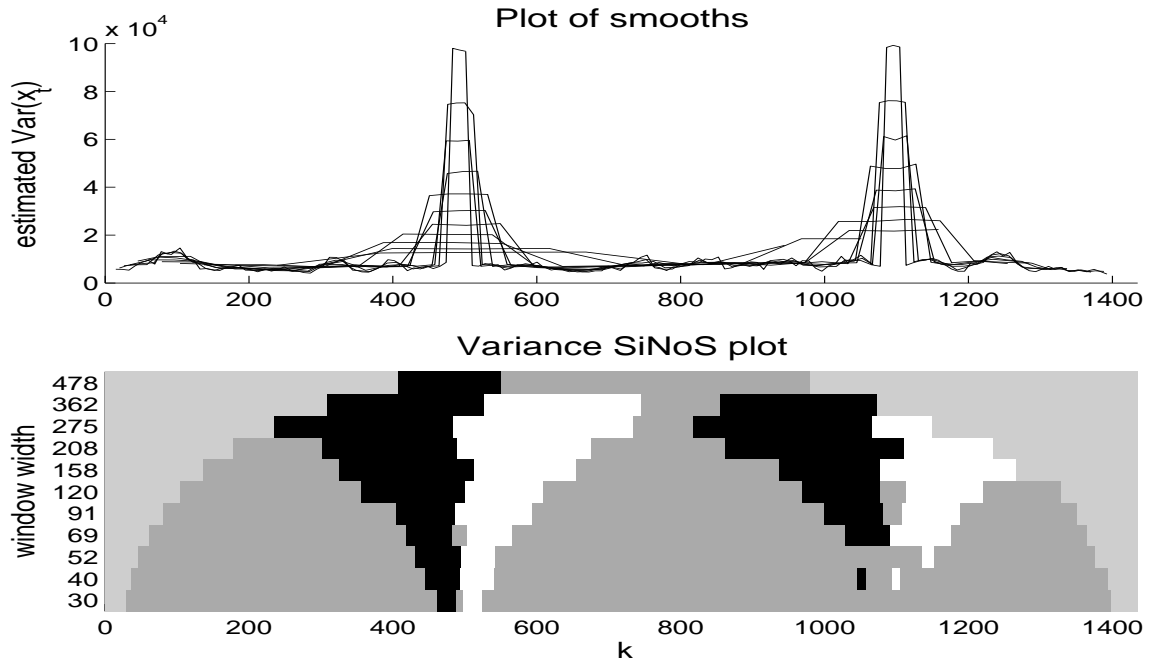
Figure 4: Upper panel is a family plot of Variance SiNos for scale $j = 9$ of the wavelet coefficients obtained from the Thu1300 data. The curves show estimated variance of the wavelet coefficients for various levels of smoothing. In the lower panel, a feature map showing significant changes in the variance of the wavelet coefficients, is given. Black and white means increasing and decreasing variance respectively. Gray is used to indicate that no significant changes are found while light gray is used in areas where no inference is being done. Combining the information in the family plot and the feature map, we conclude that there are significant changes in the variance at $k = 500$ and $k = 1100$ for this scale of the wavelet coefficients.
