Centrum voor Wiskunde en Informatica
REPORT*RAPPORT*

Outlier Detection and Localisation with Wavelet Based Multifractal Formalism

Z.R. Struzik, A.P.J.M. Siebes

Information Systems (INS)

INS-R0008 February 29, 2000

Report INS-R0008
ISSN 1386-3681

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

Outlier Detection and Localisation with Wavelet Based Multifractal Formalism

Zbigniew R. Struzik, Arno Siebes
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

ABSTRACT
We present a method of detecting and localising outliers in stochastic processes. The method checks the internal consistency of the scaling behaviour of the process within the paradigm of the multifractal spectrum. Deviation from the expected spectrum is interpreted as the potential presence of outliers. The detection part of the method is then supplemented by the localisation analysis part, using the local scaling properties of the time series. Localised outliers can then be removed one by one, with the possibility of dynamic verification of spectral properties. Both the multifractal spectrum formalism and the local scaling properties of the time series are implemented on the wavelet transform modulus maxima tree.

2000 Mathematics Subject Classification: 80-08, 82-08, 92-08
1999 ACM Computing Classification System: H.1.m, I.m, J.2, J.3
Keywords and Phrases: multifractal analysis, wavelet transform, Hölder exponent, outlier detection
Note: This work has been carried out under the Impact project.

1. Introduction

All real life stochastic signals are subject to contamination by noise. This can be relatively low level random noise or some systematic bias. But it can also be a sudden shoot off from the current value of the sensoric information - the outlier. The main difference between the noise or bias in the traditional sense and an outlier, which of course is also 'noise', is the inherently isolated and local character of the outlier. In addition, outliers will usually have high amplitudes and generally there will be relatively few of them. Therefore, non-stationarity and a highly erratic character is the main characteristic which constrasts with stationary (white) noise or linear or polynomial bias.

The outliers which are most often considered are of the isolated spike type. These are single points - Dirac delta type events, relatively far from the current 'expected' or true value of the process.[1] These may happen at random intervals and have arbitrary amplitude. The other type of outlier event is a sequence of single Dirac delta events. We will refer to it as a 'burst' type event. This may be a record of random noise instead of the process or a superposition of random noise on the process. The last type of outlier event to be classified here is the Heaviside step type. The process may be interrupted suddenly and change to another process, often to come back to the initial process after a short while. Such an event will often be accompabied by a strong baseline shift, which allows detecting and classifying the entire alien process as an outlier.

In figure 1, we present the examples of the three classes just described. The reader may say, 'It is very easy to see outliers in these plots; one can determine them with the naked eye.' We agree. The human eye (or visual system) is an outstanding pattern recognition system, very good in detecting non-stationarities. It is capable of fitting models and performing approximations in real-time using various acquired knowledge. It is, however, also an adaptive system: different runs of the same experiment will likely result in different answers. Varied magnification, resolution or normalisation

[1] whichever of the two we are able to determine, and assuming the process possesses this kind of expectancy. For example a random process with infinite variance will not have such a property. However, here we are analysing real processes where distributions are bounded, as is variance.

will also most likely affect the results. In this report, we would like to provide some *objective* criteria which should also be suitable to be exercised automatically by the computer. We will not differentiate between the outlier types but will be able to detect all three types.

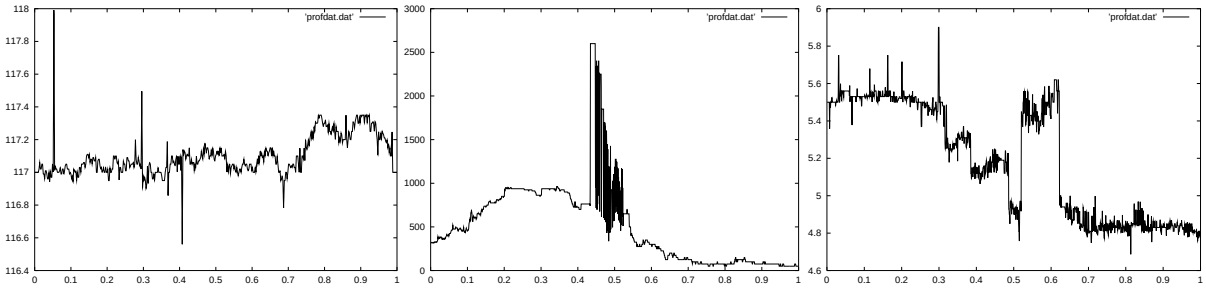

Figure 1: Left: 'First_Dirty' showing examples of outliers of the isolated Dirac delta spike type. Centre: 'Second_Dirty' showing an example of a burst type of outliers. Right: 'Third_Dirty' with an outlier consisting of a different process (baseline difference) between two Heaviside steps.

For this purpose we will need a methodology which is capable of determining the statistical nature of the non-stationary process in both a global and a local sense. The global sense will be primarily used to detect the presence of the outliers in the process analysed. The local properties will next be used to identity the offenders. This methodology stems from the wavelet based multifractal analysis first introduced by Arneodo, Bacry and Muzy [2].

We will first introduce the wavelet transform, paying special attention to detection and analysis of singular events. Then we will briefly describe the global partition function based multifractal formalism, showing how it can be applied to test the statistical integrity of the analysed process. Next we will introduce the local version of multifractal singularity strength analysis, again showing how the statistical integrity of the analysed process can be tested, now in a local fashion, allowing not only the detection but also the localisation of outliers.

Section 6 closes the paper with conclusions and suggestions for future developments.

2. Continuous Wavelet Transform and its Maxima Used to Reveal the Structure of Singularities in the Time Series

Conceptually, the wavelet transformation [5, 6] is a convolution product of the time series with the scaled and translated kernel - the wavelet $\psi(x)$, usually a $n-th$ derivative of a smoothing kernel $\theta(x)$. Usually, in the absence of other criteria, the preferred choice is the kernel, which is well localised both in frequency and position. In this paper, we chose the Gaussian $\theta(x) = \exp(-x^2/2)$ as the smoothing kernel, which has optimal localisation in both domains.

The scaling and translation actions are performed by two parameters; the scale parameter s 'adapts' the width of the wavelet kernel to the *microscopic resolution* required, thus changing its frequency contents, and the location of the analysing wavelet is determined by the parameter b:

$$Wf(s,b) = \frac{1}{s} \int_{-\infty}^{\infty} dx\ f(x)\ \psi(\frac{x-b}{s})\ ,$$

where $s, b \in \mathbf{R}$ and $s > 0$ for the continuous version (CWT).

The 3D plot in figure 2 shows how the wavelet transform reveals more and more detail while going towards smaller scales, i.e. towards smaller $\log(s)$ values. Therefore, the wavelet transform is sometimes referred to as the 'mathematical microscope', due to its ability to focus on weak transients and singularities in the time series. The wavelet used determines the optics of the microscope; its magnification varies with the scale factor s.

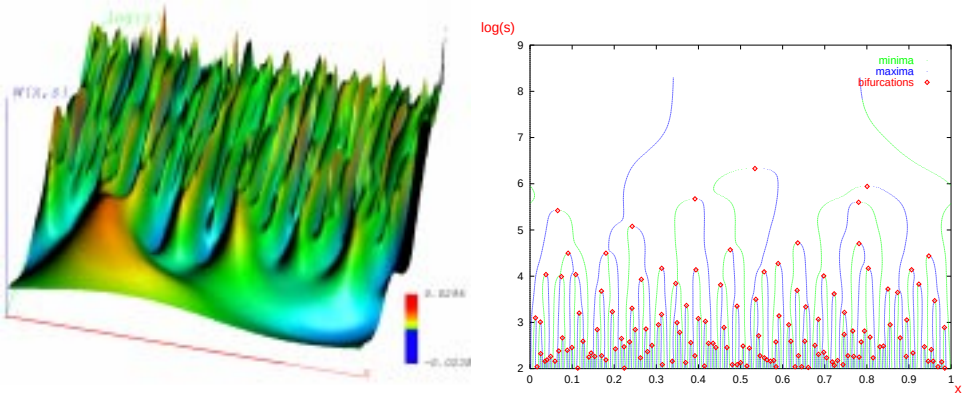

Figure 2: Left: Continuous Wavelet Transform representation of the random walk (Brownian process) time series. The wavelet used is the Mexican hat - the second derivative of the Gaussian kernel. The coordinate axes are: position x, scale in logarithm $\log(s)$, and the value of the transform $W(s,x)$. Right: the related WTMM representation.

