Wavelet Transform in Similarity Paradigm

Zbigniew R. Struzik, Arno Siebes

Centre for Mathematics and Computer Science (CWI)
Kruislaan 413, 1098 SJ Amsterdam
The Netherlands
email: `Zbigniew.Struzik@cwi.nl`

Abstract. Searching for similarity in time series finds still broader applications in data mining. However, due to the very broad spectrum of data involved, there is no possibility of defining one single notion of similarity suitable to serve all applications.

We present a powerful framework based on wavelet decomposition, which allows designing and implementing a variety of criteria for the evaluation of similarity between time series. As an example, two main classes of similarity measures are considered. One is the global, statistical similarity, which uses the wavelet transform derived Hurst exponent to classify time series according to their global scaling properties. The second measure estimates similarity locally using the scale-position bifurcation representation.

1 Introduction

Many data mining algorithms exist for (more or less) standard, relational data, see, e.g. [1]. However, in practice there is much non-relational data. The most important example is time series data. For example, banks have standard data on their clients, e.g. their names, where they live et cetera, but also a time series giving the status of their account over time.

To use existing data mining technology on such data means that the time series data has to be reduced to a fixed number of characteristics. A very simple idea would be to use the current status of the account as an extra field in the table. However, if we are going to use the data for credit scoring, the current status of the account is likely to miss out on important information. For example, two clients A and B could have both $10.000 in their account now, which is the normal status for A, whereas it is a one-time record for B. In such a case, the credit rating would be (much) higher for A than for B.

In other words, the behaviour of a time series over time is among the important characteristics of that time series. This means that we have to represent the behaviour of a time series with a finite number of characteristics. Of course, this representation should be such that two time series which show similar behaviour should be close to one another in the representation space, and vice versa.

Crucial in this statement is, of course, what similarity actualy means. The precise meaning of similarity is strongly dependent on the intended usage of the representation. Sometimes the trend of the series is the important factor in

determining similarity, whereas in other cases, it is everything but the trend. Sometimes it is global (statistical) behaviour which is important, whereas in other cases it is highly localised behaviour.

It is, therefore, not possible to define one specific measure of similarity between time series that is useful under all circumstances. Rather, a flexible toolbox, in which the user can indicate what is important in this specific case, is necessary. In this paper, we introduce a framework for such similarity measures based on wavelets. Moreover, we study the two extremes in this spectrum of possibilities in depth. The framework for similarity as developed in this paper is based on the fractal analysis of time series [2, 3, 4] [1].

The topic of the similarity of time series for data mining is not new. Important papers in this area are [5] and [6]. As an aside, note that these papers have other motives in showing why this topic is important. The most important difference between [5] and [6] on the one hand and our paper on the other is the framework. The core criterium for similarity used in [5], de facto requires a priori determining of what the time series is and what are the outliers or noise. Only then can the actual distance, ϵ sausage criterium, work.

Earlier work by the same authors [7] suggested matching in the space of the Discrete Fourier Transform representation. However, DFT in itself provides only global information. Moreover, as is also concluded in [5], this approach fails in the presence of linear bias and is rather sensitive to local outliers.

The work reported in [6] is based on local transformations of the time series. Since there is a choice of the allowed set of transformations, this approach is closer to our approach. Our approach, however, does not rely on an (implicit) underlying model, and thus is not sensitive to outliers, noise, and translations of the data.

In other words, we are not building our similarity framework using a particular similarity model. Rather we utilise a flexible hierarchical representation of a time series (up to a certain resolution). This representation in turn can be tailored to fit matching criteria required. In this approach, one can build in insensitivity to factors other than matching criteria, these often being outliers, noise, translation, scaling or polynomial bias.

We will continue with the discussion of similarity in the next section, introducing appropriate exponents for both global and local characterisation. In section 3, we will introduce the Wavelet Transform with the appropriate representations. We will elaborate on the global and local similarity measures in sections 4 and 5 respectively. In both sections examples we will given to account for the some of the most powerful abilities of the methods. Finally, a closing word will be given in section 6.