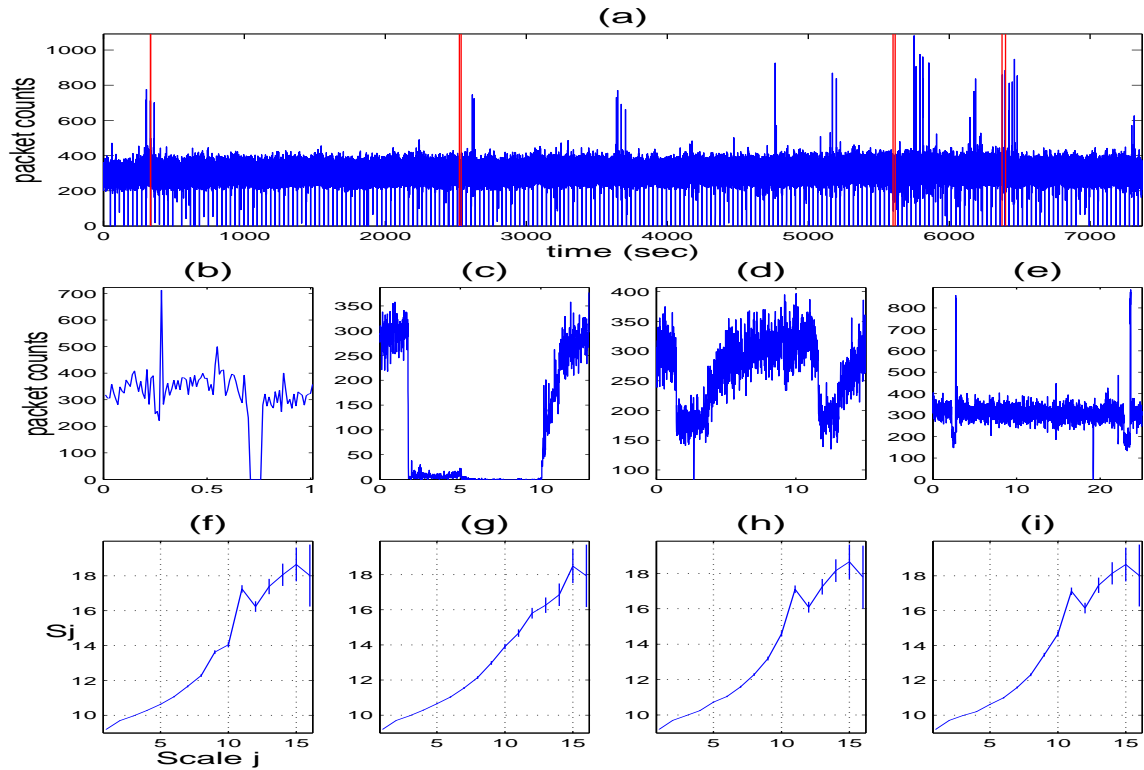
Figure 5: (a) Time series plot of the Thu1300 time series. The four intervals hightlighted by the vertical red lines correspond to the significant spikes of the wavelet SiZer in Figure 3. (b), (c), (d), and (e) show the subtraces corresponding to the four red windows in (a), respectively. (f), (g), (h), and (i) are corresponding wavelet spectra when each red window is excluded from the full time series and the remaining two parts are concatenated. This shows that the $j = 11$ bump in the spectrum was caused by the dropout shown in (c).
