Stochastic modelling and control of communication networks
Zuraniewski, P.W.

Citation for published version (APA):
Chapter 5

Hurst parameter changepoint detection

Network traffic is known to exhibit fractal characteristics such as long range dependence (LRD), which can be efficiently measured using wavelet transforms. Current estimation techniques for fractal parameters, such as the LogScale diagram, are not able to track the changes of the parameters. This chapter proposes and studies the combined use of wavelet transforms (DWT, MODWT, DTWT) and changepoint detection algorithms (ICSS, SIC) in order to detect the instants that the fractality changes noticeably. The different approaches are contrasted and statistical assessment is provided, together with results obtained when applying the procedure to synthetic traces.

1 Introduction

Network traffic exhibits some fractal properties, such as self-similarity, long-range dependence, or multifractality [16, 52]. Roughly speaking, fractality means that network traffic has similar behavior at different time scales. Traffic fractality has important implications on network performance, such as a substantial increase of the buffer overflow probability [53] (compared to what one would get by using models that do not incorporate the fractality). The Discrete Wavelet Transform (DWT)-based fractal parameter estimator, also known as the Logscale Diagram [54] is widely accepted as one of the best and most efficient estimators of fractality. It is based on a multiresolution decomposition that analyzes the input process at different time scales and computes the distribution of variance across scales.
Network traffic is known to be highly variable, and we expect its fractality to change across time. Some authors [54–58] have studied the case of time-varying fractal parameters. Detecting the changepoints between traffic segments with homogeneous fractality can be useful for particular algorithms and network mechanisms that exploit the fractal properties of network traffic, such as the TCP congestion control described in [59], a predictive bandwidth control for MPEG sources [60], or the effective bandwidth estimator described in [61], among other algorithms. When the scaling parameter of traffic changes, so does the variance structure of the signal. A changepoint detection algorithm may be applied to the output of the wavelet transform of the original signal at every scale. The change in the fractality parameters could be detected as a simultaneous variance change across scales.

We have used in our experiments the Iterated Cumulative Sum of Squares (ICSS) and the Schwarz Information Criterion (SIC). The former is based on a cumulated sum (CUSUM) technique, while the latter uses information theory concepts. The choice of the wavelet transform is important, since the statistical characteristics of its output can limit or degrade the performance of the changepoint detection procedures.

We compare the Discrete Wavelet Transform (DWT), its Maximum Overlap version (MODWT), and the Dual Tree Wavelet Transform (DTWT). The first one is an orthogonal (nonredundant) transform with poor time-location features, while the other two are redundant (in different degrees) and provide a better location of events in time. On the other hand MODWT suffers from a high correlation that makes prohibitive its direct use for changepoint detection algorithms, while DTWT offers a good trade-off between time accuracy, correlation, and complexity. This chapter describes the transforms, the algorithms, and several practical issues related. Our long-term aim is the development of an efficient, real-time, reliable monitoring algorithm capable of detecting the transitions between regions of homogeneous fractality.

The rest of the paper is organized as follows: Section 2 introduces basic concepts and previous works; Section 3 studies the limitations of the MODWT and discusses the failure of an alternative approach based on the concept of equivalent degrees of freedom. Section 4 describes the results obtained with DWT and DTWT transforms when combined with the ICSS and SIC methods. Section 5 concludes the chapter and describes future research topics.
2 Basic concepts and previous work

2.1 Long-range Dependence

A stationary stochastic process \( \{ X_n \} \), \( n \in \mathbb{N} \) is considered long-range dependent (LRD) if its autocorrelation function decays at a rate slower than a negative exponential (but also other definitions are sometimes used!). In the frequency domain LRD appears as a \( 1/f \)-like spectrum around the origin:

\[
S(f) \sim \frac{c_f}{|f|^\alpha}
\]

for \( |f| \to 0 \). The parameters of LRD are \( \alpha \) and \( c_f \). The scaling parameter \( \alpha \) is related to the intensity of the LRD phenomenon (a qualitative measure) and is usually expressed as the Hurst parameter \( H = (1 + \alpha)/2 \), while \( c_f \) has the dimension of the variance and can be interpreted as a quantitative measure of LRD.