---

[1] In this paper we refer to *time series* rather than *signals*. However, the term 'signals' is exclusively used in the signal processing literature to which reference is made. Both mean one and the same thing - the sample (not necessarily uniform) record of data.

2   Global Versus Local Similarity

The global parameter which we would like to use for characterising (the rough-
ness of) the time series should not change if we estimate it for the first or the
second half or any arbitrary part of the time series, provided the characteris-
tics of the time series do not change in time (stationarity) or with the length
of the sample. The former requires that parameters remain stable with respect
to scaling within a considerable range of scales (scaling, self-affinity). A good
parameter indication of the similarity of the time series with its parts is the
exponent with which one has to re-scale the height of the (sliding) window with
the part of the time series in order to obtain a time series similar to the one
compared. In Figure 1 below, we illustrate this concept for the case of part of
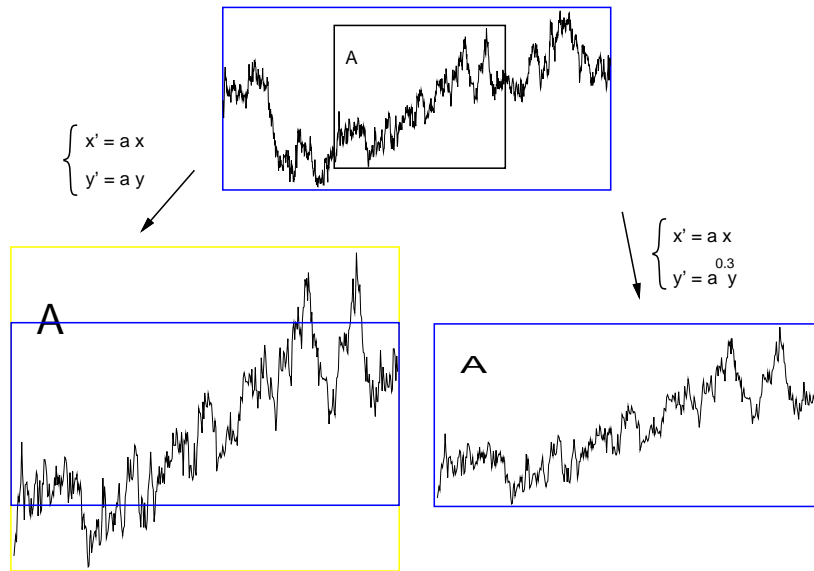the time series compared to the complete time series itself.

Fig. 1. Shows the horizontal versus vertical rescaling argument that the exponent char-
acterises the time series globally. *Similar* rescaling in the bottom left figure versus *affine*
rescaling, bottom right, of the fractional Brownian motion of $H = 0.3$. The rescaling
factor used for the affine rescaling of the (x, y) axis is $(a, a^{0.3})$, while for the similar
case both axes were rescaled using the a factor.

This concept of self-affinity and the related Hurst exponent H has been
developed within the domain of fractal geometry and is broadly applicable for
time series from sources in both natural and computer sciences. In particular, it
can be shown that the exponent $H = 0.5$ corresponds to the Brownian path (or
trail, or motion) - a random process with independent increments - the integral

of random noise. $H > 0.5$ is the evidence of a long range positive correlation in the time series, visually effecting a time series with tempered jumps. On the contrary, $H < 0.5$ gives evidence of a negative correlation, a so-called anti-correlation, which is displayed by more 'wild' behaviour. These processes are commonly refered to as the fractional Brownian motion (fBm).
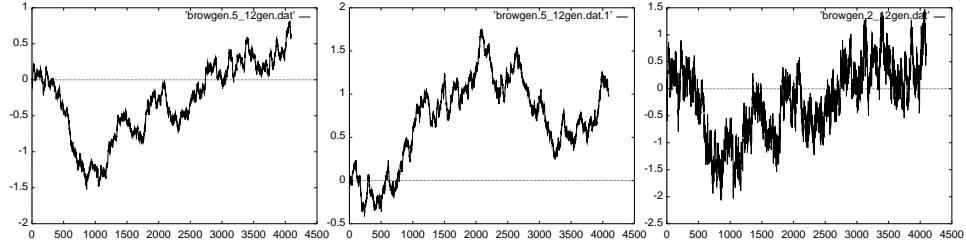

Fig. 2. Example time series, $H = 0.5$ for first two, $H = 0.2$ for the most right. Only the first two from the left are statistically similar. On the other hand, the first and the third time series are almost identical in detail if difference in scaling is neglected.