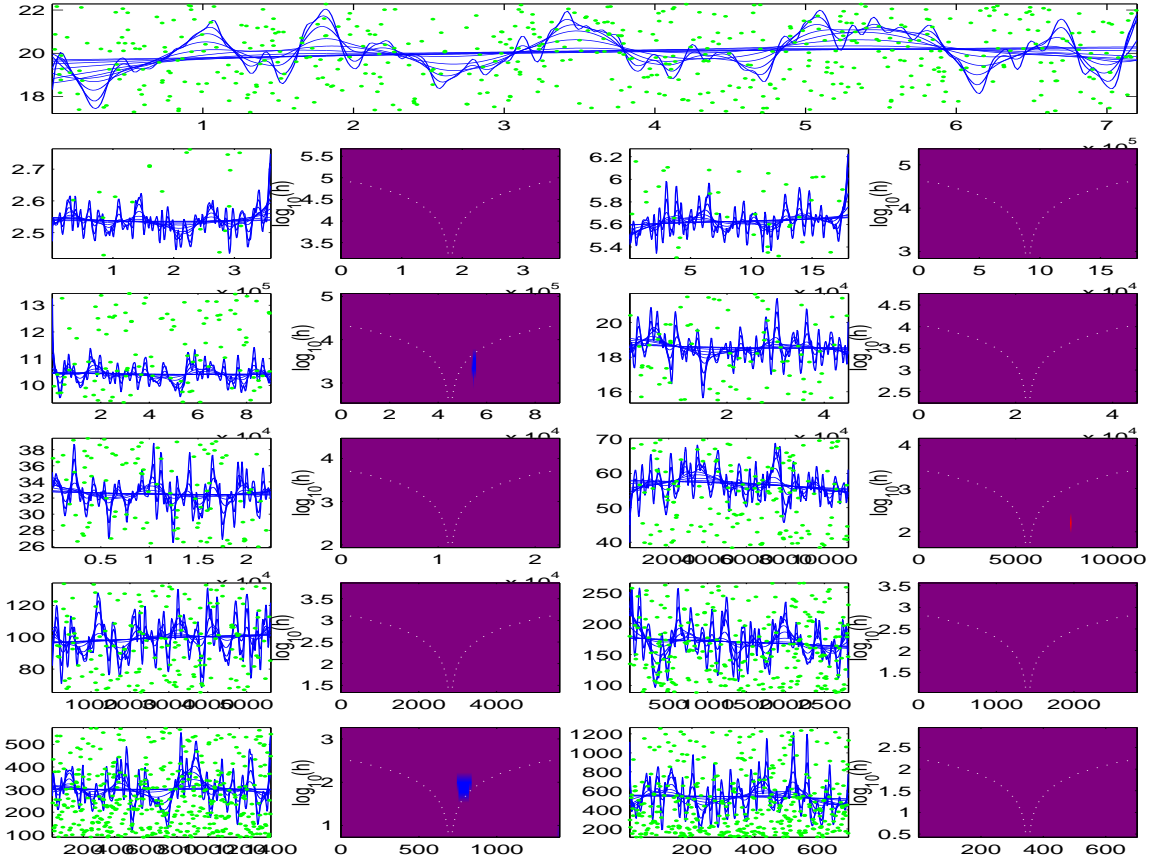
Figure 6: Wavelet SiZer plots of a simulated FGN with $H = 0.9$. At all scales, the shapes of the family of smooths look similar. Almost all SiZer maps show purple, which means the second power wavelet coefficients are not different from the white noise.
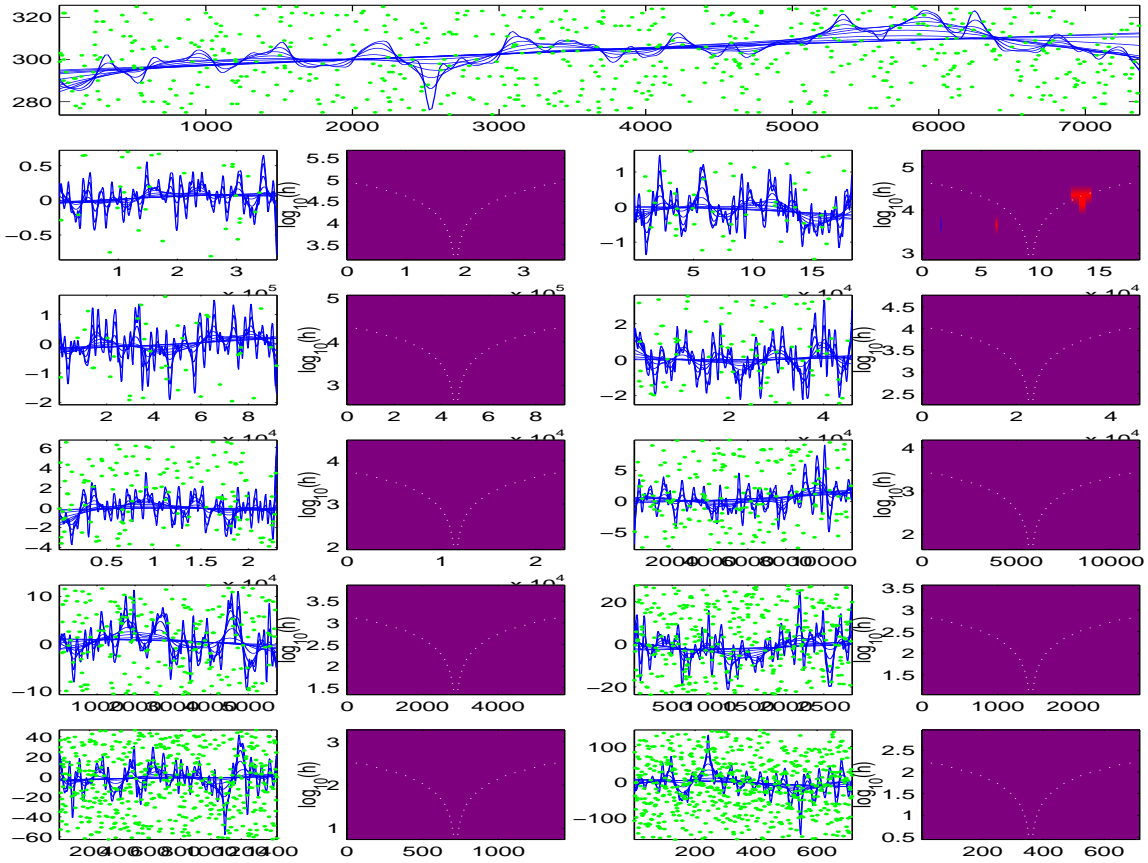
Figure 7: Wavelet SiZer plots of the Thu1300 time series with the first power wavelet coefficients. No significant features are found.