It can be shown [7] that for cusp singularities, the location of the singularity can be detected, and the related exponent can be recovered from the scaling of the Wavelet Transform, along the so-called *maxima line*, converging towards the singularity. This is a line where the wavelet transform reaches local maximum (with respect to the position coordinate). Connecting such local maxima within the continuous wavelet transform 'landscape' gives rise to the entire tree of maxima lines. Restricting oneself to the collection of such maxima lines provides a particularly useful representation [8] (WTMM) of the entire CWT. It incorporates the main characteristics of the WT: the ability to reveal the *hierarchy* of (singular) features, including the scaling behaviour. [2] Restricting oneself to the collection of such maxima lines provides a particularly useful representation of the entire CWT. In particular, we have the following power law proportionality for the wavelet transform of the cusp singularity in $f(x_0)$: [2]

$$W^{(n)}f(s,x_0) \sim |s|^{h(x_0)} \ .$$

This is on condition that the wavelet has at least n vanishing moments, i.e. it is orthogonal to polynomials up to degree n: $\int_{-\infty}^{+\infty} x^m \, \psi(x)\, dx = 0 \quad \forall m,\ 0 \le m < n$.

Moreover, the wavelet transform and its WTMM representation can also be shown to be invariant with respect to the rescaling/renormalisation operation [4, 2, 12, 11]. This property makes it an ideal tool for revealing the renormalisation structure of the (hypothetical) multiplicative process underlying the analysed time series.

2.1 Small Scaling Example

Let us consider the following set of examples of simple singular structures, see figure 3 left; a single Dirac pulse at $D(1024)$, the saw tooth consisting of an integrated Heaviside step function at $I(2048)$, and the Heaviside step function for $S(3072^+)$, where + denotes the right-handed limit. The Hölder exponent of a Dirac pulse is -1 by definition. For Hölder singularities, the process of integration and

[2] One should bear in mind that the above relation is an approximate case for which exact theorems exist [9]. In particular, we will restrict the scope of this paper to Hölder singularities for which the local and point-wise Hölder exponents are equal [10]. Thus we will not take into consideration the 'oscillating singularities' (e.g $x^\alpha sin(1/x^\beta)$) requiring two exponents [9, 11]. Nevertheless, it is sufficient for our purpose to state that the continuous wavelet transform can be used for characterising the Hölder singularities in the time series even if masked by the polynomial bias.

differentiation respectively adds and subtracts one from the exponent. We, therefore, have $h = 0$ for the right-sided step function $S(3072^+)$ and $h = 1$ for the integrated step $I(2048)$.

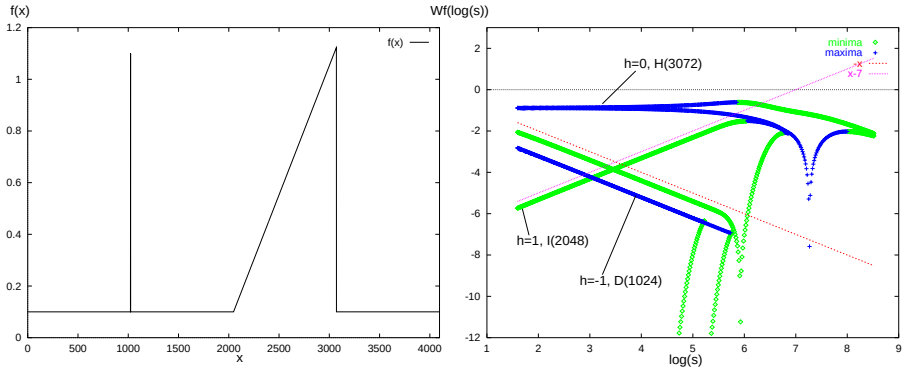

Figure 3: Left: the test signal consisting of the Dirac pulse $D(1024)$, the change in slope - integrated Heaviside step I(2048), and the Heaviside step H(3072). Right: the log-log plot of the maxima, together with their respective logarithmic derivatives, corresponding to all three singularities: $D(1024)$, I(2048) and H(3072). Lines of theoretical slope are also indicated; these are $-x$ for $D(1024)$, x for I(2048) and a constant for H(3072). The wavelet used is the Mexican hat.

These values can also be verified in the scaling of the corresponding maxima lines. We obtain the (logarithmic) slopes of the maxima values very closely following the correct values of these exponents, see figure 3 right. This, of course, suggests the possibility of the estimation of the Hölder exponent of (Hölder) singularities from the slope of the maxima lines approaching these singularities. An important limitation is, however, the requirement for the singularities to be *isolated* for this procedure to work. Note that the scaling of the maxima lines becomes stable in the log-log plot in figure 3 right only below some critical scale s_{crit}, below which the singularities effectively become isolated for the analysing wavelet. Indeed, the distance between the singular features in the test time series in figure 3 left equals 1024, which is in the order of three standard deviations of the analysing wavelet at $(\log(s_{crit}) = 5.83 = \log(1024/3)$. This example largely simplifies the issue since the singular structures are of the same size, resulting in one characteristic scale at which they appear in the wavelet transform. Also, generally, the scaling of the maxima lines for other than the presented simple examples will not follow a straight line even for isolated singularities. Still, the rate of decrease of (the supremum of) the related wavelet transform maximum will be consistent, thus allowing estimation of h.

3. Multifractal Formalism on the WTMM Tree

The WTMM tree lends itself very well for defining the partition function based multifractal formalism (MF) [2]. The MF takes the moments q of the measure distributed on the WTMM tree to obtain the dependence of the scaling function $\tau(q)$ on the moments q:

$$\mathcal{Z}(s,q) \sim s^{\tau(q)}.$$

The $\mathcal{Z}(s,q)$ is the partition function of the q-th moment of the measure distributed over the wavelet transform maxima at the scale s considered:

$$\mathcal{Z}(s,q) = \sum_{\Omega(s)} (W f \omega_i(s))^q \ , \tag{3.1}$$

where $\Omega(s) = \{\omega_i(s)\}$ is the set of all maxima $\omega_i(s)$ at the scale s, satisfying the constraint on their local logarithmic derivative in scale [13]. (The local slope bound used throughout this paper is $|\breve{h}| \leq 2$.)

Intuitively, since the moment q has the ability to select a desired range of values: small for $q < 0$, or large for for $q > 0$, the scaling function $\tau(q)$ globally captures the distribution of the exponents $h(x)$ - weak exponents are addressed with large negative q, while strong exponents are suppressed and effectively filtered out. For the large positive q, the opposite takes place (and strong exponents are addressed while week exponents are effectively filtered out).

This dependence may be linear, indicating that there is only one class of singular structures and related exponent, or it may have a slope non-linearly changing with q. In the latter case, the local tangent slope to $\tau(q_*)$ will give the corresponding exponent, i.e. $h(q_*)$, with its related dimension marked on the ordinate axis $C = D(h(q_*))$, where $\tau(q_*) = h(q_*)q_* + C$. The set of values C, i.e. dimensions $D(h(q_*))$ for each value of h selected with q_* is the so-called spectrum of the singularities $D(h)$ of the fractal signal. Formally, the transformation from $\tau(q)$ to $D(h)$ is referred to as the Legendre transformation:

$$\begin{aligned} \frac{d\tau(q)}{dq} &= h(q) \ , \\ D((h(q)) &= q\ h(q) - \tau(q) \ . \end{aligned} \tag{3.2}$$

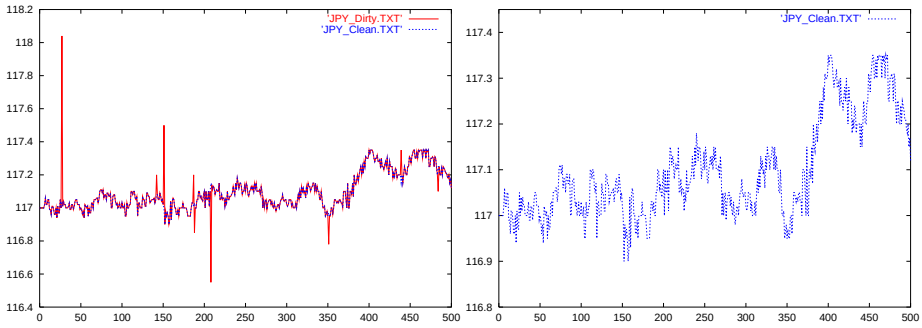

Figure 4: Left: 'First_Dirty'. Right: 'First_Clean'. The only difference between the two time series is a number of erroneous spikes which in 'First_Clean' were localised using external information and were removed by hand. The plot of 'First_Clean' still shows spikes which belong to the process investigated. The task of the methodology to be described is to detect the presence of erroneous spikes in time series and provide the means of localising them. The renormalised plot of 'First_Clean' reveals its complex form. Spiky events are still present in the time series, but these belong to the process.