As a global measure, H can be successfully used to compare time series which are *statistically* similar, provided it is estimated in a trustworthy manner (including the removal of non-stationarities in the time series and providing the limits on scaling range and the realistic error bounds). For this purpose, we will use the wavelet transform which has been shown to be a particularly successful tool in assessing the scaling behaviour of time series [3].

Two arguments for the generalisation (of the global notion of similarity) suggest themselves; one is that we could allow the Hurst exponent to vary with position or we could be interested in local rather than global similarities between time series. Both require making the characteristics of the time series local in position. The relevant concept is known as the Hölder exponent h of the function in x_0 - if there exists a polynomial $P_n(x)$ of the degree n such that:

$$|f(x) - P_n(x - x_0)| \leq C|x - x_0|^h , \qquad (1)$$

then h is said to be the local Hölder exponent of the function and it characterises the scaling of the function locally for $n < h \leq n + 1$. The polynomial P_n corresponds to the Taylor series expansion of f around x_0 up to the order n.

The Wavelet Transform (which we will describe below) has been demonstrated to be a tool exceptionally well suited to the estimation of this exponent and in fact, as we will see later, global estimates like the Hurst exponent are obtained through this local Hölder exponent by means of taking an ensemble average in an appropriate partition function.

In conclusion, we have been able to identify two major approaches to stating similarity of time series; the global (statistical) similarity and the local, feature based similarity. In the following we will demonstrate how to approach both

classes with a common formalism based on the wavelet transform decomposition of time series.

3   Continuous Wavelet Transform and its Maxima Used to Reveal the Structure of the Time Series

As already mentioned above, the recently introduced Wavelet Transform (WT), see e.g. Ref. [8], provides a way of analysing local behaviour of functions. In this, it fundamentally differs from global transforms like the Fourier Transform. In addition to locality, it possesses the often very desirable ability of filtering the polynomial behaviour of some predefined degree.

Conceptually, the wavelet transform is a convolution product of the time series with the scaled and translated kernel - the wavelet $\psi(x)$ [2].

$$(Wf)(s,b) = \frac{1}{s} \int dx\ f(x)\ U(s,b)\,\psi(x)\ . \tag{2}$$

The scaling and translation actions are incorporated as the operator $U(s,b)$; 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

$$U(s,b)\psi(x) = \psi(\frac{x-b}{s})\ , \tag{3}$$

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

Figure 3 shows how the wavelet transform reveals more and more detail while going towards smaller scales. The wavelet transform is sometimes referred to as the 'mathematical microscope', due to its ability to focus on weak transient frequencies and singularities in the time series. The wavelet used determines the optics of the microscope; its magnification varies with the scale factor s.

The analysis of the local singular properties of a function (usually it is the singularities and not the smooth behaviour which are interesting in the time series) with the wavelet transform can be illustrated by the following. As already mentioned, the singularity strength is often characterised by the so-called Hölder exponent, compare Eq.1.

If we represent the function f through its Taylor expansion around $x = x_0$:

$$f(x)_{x_0} = c_0 + c_1(x - x_0) + \ldots + c_n(x - x_0)^n + C|x - x_0|^{h(x_0)}\ .$$

It follows directly that if $h(x_0)$ is equal to a positive integer n, the function f is n times continuously differentiable in x_0. Alternatively, if $n < h(x_0) < n+1$ the function f is continuous and singular in x_0. In that case f is n times differentiable, but its nth derivative is singular in x_0 and the exponent h characterises