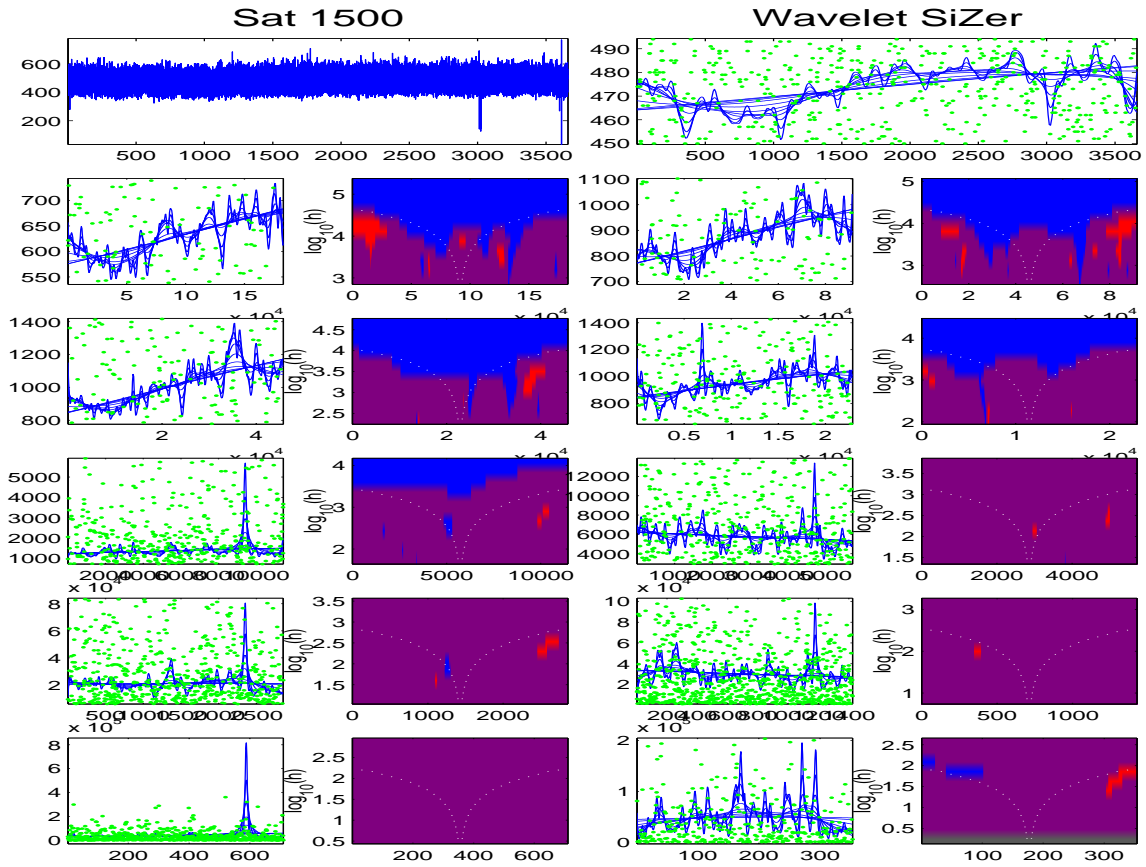
Figure 8: Wavelet SiZer plots of packet counts measured every 10 millisecond at the link of University of North Carolina, Chapel Hill (UNC) on April 12, Saturday, from 3 p.m. to 4 p.m. (Sat1500), 2003. The left of the top panel displays the original time series and the right displays its family of smooths. The wavelet SiZer shows increasing trends which correspond to blue colors from $j = 1$ to $j = 5$, and a spike from $j = 5$ to $j = 10$.