2.2 Wavelet Transforms: DWT, MODWT and DTWT

Consider some function \( \Psi(t) \) with finite support such that the family of its rescaled and dilated versions

\[
\Psi_{j,k}(t) = 2^{-j/2}\Psi(2^{-j}t - k), \quad j = 1, 2, \ldots, J; k \in \mathbb{Z}
\]

forms an orthogonal basis, known as a wavelet basis. Now for the given \( x = \{ x_n \} \) we may calculate \( d_x(j, k) = \langle x, \Psi_{j,k} \rangle \), referred to as detail coefficients, which may be regarded as a measure of correlation between the analyzed \( x \) and a given basis function. Due to the scaling and shifting governed by \( j \) and \( k \) respectively, wavelet analysis can be seen as simultaneous analysis in the time and frequency domain.

An intuitive interpretation of the DWT is a cascade of quadrature-mirror low-pass and high-pass filters (\( h \) and \( g \), respectively), where the output of the high-pass filter (the details of the signal) are the coefficients of the DWT at scale \( j \), while the output of the low-pass filter (the approximation of the signal) is filtered again iteratively. Down-sampling maintains orthogonality (\( N \) input samples produce \( N \) output coefficients) by halving the coefficients at each filter, making difficult the time location of signal changes. In terms of the signal spectrum the output of the DWT is multiresolution analysis in which the original signal is decomposed into a low-pass approximation at scale \( J \), \( a_x(J, k) \) and a set of high-pass details \( d_x(j, k) \) for each scale \( j = 1, \ldots, J \). Fig. 5.1 provides an example.
A close relative of the DWT is the Maximum-overlap Wavelet Transform (MODWT), also known as Stationary or Shift-invariant Wavelet Transform [62]. It is essentially identical to the DWT except that no downsampling is performed at any step, which results in $J \cdot N$ coefficients at the output, time-shift invariance and the possibility of accurately locating time-dependent events. On the other hand redundancy results in a higher degree of correlation between coefficients, together with an increase in complexity.

An interesting and recent development is the Dual Tree (Complex) Wavelet Transform (DTWT) [63]. The transform can be interpreted as a dual DWT-style tree in which the original samples to be analyzed in one tree are the odd positions, while the even-numbered samples are analyzed by the second tree. The complex term comes from the interpretation of every couple of coefficients from both trees ($a$ and $b$) as the real and imaginary parts of a complex number $a + jb$ ($j$ denotes the imaginary unit here). Given that the filters of the trees meet some conditions (Q-shift filters with all the filters beyond the 1st level being of even lengths and interleaved on both trees with a delay of approximately $1/4$ and $3/4$ sample [63]), the DTWT can provide an almost shift invariant analysis with almost uncorrelated coefficients at the expense of a slight increase in redundancy ($2N$ samples) and complexity when compared with DWT. See Fig. 5.2 for a DTWT decomposition example.

2.3 Wavelet-based analysis of LRD signals

Abry and Veitch developed the LogScale diagram [54], an unbiased and efficient estimator of the LRD parameters. It consists of the computation of the sample variance of the coefficients at each subband of the DWT decomposition, $\mu_j$, as an estimation of the power of the original signal at those subbands. Since the signal follows a power-law,
we have
\[ \mu_j := \mathbb{E} d_x^2(j, k) = 2^j \alpha c_f C(\alpha, \Psi) \] (5.3)

The above expression (after taking the logarithm) would suggest the use of a linear regression approach for the estimation of the LRD parameters. There are, however, two difficulties [54]: we have to estimate the second-order quantity \( d_x^2(j, k) \), and there is a bias introduced by the nonlinear operation of taking the logarithm. A possible solution is the following: define

\[ \hat{\mu}_j := \frac{1}{n_j} \sum_{k=1}^{n_j} d_x^2(j, k) \]

which is an unbiased and efficient estimator of \( d_x^2(j, k) \), with \( n_j \) being the number of the wavelet coefficients at scale \( j \). Abry and Veitch in [54] demonstrated that the following relationship holds:

\[ \log_2(\hat{\mu}_j) = j \alpha + \log_2(c_f C(\alpha, \Psi)) + g_j \] (5.4)

where \( g_j \) is a bias-correction term that reflects the fact that the expectation of the logarithm is not the logarithm of the expectation. Here \( g_j \) depends only on \( n_j \) (the number of wavelet coefficients at scale \( j \)):