From this transformation, we can directly obtain expressions for the average $h(q)$ and $D((h(q))$ in terms of the partition function over maxima values. From 3.2 and 3.1 we have:

$$h(q) = \frac{d\tau(q)}{dq} = \lim_{s\to 0} \frac{d}{dq} \frac{\log(\mathcal{Z}(s,q))}{\log(s)} = \lim_{s\to 0} \frac{1}{\log(s)} \sum_{\Omega(s)} P(s,q,\omega_i(s))\ \log(Wf\omega_i(s)) \tag{3.3}$$

where

$$P(s,q,\omega_i(s)) = \frac{(Wf\omega_i(s))^q}{\sum_{\Omega(s)}(Wf\omega_i(s))^q} \tag{3.4}$$

is the weighting measure for the statistical ensemble $\Omega(s)$ [1]. Similarly we get the expression for D(h(q)):

$$D(h(q)) = \lim_{s\to 0} \frac{1}{\log(s)} \sum_{\Omega(s)} p(s,q,\omega_i(s)) \ \log(p(s,q,\omega_i(s))) \ . \tag{3.5}$$

Usually this pair is used to obtain $D((h(q))$ spectrum in a parametric form (q is a parameter here). We will, however, also show $D(q)$ and $h(q)$ separately, since the dependence on q is crucial for our purpose.

In figure 5, we show $D(h(q))$ evaluated for the 'Dirty' time series from figure 1 and for its cleaned version where the outlier spikes were removed using external information, see figure 4.

The difference in the spectral information is striking and well reflects the high sensitivity of the partition function method to outliers. The spectrum for the clean version is narrow and focused around the main value of singularity strength $h_{mean} = 0.4$. For the dirty version, we have a very broad spectrum which gradually falls off to zero dimension values for decreasing $h < h_{mean}$. This fall off regime corresponds with positive q values which have the ability to select exponents of a relatively lower value than the h_{mean} value. As with the example time series, we are operating in the $h = 0.4$ range, capturing spikes for which $h = -1$ (in an isolated situation) is not a problem.

Let us immediately remark that processes exist which have a wide range of h among their characteristic. Such processes are generically called *multifractal*, as a fractal dimension $D(h)$ is associated with each h. Hence, if there is a multitude of meaningful values $D(h_*)$ (associated with some h_*) constituting the spectrum of the process, the process can be of multifractal type. The difference between such processes and the processes with outliers is in the relative values of $D(q)$ and the spacing between the successive q's. For the process to have a meaningful multifractal spectrum, we will require dense coverage of $h(q)$ values. Also the $D(q)$ values should be relatively large for meaningful spectra. Dense support on a line corresponds with dimension 1. Single points, that is separated point-wise events, on the other hand have support 0. Therefore, if $D(q)$ is near 0, this indicates very weakly supported events and therefore a high probability of an outlier.

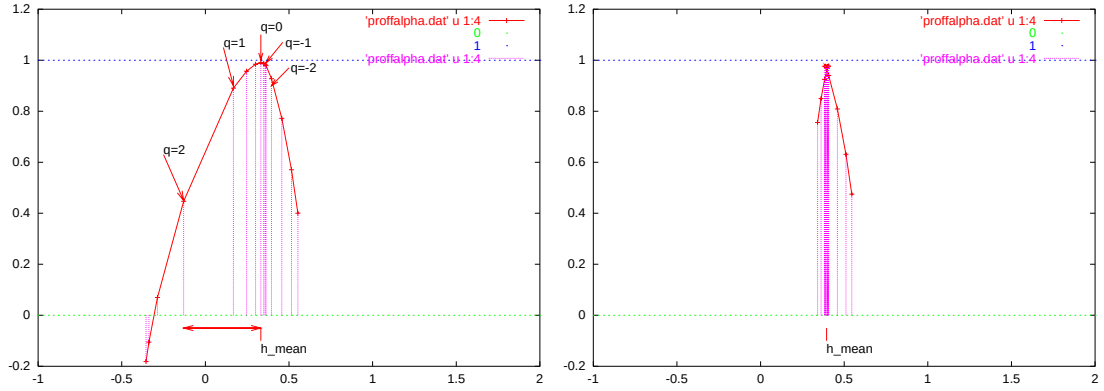

Figure 5: Left: $D(h(q))$ evaluated for the 'First_Dirty' time series from figure 1. Right: the same for the 'Clean' version of the same time series. (The outlier spikes were removed using external information.) A clear difference in D spectrum is visible.

In figures 6 and 8, we check $D(q)$ and $h(q)$ separately in order to analyse the dependence on q. The test data in figures 6 and 8 respectively is contaminated with two types of outliers: Dirac delta type and Heaviside step type respectively. For comparison we also analyse outlier-free time series. In both cases, for both types of outliers $h(q)$ evaluated from Eq. 3.3, $h(q)$ shows strong a crossover and for $q > 1$ it quickly falls off from the average h_{mean} value. Also in the case of $D(q)$ evaluated from Eq. 3.5 for the contaminated time series, a clear difference in behaviour is visible. $D(q)$ quickly approaches 0

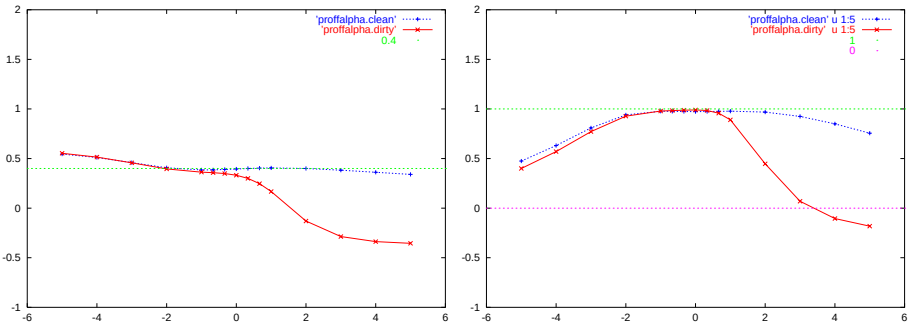

Figure 6: Left: $h(q)$ evaluated from 3.3 for both 'Clean' and 'Dirty' time series from fig 1. h shows strong crossover and for $q > 1$ quickly falls off from the average 0.4 value for this process. Right: $D(q)$ evaluated from 3.5 for the same time series. Again a clear difference in behaviour is visible. D quickly approaches 0 for q larger than 1.

for q larger than 1.

The conclusion that we can derive from these test results is that comparing the value of both the $h(q)$ and $D(q)$ for positive q's may be useful for detecting the presence of spikes in the time series. In particular the second moment $q = 2$ seems to be suitable for use as a criterium in comparison with the reference $q = 0$ moment. The rule of thumb which we suggest is that if the value of $h(q = 2)$ differs from $h(q = 0)$ by some 0.5 and if $D(q = 2)$ is about 0.5 or less, the probability of spike presence is relatively high.

For actual localisation of spikes, we need the local value of $h(x)$ instead of the global average as defined in Eq. 3.3. Such a local $h(x)$ will make possible separating outliers from the residue, using some a threshold value for h. This will be discussed further in the next section.