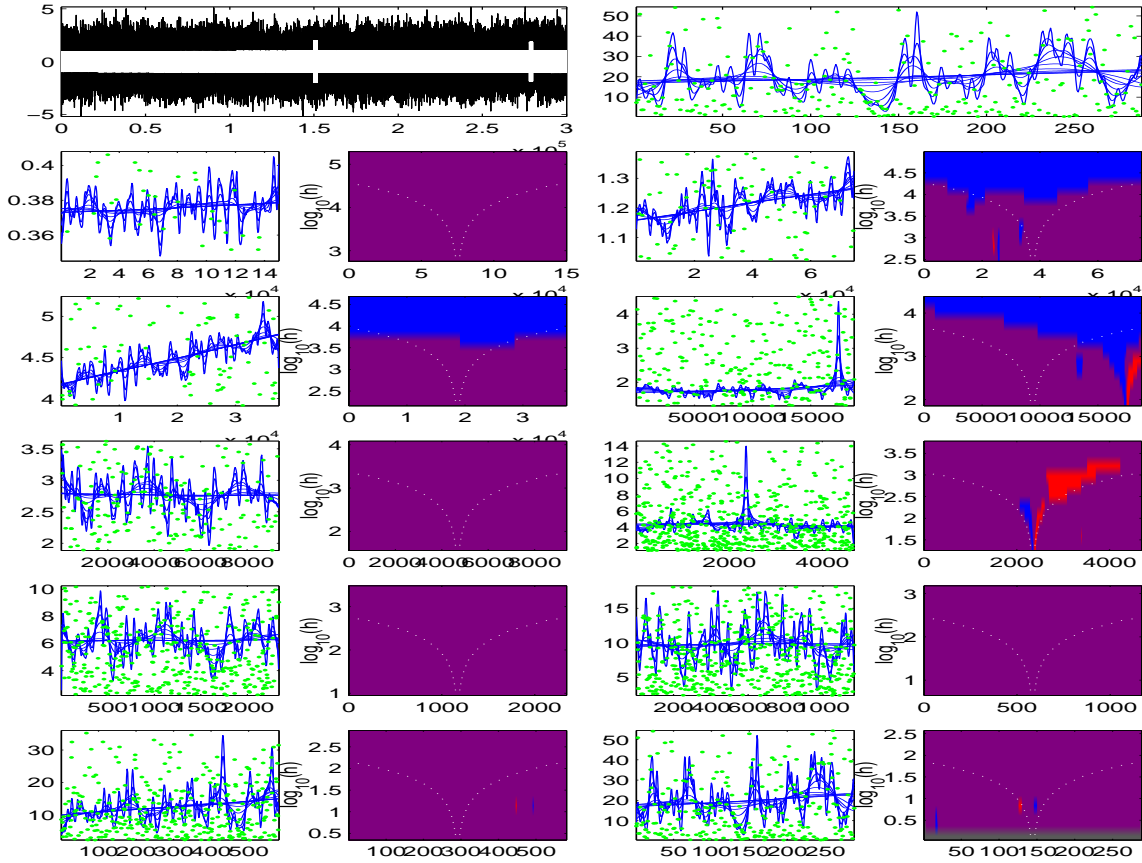
Figure 9: The wavelet SiZer of the simulated data. The left of the top panel displays the simulated time series. The sine curve with increasing variations and two twists is added to FGN with $H = 0.8$. This deterministic curve is overlayed in white. The simulated data create increasing trends at finer scales ($j = 2$ and 3) and two spikes at coarser scales ($j = 4$ and 6.