\[ g_j = \frac{\Phi(n_j/2)}{\ln 2} - \log_2(n_j/2) \] (5.5)
where $\Phi(x) = \Gamma(x)/\Gamma'(x)$ is a digamma function. Parameters $\alpha$ and $c_f C(\alpha, \Psi)$ can be estimated from (5.4) performing a weighted linear regression on $\hat{\mu}_j$, in which the weight of the regression samples is related to the number of coefficients at the corresponding scale (halved for each scale increment). Assuming $\hat{\mu}_j$ follows a Gaussian distribution, confidence intervals for the estimation are easily derived.

The expressions for the LogScale analysis with MODWT and DTWT are similar to those presented for DWT (the changes are the amount of samples available at each scale, a normalization of the variance, etc.) [62, 63].

### 2.4 Analysis of time-varying LRD traffic

There exists an on-line version of the DWT estimator [56] where the subband variance computation is performed progressively, but it is accumulative, not dynamic; it returns updated estimations performed over all available samples from time $t = 0$. The same authors studied the stationarity of the scaling exponent ($\alpha$ or $H$), developing a statistical test capable of determining its constancy [57], but the study is limited to non-overlapping segments whose lengths are equal. Traffic characteristics, however, may be expected to change at an arbitrary time instance.

The authors of [58] describe so-called ‘accuracy-based approach’ for the real-time estimation of multifractality, where the length of the windows (where fractality is assumed constant) is related to the confidence interval of the estimation, but a minimum window length is still assumed. Our approach, which is based on finding the alignment in time of variance changepoints detected across the scales provided by the wavelet transform, is more general. We also use as an auxiliary mechanism some ‘windowing’ method, because the Hough transform (see Section 2.6 below) requires providing a resolution parameter. The final result is, however, a changepoint located at any position (excluding a few ones close to the borders of the analyzed series) provided that the wavelet transform has the proper temporal resolution; MODWT has the maximum resolution at all scales, while DWT and DTWT perform downsamplings and lose resolution at higher scales). On the other hand we lose generality due to our focus on monofractality rather than multifractality, but recent works show that Internet backbone traffic can be modelled as a Poisson process for sub-second scales, while at higher scales piecewise-linear nonstationarity and long-range dependence seem to fit well with the measurements made in real networks [64]. Therefore our work can be useful for the detection of changes in the LRD regime at multi-second scales (which is the scale at which the algorithms will actually work, due to complexity constraints).
2.5 Variance changepoint detection

We begin with the statement of this classical statistical problem. Let \( x_1, x_2, \ldots, x_N \) be a sequence of independent random variables with a common mean \( \mu \) and variances \( \sigma^2_1, \sigma^2_2, \ldots, \sigma^2_N \). We want to test the null hypothesis

\[
H_0 : \quad \sigma^2_1 = \sigma^2_2 = \ldots = \sigma^2_N = \sigma^2
\]

versus the alternative

\[
H_1 : \quad \sigma^2_1 = \ldots \sigma^2_k \neq \sigma^2_{k+1} = \ldots \sigma^2_N
\]

where \( q \) is the unknown number of changepoints and

\[1 < k_1 < k_2 < \ldots < k_q < N\]

are the unknown positions of the changepoints.

Two variance changepoint detection algorithms have been tested: the Iterated Cumulated Sum of Squares (ICSS) by Inclan and Tiao [65], and the Schwarz Information Criterion (SIC) by Chen and Gupta [25]. The ICSS computes sample by sample \( (n) \) the cumulative sum of the squares of the \( N \) samples and decides that a change is present when the cumulative sum deviates notably from the homogeneous variance case (a \( n/N \)-like cumulative sum). For a series \( \{x_n\}_{n=1}^{N} \), the statistic is defined as:

\[
\text{ICSS}(n) = \left| \frac{\sum_{j=0}^{n} x_j^2}{\sum_{j=0}^{N} x_j^2} - \frac{n}{N} \right|.
\]

The ICSS absolute (and local) maxima and minima are the candidates for variance changepoints. They are iteratively validated or rejected by segmenting the time series between these points and reevaluating the candidates till the algorithm converges. The reevaluation phase rejects spurious candidates. The critical values of the statistic are derived from Monte-Carlo simulations, as shown in [65].