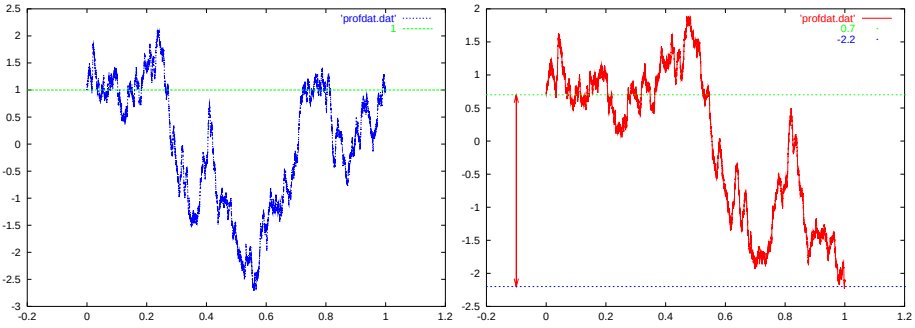

Figure 7: Left: fractional Brownian motion fBm with $H = 0.6$. This sample is cut out so that both ends meet when put together in a wrap around fashion. No 'end of the sample artifact' will be detected while processing. Right: the first half of the same fBm sample as in the left plot. When processed in a wrap around fashion this sample will have a strong jump indicated. The corresponding 'outlier' artifact will be detected.

4. Estimation of the Local, Effective Hölder Exponent Using the Multiplicative Cascade Model

Note that even though the partition function method (discussed thus far) uses the maxima tree containing full local information about the singularities, this is lost at the very moment the partition function is computed. Therefore, there is no explicit local information present in the scaling estimates;

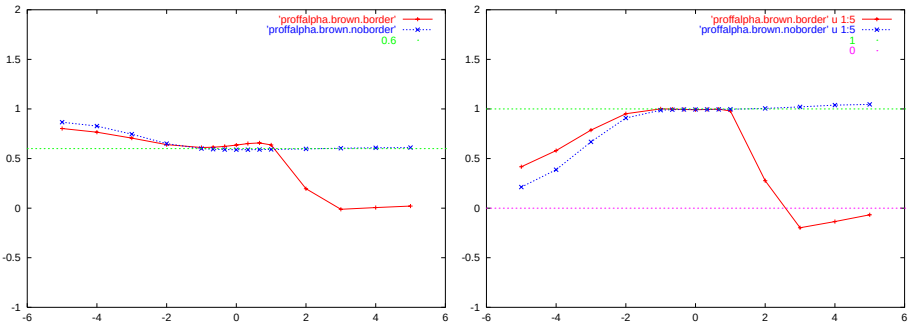

Figure 8: Left: $h(q)$ evaluated from 3.3 for both samples of fBm from the previous figure. For the sample with the Heaviside step outlier, h shows strong crossover and for $q > 1$ quickly falls off from the average 0.6 value for this process. Right: $D(q)$ evaluated from 3.5 for the same two samples. Again a clear difference in behaviour is visible. D quickly approaches 0 for q larger than 1.

τ, h or D, and all these are global statistical estimates. This is also where the strength of the partition function method lies - global averages are much more stable than local information and in some cases all that it is possible to obtain.

Indeed, it is generally not possible to obtain local estimates of the scaling behaviour other than in the case of isolated singular structures from the WT. A typical example of the evolution of the maximum line along scale is shown in figure 9. It is not possible to evaluate the slope of the plot, not even on the selected range of scales. This is why we introduced [3] an approach circumventing this problem while retaining local information - a local effective Hölder exponent, in which we model the singularities as created in some kind of a collective process of a very generic class.

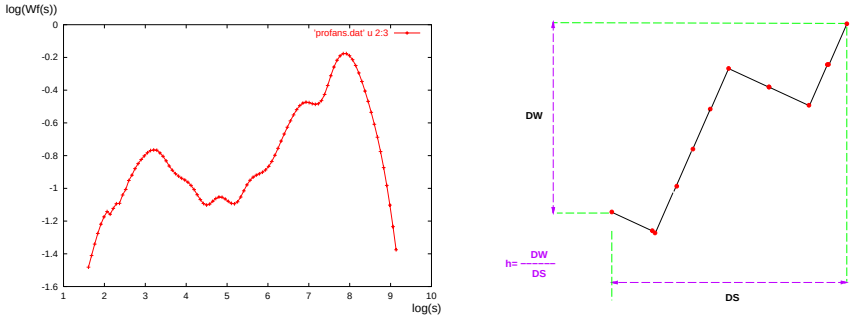

Figure 9: Left: it is impossible to evaluate the scaling exponent for an arbitrary maximum line participating in a complex process: a real file example of a maximum line. Right: the local effective Hölder exponent estimate takes the effective difference in the logarithm of the density of the process with respect to the logarithm of the scale difference gained along the process path.

While we can use the average $h(q)$ in globally detecting the presence of spikes in the time series, in order to localise them we need the local h estimate. The local effective h approach will make this possible.

We have shown in the previous section that the wavelet transform and in particular its maxima lines can be used in evaluating the Hölder exponent in isolated singularities. In most real life situations, however, the singularities in the time series are not isolated but densely packed. The logarithmic rate of increase or decay of the corresponding wavelet transform maximum line is usually not stable but fluctuates, following the action of the (hypothetical, multiplicative) process involved.

To capture the fluctuations and estimate the related exponents (to which we will refer as an *effective*

Hölder exponent of the singularity), we will model the singularities as created in some kind of a collective process of a very generic class - the multiplicative cascade model. Each point of this cascade is uniquely characterised by the sequence of weights $(s_1 ... s_n)$, taking values from the (binary) set $\{1, 2\}$, and acting successively along a unique process branch leading to this point. Suppose that we denote the density of the cascade at the generation level F_i (i running from 0 to max) by $\kappa(F_i)$, we then have

$$\kappa(F_{max}) = p_{s_1} \ldots p_{s_n} \; \kappa(F_0) = P_{F_0}^{F_{max}} \; \kappa(F_0)$$

and the local exponent is related to the rate of increase of the product $P_{F_0}^{F_{max}}$ over the gained scale difference. In any experimental situation, the weights p_i are not known and h has to be estimated. This can be simply done using the fact that for the multiplicative cascade process, the effective product of the weighting factors is reflected in the difference of logarithmic values of the densities at F_0 and F_{max} along the process branch:

$$h_{F_{max}}^{F_0} = \frac{\log(\kappa(F_{max})) - \log(\kappa(F_0))}{\log((1/2)^{max}) - \log((1/2)^0)} \; .$$

The densities along the process branch can be estimated with the wavelet transform, using its remarkable ability to reveal the entire process tree of a multiplicative process [4]. It can be shown that the densities $\kappa(F_i)$ corresponds with the value of the wavelet transform along the maxima lines belonging to the given process branch. The estimate of the effective Hölder exponent becomes:

$$\hat{h}_{s_{lo}}^{s_{hi}} = \frac{\log(Wf\omega_{pb}(s_{lo})) - \log(Wf\omega_{pb}(s_{hi}))}{\log(s_{lo}) - \log(s_{hi})} \; ,$$

where $Wf\omega_{pb}(s)$ is the value of the wavelet transform at the scale s, along the maximum line ω_{pb} corresponding to the given process branch. Scale s_{lo} corresponds with generation F_{max}, while s_{hi} corresponds with generation F_0, (simply the largest available scale in our case).

4.1 Estimation of the Mean Hölder Exponent

For a multiplicative cascade process, a mean value of the cascade at the scale s can be defined as:

$$\mathcal{M}(s) = \frac{\sum_{\Omega(s)} \log(Wf\omega_i(s)) \; ,}{\mathcal{Z}(s, 0)} \; , \tag{4.1}$$

where the $\mathcal{Z}(s, 0)$ is the partition function Eq. 3.1 for $q = 0$ and corresponds with the number of maxima at the scale s considered.

This mean is compatible with the canonical formalism based spectrum,[3] see Eq. 3.3, and gives the direct possibility of estimating the mean value of the local Hölder exponent as a linear fit to $\mathcal{M}$:

$$\log(\mathcal{M}(s)) = \bar{h} \log s + C \; . \tag{4.2}$$

Therefore, we estimate our mean Hölder exponent $\bar{h}$ from 4.2 by substituting $\mathcal{M}$ with $\mathcal{M}$. The estimate of the local Hölder exponent, from now on to be denoted as $\hat{h}(x_0, s)$ or just $\hat{h}$, now becomes:

$$\hat{h}_{s_{lo}}^{s_{SL}} \cong \frac{\log(Wf(s_{lo})) - (\bar{h} \log(s_{SL}) + C)}{\log(s_{lo}) - \log(s_{SL})} \; .$$

[3] In Ref. [3], we used a different, 'micro-canonical' mean which in general differs slightly from the canonical one. For reasons of the compatibility of the spectra, we take the canonical mean here.