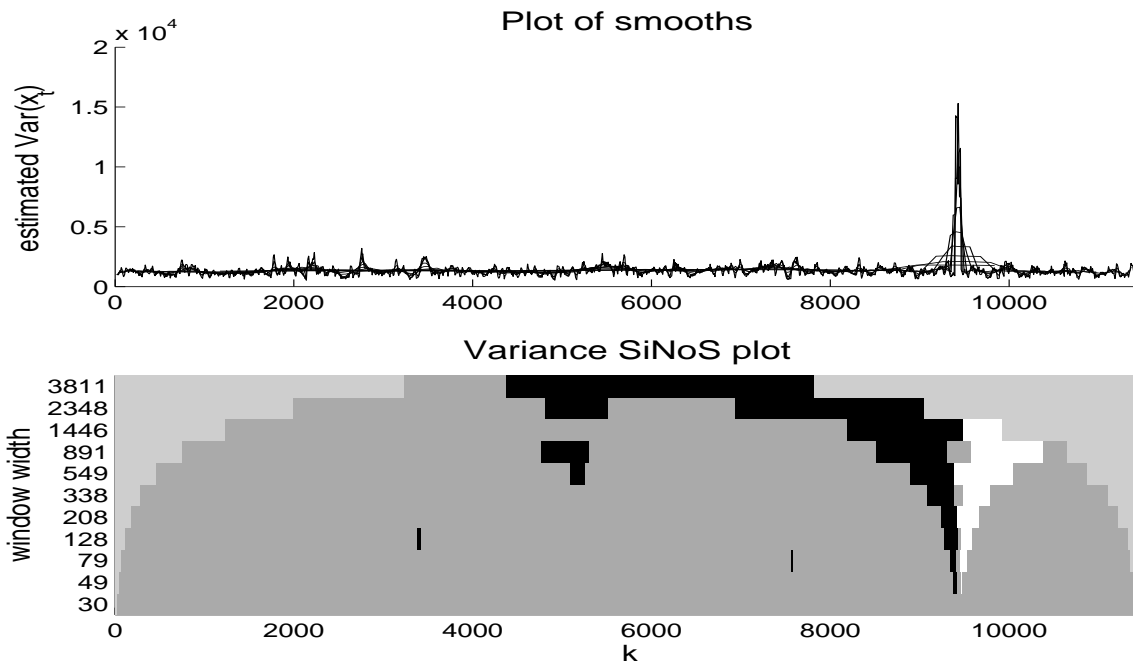
Figure 10: The upper panel is a family plot of Variance SiNos for scale $j = 5$ of the wavelet coefficient obtained from the Sat1500 data. In the lower panel a feature map, showing significant changes in the variance of the wavelet coefficients, is given. Note that a peak around $k = 9500$ is detected for a broad range of window widths.
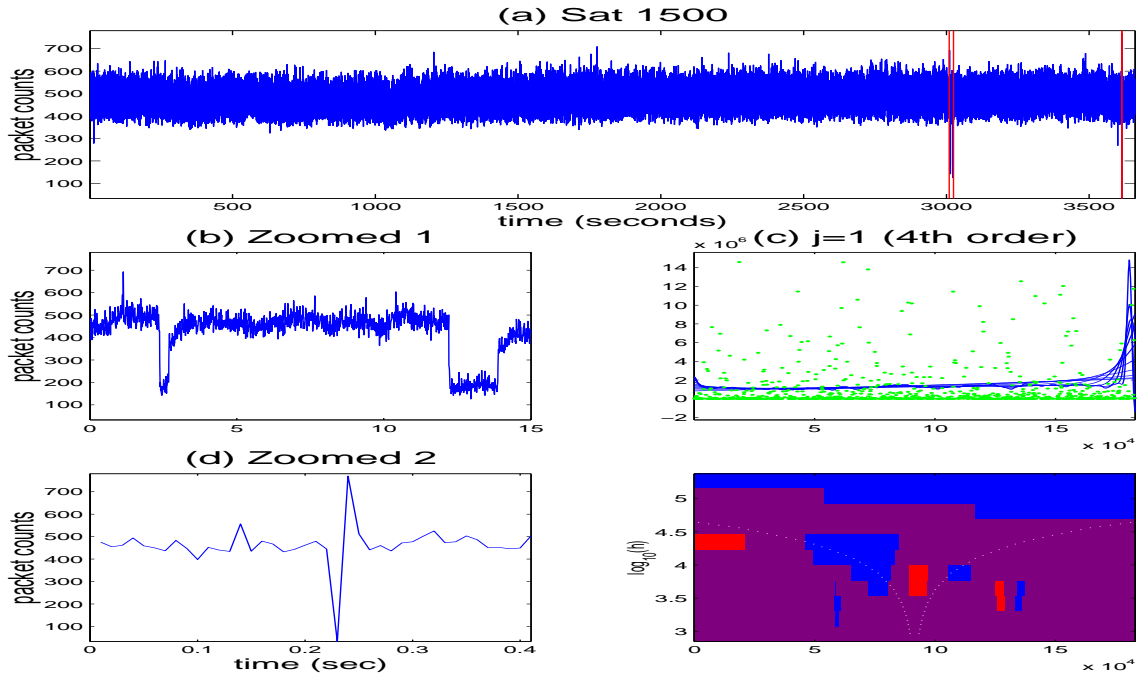
Figure 11: (a) Time series of the Sat1500 time series. The first red window corresponds to the spike in wavelet SiZer in Figure 8. (b) shows the subtrace corresponding to the first red window in (a). There are two drops, which is similar to a pattern in the Thu1300 time series. (c) shows the SiZer plot of scale $j = 1$ of the fourth power wavelet coefficients. There is a spike at the end and this corresponds to the second red window in (a). (d) shows the subtrace corresponding to the second red window in (a). A short burst is observed.
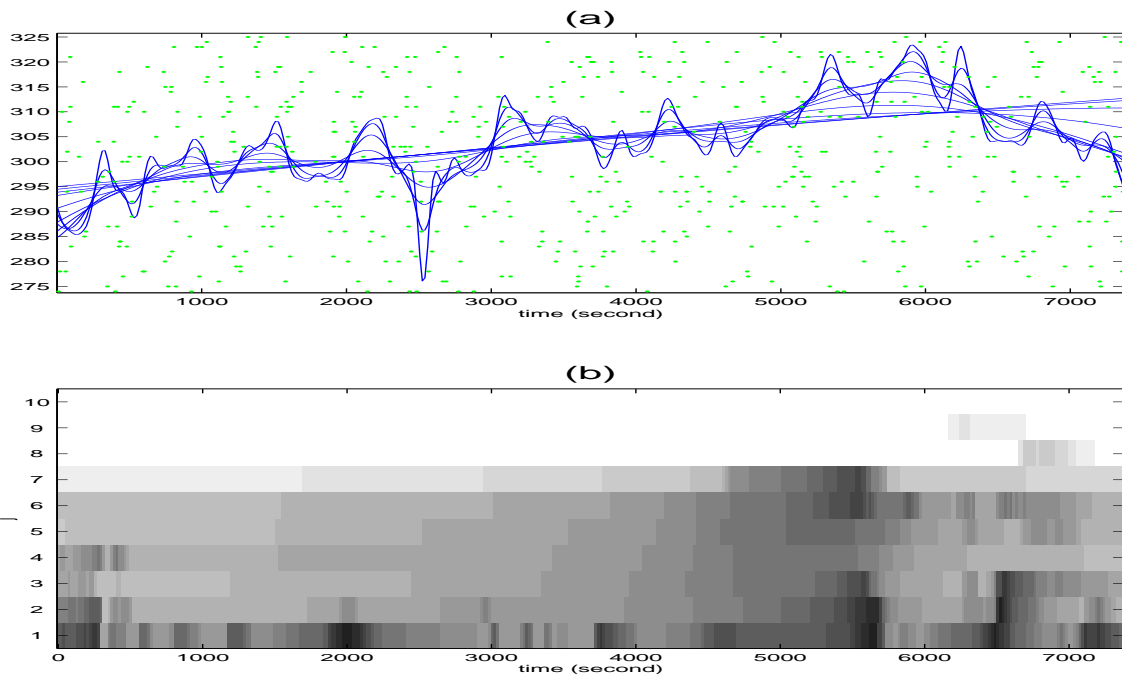
Figure 12: The upper panel is a family plot of the observed time series. The lower plot is a feature map summary for scales $j = 1, 2, \ldots, 10$ and time points. See text for more explanation about interpretation of this feature map.

34