SIC is based on information theory concepts. It can only detect the presence of a single changepoint, but can be easily generalized thanks to the binary segmentation procedure, iterating the process in the two subsequences that surround the first detected changepoint. If a change is detected in the subsequence, one should split again and iterate the process until no more changes are found in any of the subsequences (note that \( N \) may in this situation mean the length of the currently processed subsequence).
Therefore, our problem is reduced to testing the null hypothesis (5.6) against the alternative

\[ H'_1 : \sigma_1^2 = \ldots = \sigma_{k_0}^2 \neq \sigma_{k_0+1}^2 = \ldots = \sigma_N^2 \]  

(5.9)

The intuition behind SIC is that a sequence with a variance changepoint has higher entropy (disorder) than a sequence with constant variance. Therefore, a way to detect simultaneously the presence and the location of a single changepoint is to compute the entropy of the sequence (samples \( i = 1, \ldots, N \)) and of the pairs of subsequences \((i = 1, \ldots, k \) and \(i = k + 1, \ldots, N, 1 < k < N)\), compare the values, and decide that a change is present at position \( k \) if the entropy of the subsequences is lower than the entropy of the sequence.

SIC is derived from the Akaike Information Criterion (AIC) for model selection and is defined as follows:

\[ SIC(k) = -2 \log L(\hat{\Theta}) + p \log k \]  

(5.10)

where \( L(\hat{\Theta}) \) is the Gaussian maximum likelihood function for the model (see Section 5 for comments on non-normality), \( p \) is the number of free parameters in the model, and \( k \) is the sample size. The models correspond to the null and the alternative hypotheses. The decision will be taken following the principle of minimum information: no evidence to reject \( H_0 \) is found if \( SIC(N) \leq \min_k SIC(k) \), and \( H_0 \) is rejected if \( SIC(N) > SIC(k) \) for some \( k \). The estimated position of the change \( \hat{k} \) is such that

\[ SIC(\hat{k}) = \min_{1 \leq k < N} SIC(k) \]  

(5.11)

where \( SIC(N) \) is the SIC statistic under \( H_0 \) and \( SIC(k) \) is the SIC under \( H_1 \) for \( k = 1, \ldots, N - 1 \). Then, expression (5.10) can be rewritten as follows:

\[ SIC(N) = N \log 2\pi + N \log \hat{\sigma}_2^2 + N + \log N \]
\[ SIC(k) = N \log 2\pi + k \log \hat{\sigma}_1^2 + (N - k) \log \hat{\sigma}_2^2 + N + 2 \log N \]  

(5.12)

where \( \hat{\sigma}_2^2, \hat{\sigma}_1^2 \) and \( \hat{\sigma}_2^2 \) are the maximum likelihood estimators of the variances of the whole sequence and the subsequences, respectively (note that due to the properties of wavelets, the mean is assumed to be zero). The position of the changepoint is estimated by the \( \hat{k} \) such that

\[ SIC(\hat{k}) = \min_{2 \leq k < N-1} SIC(k). \]  

(5.13)

The introduction of a significance level \( \alpha \) (which should not be confused with the LRD parameter) and its associated critical value \( u_\alpha \) allows us to filter out the changepoints.
which are believed to be spuriously detected because of the normal statistical fluctuations of the SIC value:

\[ 1 - \alpha = \mathbb{P} \left\{ \text{SIC}(N) < \min_{1 < k < n} \text{SIC}(k) + u_\alpha \mid H_0 \right\}. \] (5.14)

The values of the critical levels \( u_\alpha \) and a discussion on the consistency of \( \hat{k} \) can be found in [25]. The SIC method will be used after some modifications with MODWT and directly with both DWT and DTWT.

### 2.6 Finding alignments: the Hough transform

After obtaining the changepoints for every scale, an alignment-detection algorithm is needed, in order to observe whether a change is present in a sufficient number of scales. The first step is to detect the alignment of the changepoints at the different scales but close enough in time. The Hough Transform [66] is used in signal processing in order to detect features of a particular shape within images. It maps a line in the space domain \((x, y)\) to a point in the polar representation \((\rho, \theta)\); inversely, a point \((x_0, y_0)\) in the space domain is mapped to a cosinusoid in the Hough domain. The alignment of points is detected in the Hough space as the intersection of several cosinusoids. A practical implementation is based on the discretization of the Hough space into bins and the accounting of the amount of ‘votes’ (changepoints) that each bin receives. The procedures need a couple of parameters: the bin size for the discretization and the voting quorum.