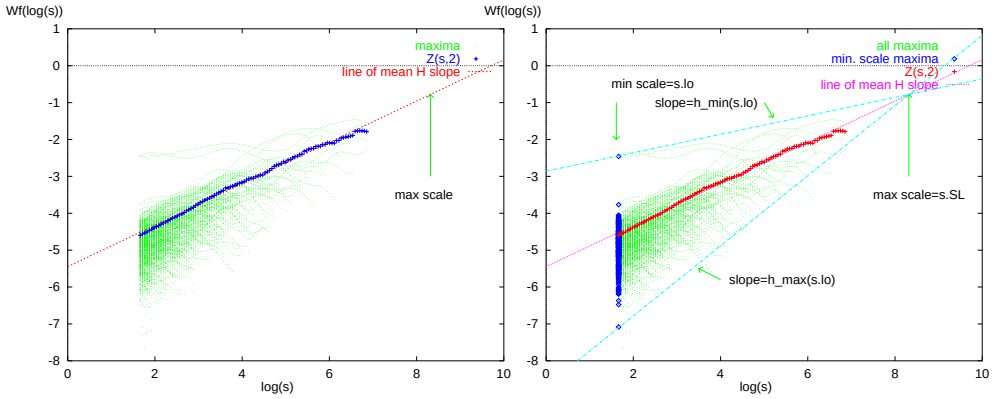

Figure 10: Left: the projection of the maxima lines of the WT along time. The mean value of the Hölder exponent can be estimated from the log-log slope of the line shown. Also, the beginning of the cascade at the maximum scale s_{hi} is indicated. Right: the maxima at the smallest scale considered are shown in the projection along time. The effective Hölder exponent can be evaluated for each point of the maximum line at s_{lo} scale. Two extremal exponent values are indicated, for minimum and maximum slope.

4.2 Employing the Effective Hölder Exponent in Local and Global Spectra Estimation Examples
Such an estimated local $\hat{h}(x,s)$ is a function of the same x parameter (time, position) as the analysed function $f(x)$, and can be analysed in a local fashion or histogrammed in order to study its distribution properties. In figure 11, we depict it in the temporal fashion for the time series from figure 7 - the computer generated sample of fractional Brownian motion with $H = 0.6$. The local $\hat{h}(x,s)$ plot shows almost 'monochromatic' behaviour, centred at $H = 0.6$. There are, however, two instances, at both ends of the time series, where the exponent drops significantly. These drops are caused by the border effect and indicate that the time series is corrupted by cutting away the earlier and later parts. Both cut-offs also show in the log-histogram[4] of the values of the $\hat{h}(x,s)$ as peaks separate from the entire (rather narrow) bulk of exponents.

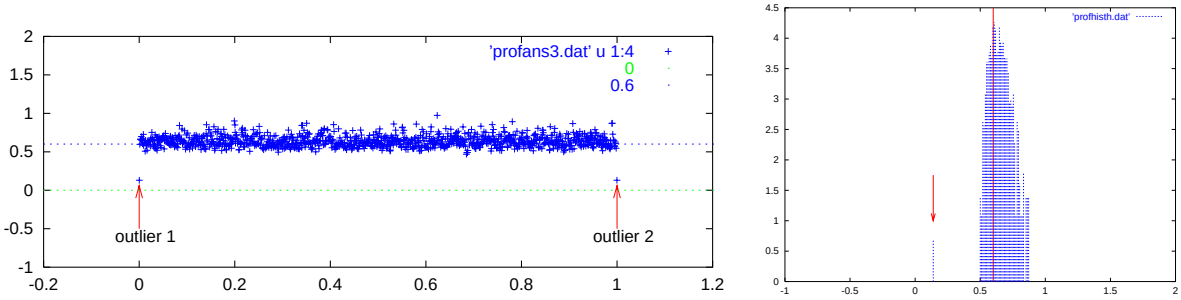

Figure 11: Left: local Hölder exponent for the time series from figure 7 right. Two ends of the sample outliers indicated. Right: the corresponding log-histograms of the local Hölder exponent (outliers indicated.)

The second example, in figure 12 is a record contaminated with the Dirac type events, see figure 5.

[4] They are made by taking the logarithm of the measure in each histogram bin. This conserves the monotonicity of the original histogram, but allows us to compare the log-histograms with the spectrum of singularities $D(h)$. By following the evolution of the log-histograms along scale, we will be able to extract the spectrum of the singularities $D(h)$.

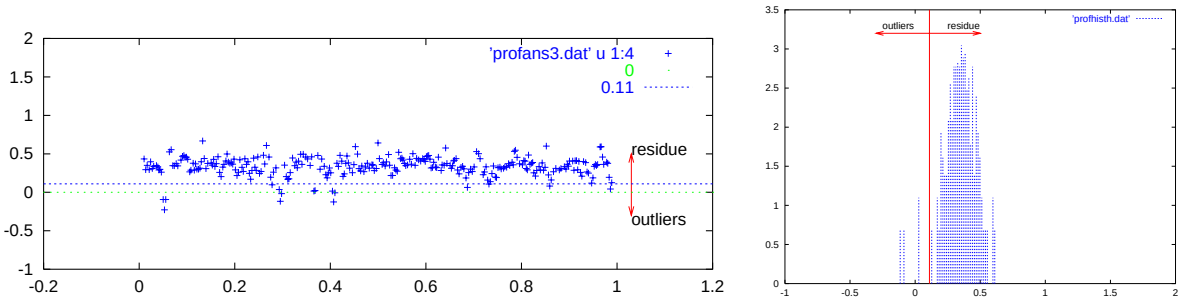

Figure 12: Left: local Hölder exponent for the time series from figure 5 left. Right: the corresponding log-histograms of the local Hölder exponent. Thresholding on h separates outliers from residue.

The local $\hat{h}(x, a = \log(10))$ plot is now more varied, but still clustered around the $h_{mean} = 0.4$. There are several 'drop down' events present, indicating the presence of strong singular events. These can be selected using an appropriate threshold on h.

Generic criteria for threshold choice can be obtained from the statistical distribution of the h exponent. For this purpose, the histogram of h can be analysed. For noisy time series, the corresponding histograms show considerable widening, often accompanied by visible fragmentation (discontinuity).

Determining the 'noiseless' width of the histogram and the point above/below which 'noise' starts is the crucial task in this procedure. The histogram can provide a good insight but it is somehow hard to automatize the choice of the threshold level and make it arbitrary. We have therefore resorted to a simple heuristic.

A good threshold value[5] which worked well in our tests (to follow in the experimental section) was determined using the scaling exponent of (the square root of) the second moment of the measure $\mathcal{Z}(s, 2)$:

$$\mathcal{M}'(s) = \sqrt{\frac{\mathcal{Z}(s,2)}{\mathcal{Z}(s,0)}}\,, \tag{4.3}$$

and the thresholding $\hat{h}$ exponent is then determined from the linear fit:

$$\log(\mathcal{M}'(s)) = \hat{h}\log(s) + C'\,. \tag{4.4}$$

In cases where there are outliers in the time series, this quantity, the 'micro-canonical' geometric mean, $\hat{h}$ does not coincide with the h_{mean} mode value of the $D(h)$ distribution. In such cases it can be used for the threshold between the outliers and the residue, as in figure 12. In fact, the $\hat{h}$ value can also be used as an additional criterium (besides that suggested in the previous section) for testing for the presence of outliers. In cases where the h_{mean} differs significantly from $\hat{h}$, the probability of outliers is high.

Finally, the histogram of the h exponent is shown in figure 12, with the tail to the lower h values clearly visible. The same threshold $\hat{h}$ as in the local plot separates the residual bulk and the tail outlier events.

[5] There is no 'absolute' criterium for the threshold value (at least we do not have one); the mean $\hat{h}$ is just one possibility which worked for us.