---

[2] The power given to the normalising factor s, it is often chosen to serve a particular purpose. In this work, we choose a default factor s^{-1}, which conserves the integral $\int dx\,|\psi(x)|$ and thus leaves the L^1 measure invariant.
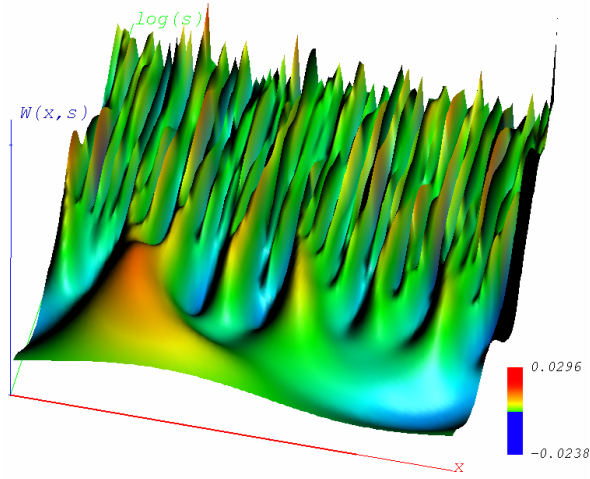
Fig. 3. WT representation of the time series from Figure 2 leftmost. The wavelet used is the Mexican hat.

this singularity. The exponent h, therefore, gives an indication of how regular the function f is in x_0, that is the higher the h, the more regular the function f.

The wavelet transform of the function f in $x = x_0$ with the wavelet of at least n vanishing moments, i.e. orthogonal to polynomials up to (maximum possible) degree n:

$$\int_{-\infty}^{+\infty} x^m \, \psi(x) \, dx = 0 \quad \forall m, \, 0 \leq m < n \,,$$

reduces to

$$W^{(n)} f(s, x_0) \sim C \int \psi(x) |s \, x|^{h(x_0)} \, dx \sim C \, |s|^{h(x_0)} \int \psi(x') |x'|^{h(x_0)} \, dx' \,.$$

Therefore, we have the following proportionality of the wavelet transform of the singularity $n \leq h \leq n + 1$, with the wavelet with n vanishing moments:

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

Thus the continuous wavelet transform can be used for detecting and representing the singularities in the time series even if masked by the polynomial bias. This ability is inherited by the more efficient representation based on modulus maxima of CWT, on which we will shortly expand in the following.

3.1 Wavelet Transform Modulus Maxima Representation and Bifurcation Representation

The continuous wavelet transform described in Eq. 2 is an extremely redundant representation. Therefore, other, less redundant representations, are frequently

used, including orthogonal representations and a variety of frames (almost orthogonal representations) [8]. For our purpose of comparison of features in the time series, one critical requirement is the translation shift invariance of the representation; no other than the boundary coefficients of the representation should change, if the time series is translated by some Δt. A useful representation satisfying this requirement and of much less redundancy than the CWT is the Wavelet Transform Modulus Maxima (WTMM) representation, introduced by Mallat [9]. It is derived from the CWT representation by extracting lines of maxima of the modulus of the wavelet transform. An example WTMM tree is shown in Figure 4, together with the highlighted bifurcations of the maxima lines [4].
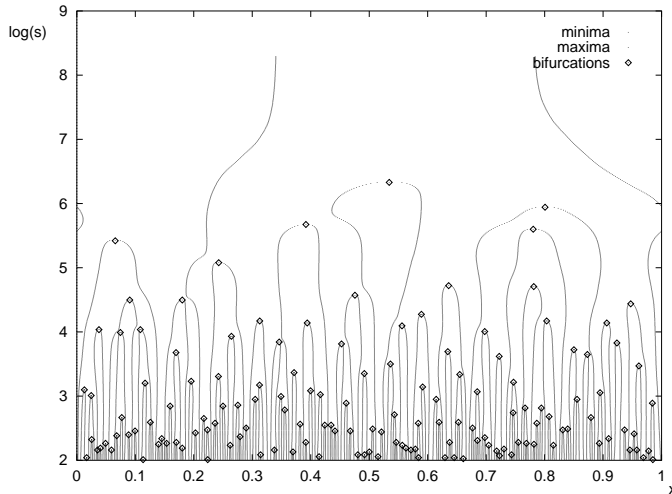

Fig. 4. WTMM representation of the time series and the bifurcations of the WTMM tree.