### 3 Limitation of the MODWT-SIC algorithm

In [67, 68] the authors examine the possibility of using a combination of SIC and MODWT for changepoint detection of the Hurst parameter \( H \). They hope that a transform that lacks a decimation step will be a remedy for having a low number of coefficients at higher scales, which happens in case of using DWT. MODWT, compared to DWT, is also known for its good localization of events in the analyzed time series. However, the price we pay for that is that the autocorrelation of the wavelet coefficients increases with every scale. The SIC method, on the other hand, requires uncorrelated input; to compensate correlations somewhat, in [68] some uncommon values of certain parameters (such as the significance level \( \alpha \)) were used. We tried to attack the problem by using the concept of equivalent degrees of freedom (EDOF), which may be helpful for wavelet-based variance estimation in case of correlation due to for example the use of MODWT.
Following [62], the wavelet variance for the stationary process \( \{X_n\}, n \in \mathbb{N} \) is defined for scale \( j \) as

\[
\nu^2(j) = \text{Var}(\tilde{d}(j, k))
\]  

(5.15)

where \( \tilde{d}(j, k) \) denote the MODWT coefficients obtained by filtering \( \{X_n\}, n \in \mathbb{N} \). A fundamental property of wavelet variance is its ability to decompose the original process variance \( \sigma_X^2 \) across scales:

\[
\sum_{j=1}^{\infty} \nu^2(j) = \sigma_X^2.
\]  

(5.16)

An unbiased estimator of \( \nu^2(j) \) is given by

\[
\hat{\nu}^2(j) = \frac{1}{M_j} \sum_{k=L_j}^{N} \tilde{d}^2(j, k)
\]  

(5.17)

where \( M_j := N - L_j + 1 \) and \( L_j := (2^j - 1)(L - 1) + 1 \) for a wavelet filter of length \( L \). If the \( \tilde{d}(j, k) \) had a Gaussian distribution and were uncorrelated, their sum of squares would follow (with a proper renormalization) a chi-square distribution with \( M_j \) degrees of freedom. Due to the correlation, however, some modification is required, for example by adjusting the degree of freedom and by approximating

\[
\frac{\eta \hat{\nu}^2(j)}{\nu^2(j)} \approx d \chi^2
\]  

(5.18)

with \( \eta \) being a constant known as the ‘equivalent degree of freedom’ [62]. One of the possible estimators of \( \eta \) is

\[
\hat{\eta} = \frac{M_j \hat{\nu}^4(j)}{\hat{A}_j}
\]  

(5.19)

for

\[
\hat{\eta} = \frac{\hat{\nu}^4(j)}{2} + \sum_{t=1}^{M_j-1} \hat{s}_{j,t}^2
\]  

(5.20)

and \( \hat{s}_{j,t}^2 \) being the usual biased estimator of the autocovariance of \( \tilde{d}(j, k) \). In [69] one can find a study where in the ICSS test statistics, used for the variance changepoint detection, the sample size \( N \) was substituted with \( \eta \) to compensate the correlation resulting from using MODWT. We tried to employ a similar technique by rewriting the equations (5.12) to the form

\[
\hat{SIC}(N) = \hat{\eta} \log 2\pi + N \log \hat{\sigma}^2 + \hat{\eta} + \log \hat{\eta}
\]

\[
\hat{SIC}(k) = \hat{\eta} \log 2\pi + k \log \hat{\sigma}_1^2 + (N - k) \log \hat{\sigma}_2^2 + \hat{\eta} + 2 \log \hat{\eta}
\]  

(5.21)