5. Experiments

We analysed the example time series from figure 1 for the presence of spikes of the sorts described in the introduction:

- Dirac delta type;
- Heaviside step type;
- Burst type.

The only preprocessing of the data was the linear interpolation and resampling into 4096 points. This was necessary in order to access small resolutions with the Mexican hat wavelet. The tests were performed with the entire length of the sample and the top scale of the maxima decomposition starting at $a = 1000$, i.e. one quarter of the (resampled) sample length. The lowest resolution was set to $a = 10$. We were thus able to access almost three decades of scaling. In real life, shorter scaling ranges will be required. The tests performed indicate that reducing the scaling range by half should not affect the resolution.

The local effective Hölder exponent h was extracted for all the time series at $a = 10$. Histograms of h were made for the two and the twenty lowest scales (in total we calculated 100 equidistant scales between $a = 10$ and $a = 1000$). For all noisy time series, the corresponding histograms show considerable widening, often accompanied by visible fragmentation. For the 'noise' thresholds shown in local h plots, we used the heuristic described in section 4 - whenever the second moment based mean $\hat{h}$ was substantially different from the mode h_{mean} of the h distribution, we took the value of $\hat{h}$ as the threshold.

$D(q)$ and $h(q)$ were also computed for each time series for all integers in the $q = -5..5$ range plus $\pm 1/3$ and $\pm 2/3$. All plots support our rule of thumb for outlier presence suggested in section 3. For specific results in each of the four cases presented, we refer the reader to the figure captions.

6. Conclusions

We have presented a method of detecting and localising outliers in stochastic processes. The method checks the internal consistency of the scaling spectrum of the process within the paradigm of the multifractal spectrum. Deviation from the expected spectrum is interpreted as the potential presence of outliers. We suggested a rule of thumb to detect the presence of outliers based on the global, partition function based multifractal formalism. The detection part is then supplemented by the localisation analysis part, using the local scaling properties of the time series. Each data point has an (effective) Hölder exponent associated with it, and based on some threshold level, the outliers can be separated from the residue of the exponent distribution. In particular, localised outliers can then be removed one by one, with the possibility of the dynamic verification of spectral properties (in a dynamic programming approach).

A possible extension to the methodology is in estimating the 'size' of outliers from the scale at which the associated maxima lines appear. Properties of the size distribution of all singular events could then be used with the related Hölder exponent distribution for outlier discrimination and localisation.

Another direction is the real time fast computation of time series local effective Hölder exponent and scaling properties with the use of the Haar wavelet. Testing for the likelihood of the current sample (that last acquired) being an outlier is crucial in some applications. Initial work on the Haar wavelet for this task has been done following the approach reported in [14].

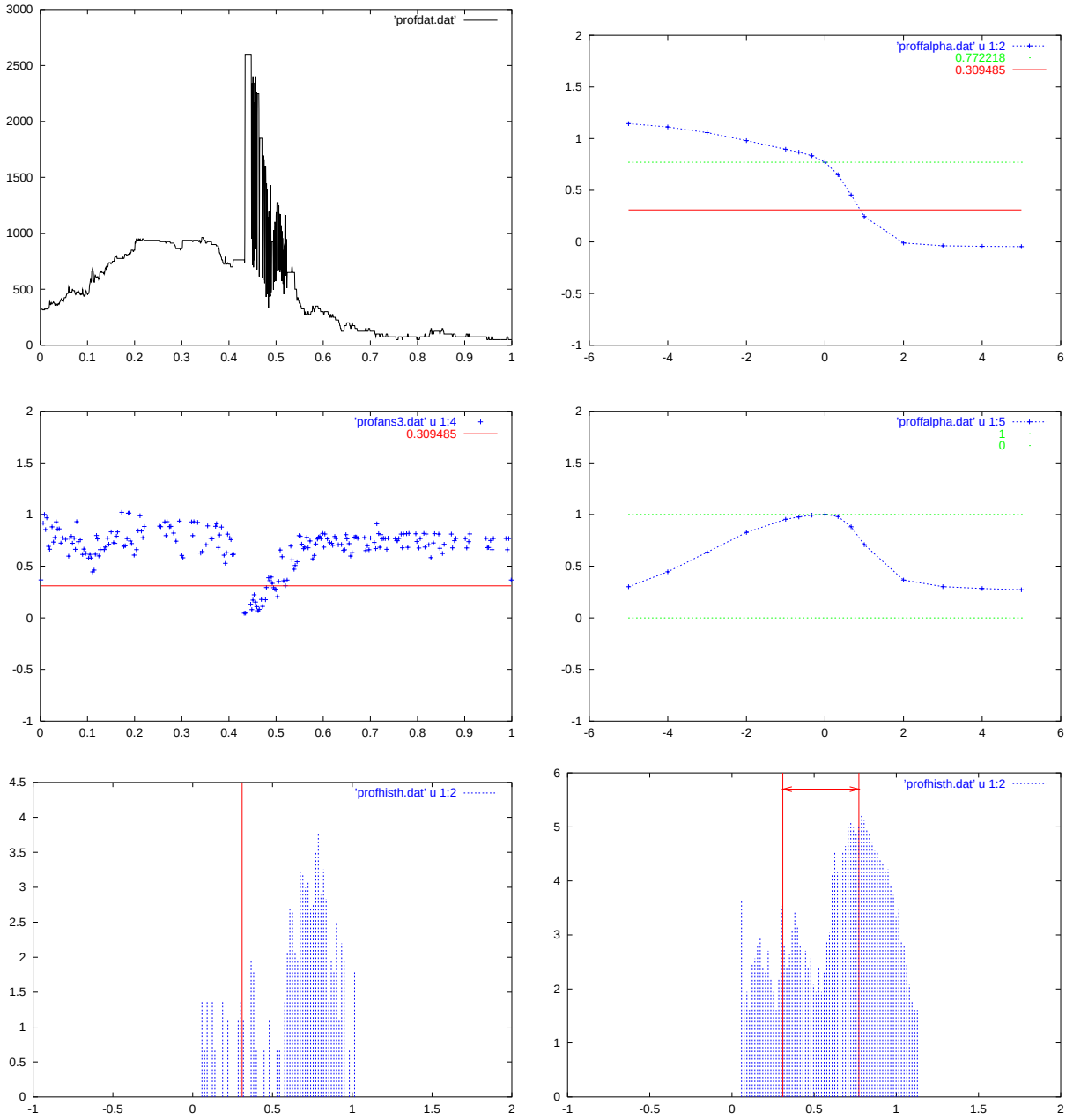

Figure 13: 'Second_Dirty'. Anti-clockwise: i) The input signal. ii) The local effective Hölder exponent at the scale $a = 10$ with the $\hat{h} \sim 0.31$ threshold. Positions corresponding with h lower than the threshold indicate localisation of the outliers. iii) The histogram of the local effective Hölder exponent (for the two lowest scale levels). Again the threshold is shown. iv) The histogram of the local effective Hölder exponent (for the twenty lowest scale levels). Both the threshold $\hat{h}$ and the main mode h_{mean} of the distribution is shown. The large discrepancy/distance between the two indicates the presence of outliers. v) $D(q)$ plot reaches below 0.5 for $q = 2$ and vi) $h(q)$ plot with both the threshold $\hat{h}$ and the main mode h_{mean} of the distribution is shown. The distance between $h(q = 2)$ and h_{mean} is larger than 0.5, indicating the presence of outliers.