4 A Global (Statistical) Estimation of the Similarity of Time Series in Presence of Scaling, Translation and Polynomial Bias

It has been shown by Arneodo et al [3], that the WTMM tree is particularly useful for estimating the scaling parameters of functions. In particular, the Hurst exponent is related to the $q = $ 2nd moment (correlation) of the scaling of the measure on the WTMM maxima tree:

$$s^{2(H+P_n)-1} \sim \sum_{all \ maxima \ at \ scale \ s} \mu^2(s) \qquad (4)$$

where $\mu(s)$ is the amplitude of the maximum of the WT at the corresponding scale, and the sum - the partition function - is taken over all the maxima at the given scale s. P_n indicates the degree of the polynomial offset of the time series.

This relation (for $P_n = 1$) can be easily verified in Figure 5 where in log-log coordinates the power law relation 4 should result in a straight line.

We show the same, second moment for two examples of random noise and anti-correlated fractional noise. The Hurst exponent can easily be estimated from the slope of the linear fit to the scaling portion of the plot.
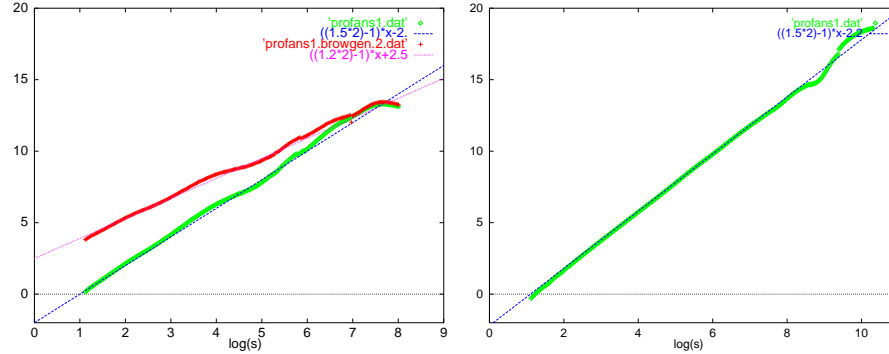

Fig. 5. Left: the second moments of the modulus maxima representation of the noise samples from Figure 2 leftmost and rightmost. Right: the same for the real life sample of a financial index shown in Figure 11 left.

In Figure 5 right, we show the same second moment estimated for the record of the financial index. It falls very well into the same category as the simulated Brownian path - indeed financial records are known to follow the H=0.5 law very closely, see e.g. [10]. The exponent H can thus be used to classify global similarity between time series or categorize them on the statistical grounds.

By design, H is limited to take values from the interval $0 \leq H \leq 1$. With the WTMM based formalism, we are able to estimate not only the fractional scaling part but also the degree of the underlying polynomial Pn. For fBm trails $Pn = 1$, for the noise record it would be $Pn = 0$. It is therefore more convenient to take the complete exponent $\beta = H + Pn$ as the (correlation) exponent representative to our time series.