(note that we kept \( N \) in estimating \( \hat{\sigma}^2, \hat{\sigma}_1^2 \) and \( \hat{\sigma}_2^2 \) and when using their logarithms, because the changepoint \( k \) may appear at positions 2, \ldots, N - 2). We conducted a
small numerical study (100 independent repetitions) using a fractional Gaussian noise generator [70, 71], choosing a sample of size $N = 2048$, Hurst parameter 0.6 and 0.9 in the segments of length 1000 and 1048 respectively. The variance of the whole process was kept constant and equal to 1, the significance level $\alpha$ was set to 0.1 (for the values of 0.05 and 0.01 the outcomes did not differ) and Daubechies’ ‘db4’ filter was used (no improvement observed with the ‘symlet4’ filter). The results obtained were not appealing: on the scales $j = 3, \ldots, 6$ we observed more or less uniformly distributed spurious changepoints instead of the desirable sharp peak around the true value (see Fig. 5.3).

![Figure 5.3: Failure of SIC-EDOF-MODWT: real changepoint at 1000, max number of detections could be 100, scale $j = 6$.](image)

A possible reason of the failure of this method could lie in the use of critical values of the distribution (derived under an independence assumption in [25]) for the correlated data. Another explanation might be the bias (due to the correlation) of the variance estimators in the SIC calculations. Perhaps a more sophisticated modification of the SIC method than ours would give better results.

## 4 Results with DWT and DTWT

The high degree of correlation detected in the output of the MODWT makes the transform not suitable for its joint use with changepoint detection algorithms, which expect their inputs to be independent random variables. DWT is still an option, but its poor resolution at higher scales, as well as its lack of shift invariance, are inconvenient. That is why we turned our attention to the Dual Tree DWT. This transform retains almost all the good properties of the MODWT at a lower redundancy expense, and what is more important, providing an almost uncorrelated output which makes it suitable for using ICSS and SIC.
In order to provide some statistical analysis of the performance of the algorithms we designed two different scenarios and analyzed them with four combinations: DWT-ICSS, DWT-SIC, DTWT-ICSS and DTWT-SIC. Because of its phase properties, the ‘symlet4’ filter was used for DWT and ‘near_sym_b’ along with ‘qshift_b’ were our choice for DTWT to achieve good results in terms of changepoint localization. All the tests were performed with fractional Gaussian noise (FGN) segments; we use FGN because it seems to be one of the preferred models for network traffic [16]. We do not provide any test with real traffic, since in this study we are rather interested in comparing the algorithms from a statistical point of view, and this can only be achieved with synthetic traces where the true changepoints are known.

In [67, 68] it was shown with DWT-ICSS and DWT-SIC that real traffic traces exhibit highly variable values of both the variance (traffic volume) and Hurst parameter. The first scenario includes an 8192-sample FGN trace with constant Hurst parameter ($H = 0.8$) and three segments with different variances: 1, 1.75 and 6, providing an example of a smooth variance change at position 2001 and an abrupt one at position 5001. The second scenario simulates a constant variance situation with a number of Hurst parameter changes. The trace starts with $H = 0.8$, changes to $H = 0.9$ (smooth change) at 2001, and then changes again to $H = 0.6$ (abrupt change) at sample 5001. For each combination of transform and changepoint detection algorithm we performed a Monte Carlo test with 1000 runs with significance level $\alpha = 0.05$. The Hough transform-based clustering procedure was included in the tests. By a successful detection we regarded a presence of the detected changepoint in a neighborhood of length 129, centered in the true value of the estimated parameter ($\pm 64$ points around the true changepoint position). An interval of this size is about 1.5% of the length of the whole time series.

Observe that for the DTWT method, instead of processing approximation and detail coefficients $a_{xa}(j, k), d_{xb}(j, k)$ from both trees separately, the quantity

$$z_x(j, k) = \sqrt{a_{xa}^2(j, k) + d_{xb}^2(j, k)}$$

is used as an input for the ICSS/SIC algorithm. $z_x(j, k)$ may be regarded as a modulus of a complex number. [72] provides an explanation of this approach along with a shift invariance analysis. An alternative approach, i.e., combining sets of detected candidate changepoints and doubling the threshold for deciding which candidates become valid changepoints, is very conservative and results in a relatively low power of the test, as reported in [14, 72].

The parameters to be studied are the ratio of successfully detected changepoints (a
rough measure of the power of the test), the mean and median position of the change-
point found by the algorithms, the variance around the true changepoint, and the skew-
ness also with respect to a true value.