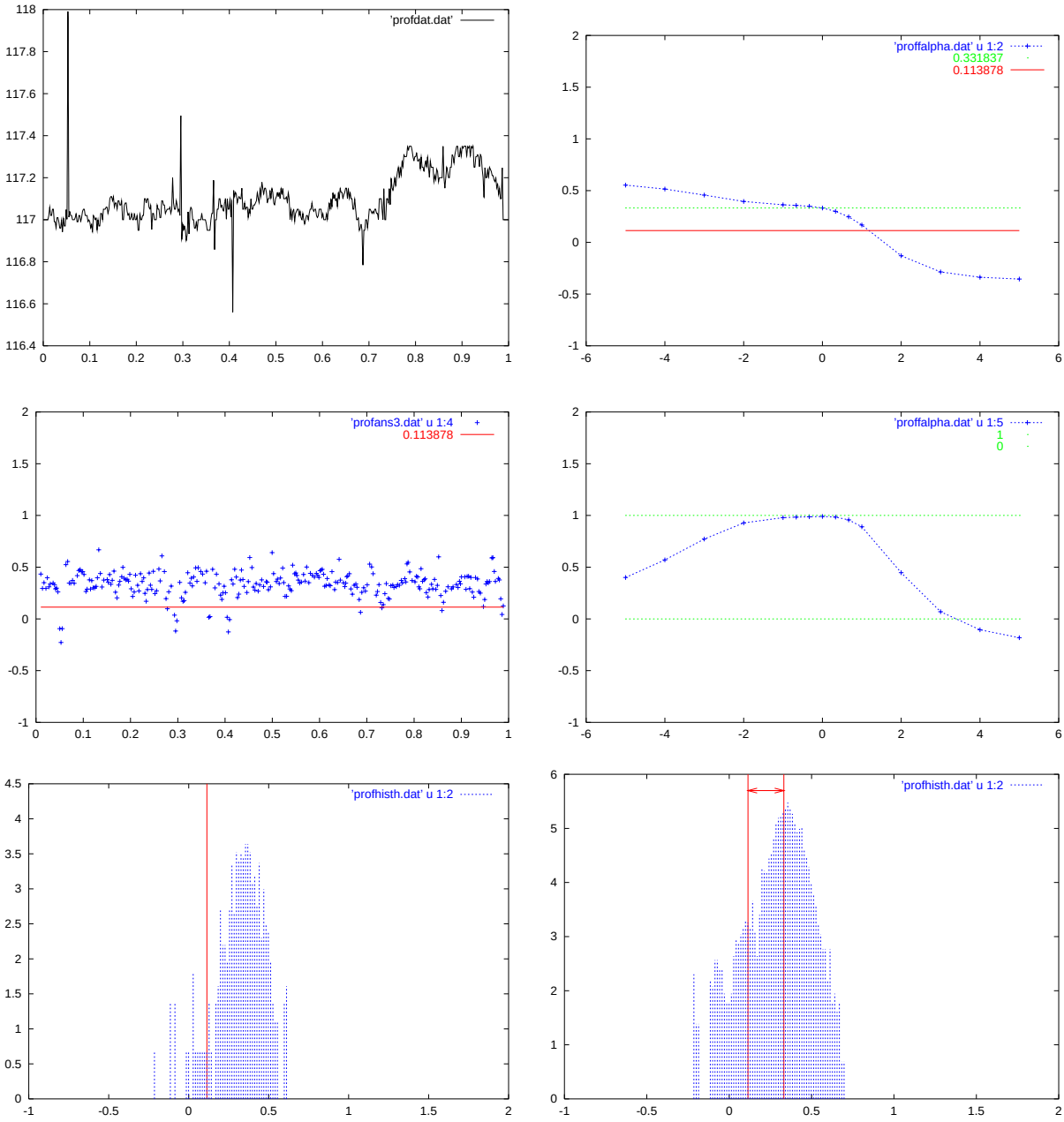

Figure 14: 'First_Dirty'. Anti-clockwise: i) The input signal. ii) The local effective Hölder exponent at the scale $a = 10$ with the $\hat{h} \sim 0.11$ threshold indicated identifying the outliers. iii) The histogram of the local effective Hölder exponent (for the two lowest scale levels). Again the threshold is shown. iv) The histogram of the local effective Hölder exponent (for the twenty lowest scale levels). Both the threshold $\hat{h}$ and the main mode h_{mean} of the distribution is shown. The discrepancy/distance between the two is relatively large and together with the multi-modality of the distribution, it indicates the presence of outliers. v) $D(q)$ plot reaches below 0.5 for $q = 2$ and vi) $h(q)$ plot with both the threshold $\hat{h}$ and the main mode h_{mean} of the distribution is shown. The distance between $h(q = 2)$ and h_{mean} is larger than 0.5, indicating the presence of outliers.

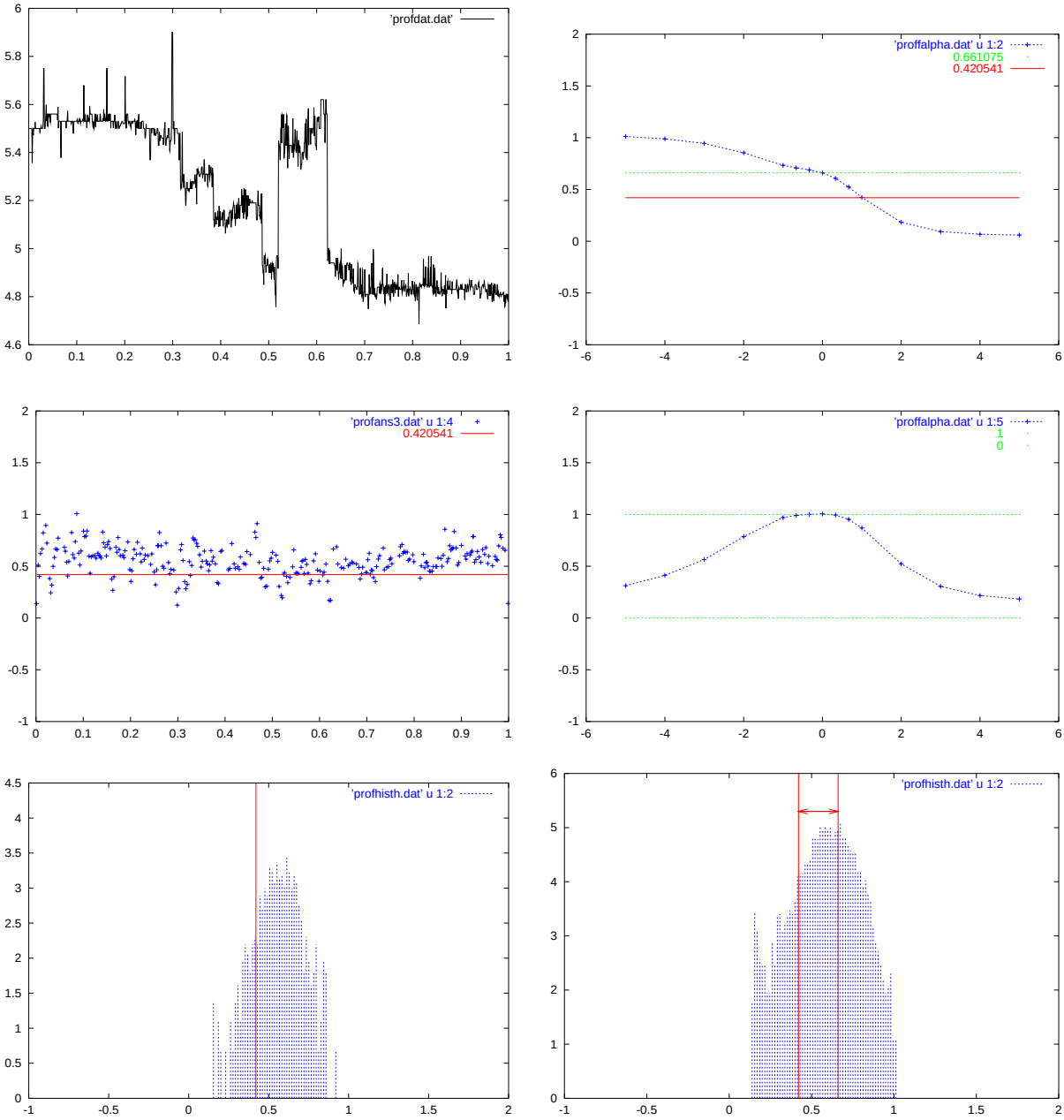

Figure 15: 'Third_Dirty'. Anti-clockwise: i) The input signal. ii) The local effective Hölder exponent at the scale $a = 10$ with the $\hat{h} \sim 0.42$ threshold indicated. A large number of outliers can be identified. iii) The histogram of the local effective Hölder exponent (for the two lowest scale levels). Again the threshold is shown. iv) The histogram of the local effective Hölder exponent (for the twenty lowest scale levels). Both the threshold $\hat{h}$ and the main mode h_{mean} of the distribution are shown. The discrepancy/distance between the two is still large but the distribution is rather compact. The probability of outliers is high but should be treated with caution. Either there are very many outliers at various levels or this is a multifractal process. v) $D(q)$ plot reaches just below 0.5 for $q = 2$ and vi) $h(q)$ plot with both the threshold $\hat{h}$ and the main mode h_{mean} of the distribution is shown. The distance between $h(q = 2)$ and h_{mean} is just larger than 0.5. This indicates the presence of outliers, but should be treated with caution.