With this exponent, we are able to distinguish between various categories of time series for $0 < \beta \leq 1$, $1 < \beta \leq 2$, $2 < \beta \leq 3$ or higher. As an illustration, see Figure 6, where we show the scaling of the WTMM correlation dimension evaluated for random noise, its integral - Brownian trail - and, again, its integral. For each integration step, the increase of the slope of the second moment is two, and the corresponding increase of the correlation dimension is one. Note that WTMM based formalism will correctly estimate the correlation exponent β only if the wavelet used has enough vanishing moments. In most practical situations, this condition is satisfied with $n = 2$ or $n = 3$. For example, $n = 1$ is enough for noise record $0 < \beta \leq 1$, like the leftmost example in Figure 6. But we need a wavelet with at least $n = 2$ for the $1 < \beta \leq 2$ class (central example in Figure 6) and with at least $n = 3$ for the $2 < \beta \leq 3$ class (rightmost example in Figure 6).
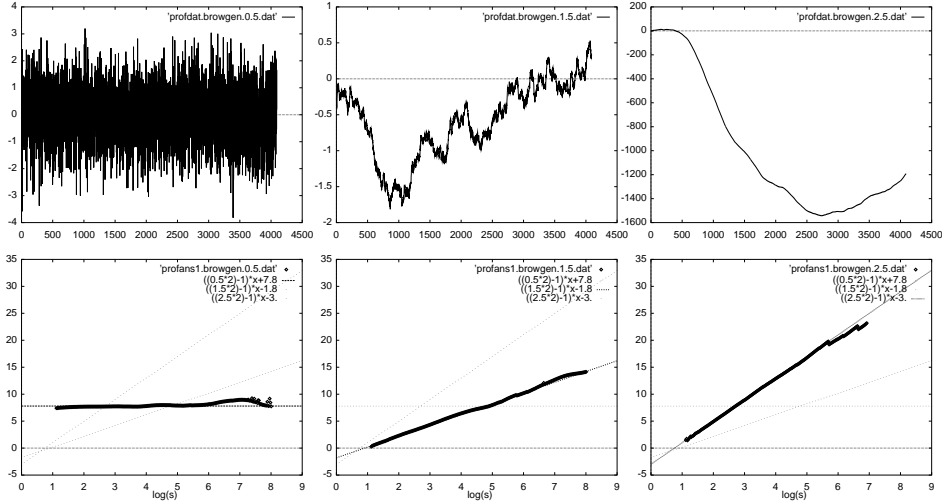
Fig. 6. Above, from the left: the random noise sample, its integral - Brownian trail and again its integral. Below from the left: corresponding scaling of the second moments of WTMM maxima.

5 Local Similarity Estimation in the Presence of Scaling, Translation and Polynomial Bias

In order to evaluate the local similarity, we will turn to local features of the time series, the bifurcations of the WTMM tree. Bifurcations [4] form a set of highly sensitive 'landmarks' in the WT landscape of the decomposed function. By reflecting the scale-position development of the maxima tree, they capture the singular structure of the time series. Each bifurcation can be represented with its position and scale coordinates, plus the corresponding value of the WT in the bifurcation point.

Just like the wavelet transform itself, the bifurcations can be evaluated for the time series up to a certain resolution, meaning that only the coarse features are taken into account. Alternatively, a range of scales can be determined in the application to be covered by the WTMM tree and its bifurcations. The numbers used for our experiments ranged from 20-100 bifurcations, covering the span of a maximum of two decades of scales from the highest resolution available.

We will use bifurcations obtained from the WTMM tree of the wavelet transform of the time series with the Mexican hat wavelet, the second derivative of the Gaussian kernel. This means that the maxima lines follow singular features in the second derivative of the function and the bifurcation representation reflects the structure of these features. This will allow for looking for similarities with respect to linear bias - such bias is filtered out from the time series by the wavelet with two vanishing moments. One can, of course, use wavelets with fewer or more vanishing moments to suit one's particular needs.

5.1 Local Distance (Similarity) Measure between Two (or More) Bifurcations

Essentially, the method uses the bifurcation representations of two time series to be compared and estimates the degree of similarity of these representations. Additionally one can shift both representations with respect to one another in order to find whether there is a better match if shift and scaling are involved. Note that a shift in the logarithmic scale corresponds with the scaling operation in the original time series.

The simplest but quite reliable measure of the similarity of two sets of bifurcations is given by the occurrence of a bifurcation in one representation, within a distance ϵ of some 'reference' bifurcation in the 'reference' set of bifurcations. Counting the fraction of reference bifurcations which have such a matching counterpart in the bifurcation set compared gives the estimate of the similarity between the two sets - a number from the range 0..1. A useful improvement of this scheme is easily made by counting only bifurcations in which WT has the same sign. This procedure, subject to one ϵ parameter only, gives good results.

A straightforward extension to the box of ϵ size is a two-dimensional correlation function which has a smooth decay of the distance between the bifurcations. As the measure of correlation between two bifurcation points: $biff_a(a, \sigma_b)$ and $biff_b(b, \sigma_b)$, we took for our experiments the auto-correlation function of two Gaussian kernels $C(biff_a, biff_b)$. This correlation can be parametrized by the additional shift in (logarithmic) scale and position Δs and Δx respectively, see [11] for more details.