For the first changepoint in the first scenario (Table 5.1, Fig. 5.4), DTWT seems to lo-
calize the changepoint with a higher precision, which was to be expected due to its
property of being time shift invariant. Lack of this property for DWT-ICSS results in
most of the detected changepoints being shifted with respect to the true changepoint at
position of 2000, which is reflected in the mean and median values as well as high vari-
ance rate as reported in Table 5.1. Regarding the power of the test, note the possibility
of increasing it if we count as the correct detection all the changepoints located in the
neighborhood of the true changepoint larger than used \(2000 \pm 64\), cf., \(2000 \pm 128\). This
effect is not that much pronounced for the DTWT case. For the second changepoint,
all the methods perform very well, but again DWT-ICSS suffers from misalignment
(mean and median about 20 points away from the true value) which is reflected in a
higher variance. All the methods have a comparable rate of asymmetry calculated by
the skewness parameter.

For the second scenario (Table 5.2) we encounter a problem with detecting the first
changepoint while still being able to locate the second one. The transition of the \(H\)
parameter produces ‘blind zones’; i.e., after detecting a change at certain scales (in our
case for the second changepoint at \(j = 1, 2\)), no changepoints are observed at higher
scales \((j = 3, 4)\) and reappear later \((j = 5, 6)\). See Fig. 5.5 for an illustration of this
phenomenon.

For the first changepoint, however, a high detection rate was present only for \(j = 1, 2\),
and to a very small extent on the sixth scale. With a minimum number of votes equal to
Chapter 5. Hurst parameter changepoint detection procedure

3, it resulted in an overall lack of power of the test. On a LogScale diagram, the ‘blind zone’ would correspond to the situation when the linear regression lines are close to each other. It may be likely that a larger sample size would help in this situation, but even if this was the case, this example shows the undesired possibility of confusing a process variance change with an $H$ transition. To distinguish between these situations one should possibly keep track of the scales at which changepoints were detected, and try to use the ‘blind zone’ concept.

For the second changepoint (Fig. 5.6) we observe overall good performance of all methods. One may notice that DWT offers a higher detection ratio at the expense of the shift with respect to the true changepoint, reflected again in a higher variance.

5 Conclusions and future research

We showed that the approaches based on the use of the MODWT, which seemed attractive because of their very good ability to localize events in a time series, suffer from the presence of correlation at the output of the transform. Hence the use of variance changepoint detection algorithms such as SIC or ICSS, which assume uncorrelated inputs, is not mathematically correct. We presented an alternative complex transform, the Dual Tree DWT, which retains some of the best properties of the MODWT such as perfect reconstruction and (almost) shift invariance, but avoids its inherent correlation.
Chapter 5. *Hurst parameter changepoint detection procedure*  

Figure 5.6: Scenario 2 (Hurst parameter change), 2nd changepoint. Left: DWT-SIC, right: DTWT-SIC.

**Table 5.1**: Results for Scenario 1 (1000 independent runs, variance change).

<table>
<thead>
<tr>
<th></th>
<th>1st changepoint</th>
<th>Ratio</th>
<th>Mean</th>
<th>Median</th>
<th>Var.</th>
<th>Skew.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWT-ICSS</td>
<td></td>
<td>28%</td>
<td>2037</td>
<td>2040</td>
<td>1664</td>
<td>1.19</td>
</tr>
<tr>
<td>DWT-SIC</td>
<td></td>
<td>27%</td>
<td>2032</td>
<td>2032</td>
<td>1424</td>
<td>1.18</td>
</tr>
<tr>
<td>DTWT-ICSS</td>
<td></td>
<td>28%</td>
<td>2002</td>
<td>2002</td>
<td>523</td>
<td>0.87</td>
</tr>
<tr>
<td>DTWT-SIC</td>
<td></td>
<td>34%</td>
<td>1993</td>
<td>1994</td>
<td>417</td>
<td>-0.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2nd changepoint</th>
<th>Ratio</th>
<th>Mean</th>
<th>Median</th>
<th>Var.</th>
<th>Skew.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWT-ICSS</td>
<td></td>
<td>97%</td>
<td>5021</td>
<td>5020</td>
<td>767</td>
<td>1.40</td>
</tr>
<tr>
<td>DWT-SIC</td>
<td></td>
<td>99%</td>
<td>5012</td>
<td>5011</td>
<td>513</td>
<td>1.19</td>
</tr>
<tr>
<td>DTWT-ICSS</td>
<td></td>
<td>95%</td>
<td>4999</td>
<td>4999</td>
<td>130</td>
<td>-0.27</td>
</tr>
<tr>
<td>DTWT-SIC</td>
<td></td>
<td>97%</td>
<td>4995</td>
<td>4996</td>
<td>172</td>
<td>-1.48</td>
</tr>
</tbody>
</table>