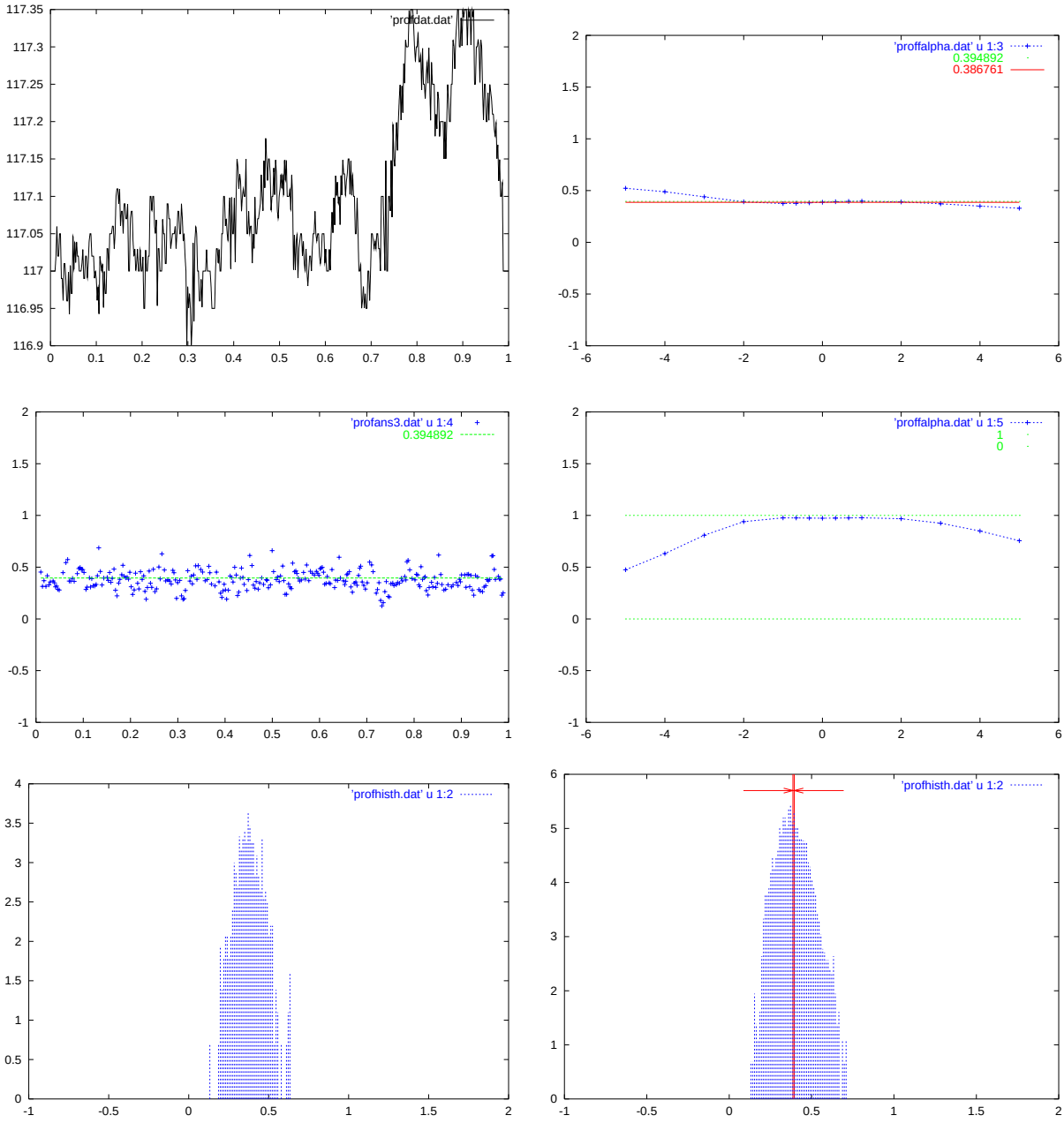

Figure 16: 'First_Clean'. Anti-clockwise: i) The input signal. ii) The local effective Hölder exponent at the scale $a = 10$. iii) The histogram of the local effective Hölder exponent (for the two lowest scale levels). iv) The histogram of the local effective Hölder exponent (for the twenty lowest scale levels). Both the threshold $\hat{h}$ and the main mode h_{mean} of the distribution are shown, but they nearly overlap. v) $D(q)$ plot remains high for $1 \leq q \leq 5$. vi) $h(q)$ plot shows almost no difference between h_{mean} and $h(q = 2)$.

References

1. A. Arneodo, E. Bacry and J.F. Muzy, Wavelet Analysis of Fractal Signals: Direct Determination of the Singularity Spectrum of Fully Developed Turbulence Data
2. A. Arneodo, E. Bacry and J.F. Muzy, The Thermodynamics of Fractals Revisited with Wavelets. *Physica A*, **213**, 232 (1995).
   J.F. Muzy, E. Bacry and A. Arneodo, The Multifractal Formalism Revisited with Wavelets. *Int. J. of Bifurcation and Chaos* **4**, No 2, 245 (1994).
3. Z. R. Struzik, Local Effective Hölder Exponent Estimation on the Wavelet Transform Maxima Tree, in *Fractals: Theory and Applications in Engineering*, Eds: M. Dekking, J. Lévy Véhel, E. Lutton, C. Tricot, Springer Verlag, (1999).
4. Z.R. Struzik The Wavelet Transform in the Solution to the Inverse Fractal Problem. *Fractals* **3** No.2, 329 (1995).
   Z.R. Struzik, From Coastline Length to Inverse Fractal Problem: The Concept of Fractal Metrology. *Thesis*, University of Amsterdam. (1996).
   Z.R. Struzik, Fractals under the Microscope or Reaching Beyond the Dimensional Formalism of Fractals with the Wavelet Transform. *CWI Quarterly*, **10**, No 2, 109 (1997).
5. I. Daubechies, *Ten Lectures on Wavelets*, (S.I.A.M., 1992).
6. M. Holschneider, *Wavelets - An Analysis Tool*, (Oxford Science Publications, 1995).
7. S.G. Mallat and W.L. Hwang, Singularity Detection and Processing with Wavelets. *IEEE Trans. on Information Theory* **38**, 617 (1992).
8. S.G. Mallat and S. Zhong Complete Signal Representation with Multiscale Edges. *IEEE Trans. PAMI* **14**, 710 (1992).
9. S. Jaffard and Y. Meyer, Wavelet Methods for Pointwise Regularity and Local Oscillations of Functions. *Memoirs of AMS*, 123 (1996).
10. B. Guiheneuf and J. Lévy Véhel, 2-Microlocal Analysis and Application in Signal Processing, in *Proc. of Int. Wavelets Conference*, Tangier (1998).
11. A. Arneodo, E. Bacry and J.F. Muzy, Oscillating Singularities in Locally Self-Similar Functions. *PRL*, **74**, No 24, 4823 (1995).
12. A. Arneodo, E. Bacry and J.F. Muzy(1994): Solving the Inverse Fractal Problem from Wavelet Analysis. *Europhysics Letters*, **25**, No 7, 479–484.

13. Z.R. Struzik, Removing Divergences in the Negative Moments of the Multi-Fractal Partition Function with the Wavelet Transformation. *CWI Report*, **INS-R9803**. Also see 'Fractals and Beyond - Complexities in the Sciences', M.M. Novak, Ed., World Scientific, 351 (1998).
14. Z.R. Struzik, A. Siebes, The Haar Wavelet Transform in the Time Series Similarity Paradigm, in *Principles of Data Mining and Knowledge Discovery*, Eds J.M. Żytkow, Jan Rauch. Lecture Notes in Artificial Intelligence **1704**, Springer, (1999).