**Table 5.2**: Results for Scenario 2 (1000 independent runs, $H$ change). Estimators based on too small sample size skipped as being meaningless.

<table>
<thead>
<tr>
<th></th>
<th>1st changepoint</th>
<th>Ratio</th>
<th>Mean</th>
<th>Median</th>
<th>Var.</th>
<th>Skew.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWT-ICSS</td>
<td></td>
<td>8%</td>
<td>2022</td>
<td>2029</td>
<td>1362</td>
<td>0.71</td>
</tr>
<tr>
<td>DWT-SIC</td>
<td></td>
<td>2%</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>DTWT-ICSS</td>
<td></td>
<td>9%</td>
<td>1986</td>
<td>1985</td>
<td>459</td>
<td>-1.51</td>
</tr>
<tr>
<td>DTWT-SIC</td>
<td></td>
<td>1%</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2nd changepoint</th>
<th>Ratio</th>
<th>Mean</th>
<th>Median</th>
<th>Var.</th>
<th>Skew.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWT-ICSS</td>
<td></td>
<td>66%</td>
<td>5009</td>
<td>5010</td>
<td>661</td>
<td>0.92</td>
</tr>
<tr>
<td>DWT-SIC</td>
<td></td>
<td>73%</td>
<td>5014</td>
<td>5015</td>
<td>787</td>
<td>1.06</td>
</tr>
<tr>
<td>DTWT-ICSS</td>
<td></td>
<td>55%</td>
<td>4995</td>
<td>4997</td>
<td>244</td>
<td>-1.28</td>
</tr>
<tr>
<td>DTWT-SIC</td>
<td></td>
<td>52%</td>
<td>4996</td>
<td>4997</td>
<td>226</td>
<td>-1.16</td>
</tr>
</tbody>
</table>
We performed some statistical analysis of the performance of both DWT and DTWT with SIC and ICSS. The DTWT detection method is attractive because of its ability to precisely locate events in a time series. Using twice as much memory and computation power as DWT seems to be a fair justification.

Although the results presented in this paper are promising, there are still some open issues to be addressed in future studies. The Hough transform in general seems to perform well but some changepoints correctly recognized by SIC/ICSS may not be counted as valid ones. This is because at chosen resolution they fall just outside the bin border. Another problem is how to distinguish non-stationarities like shift in the mean or variance change from the $H$-transition points. These are important issues since LRD estimators can be ‘fooled’ by this type of behavior [55, 73].

Another open path is testing the algorithms with other wavelet transforms. Recent developments in wavelet theory have brought some options that could improve DWT and DTWT performance in our application scenarios. Examples could be Double Density DWT (DDDWT) [74] or Double Density Dual Tree Complex Wavelet Transform [75].

Finally, another area of future research relates to studying how departures from the ideal situation (non-Gaussianity, nonstationarity of the mean) influence the changepoint detection. Regarding the non-Gaussian distributions of the wavelet coefficients of real traffic traces, the Generalized Gaussian Distribution (which unifies Laplacian, Gaussian, Dirac’s delta and Uniform distributions) seems to be the most appropriate choice, as recently reported in [76], where the wavelet coefficients of real traffic traces show a mixture of Laplacian and Gaussian distributions. Chen and Gupta developed a generalization of the SIC algorithm for non-Gaussian samples [77]. It seems rather general, but imposes some conditions on the higher moments of the distribution (i.e., the distribution is identical to a standard normal distribution up to the fourth moment). One has to assess if this assumption could be relaxed. Our long-term efforts are focused towards designing a real-time characterization of fractal traffic which, in turn, can lead to the development of new fractal-aware network algorithms.