With this measure for the correlation of two (or more) bifurcations, we can now estimate the total correlation of two bifurcation representations of the time series we want to compare. The most straightforward measure is simply

$$M(\Delta x, \Delta s) = \frac{1}{norm(N_1, N_2, \Delta x, \Delta s)} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} C(biff_i, biff_j)(\Delta x, \Delta s) \ , \quad (5)$$

where N_1 and N_2 are the respective numbers of bifurcations in both compared time series representations. Generally, *norm* should also be an appropriate function of the scale and position shift in order to compensate for the change in scale-position overlap of compared bifurcation representations.For more details on the measure and algorithm design, the reader can refer to [11].

In the rest of this section we will demonstrate the ability of the method to localise similarity in time series using the measure just designed, Eq. 5, to compare their intricate structure - the scale-position behaviour of the second derivative of the time series. This structure is captured by the bifurcation representation obtained with the wavelet orthogonal to linear bias ($n = 2$), which we will use in all the examples in the rest of this paper. It is, of course, possible to use a structure of a different derivative of the time series or the time series itself by taking a wavelet with an appropriate n - number of vanishing moments.

5.2 The Effect of Translation and Scaling

In this example, we demonstrate how similarities can be found for the time series which is scaled and translated. Let us take the example time series record, see Figure 7.

Using this example, we will demonstrate that the algorithm using the measure 5 to compare two bifurcation representations is capable of finding similarities between the time series (or their parts), with respect to the operations of translation and scaling.
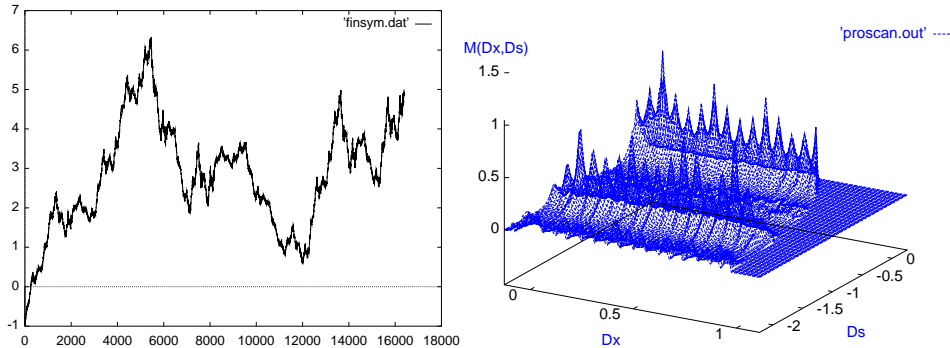

Fig. 7. Example fractal time series consisting of four self-affine sub-parts (left). The measure shows a high regularity and reveals the scaling and translation - elements of the invariance of the time series.

Instead of comparing one time series with another, in this test we will actually compare the time series with itself. Covering an adequate scope of values $\Delta x, \Delta s$ reveals in the measure $M(\Delta x, \Delta s)$ the presence of a very structured self-affinity relation within this time series - there are four main peaks located at $\Delta x = 0$, $\Delta x = 1/4$, $\Delta x = 1/2$ and $\Delta x = 3/4$. In between these, there are lower peaks, starting from the big peak at $\Delta x = 0$. The next smaller peak is at $\Delta x = 1/16$, followed by another at $\Delta x = 2/16$, and the next at $\Delta x = 3/16$. This sequence is repeated at $\Delta s = -1.39 = \log(1/4)$.

This, in fact, goes to show that we discovered that within the time series there are four similar parts, which in sequel contain four similar parts, etc. This similarity is evaluated with respect to the second derivative of the time series - the analysing wavelet is the Mexican hat - all the masking linear trend at different scales has been removed. Indeed, the test time series is an IFS fractal [12] with four non-overlapping self-affine transformations as the construction rule.

5.3 The Effect of Random Noise

Speaking of the influence of random noise has, of course, a very special meaning in the context of our analysis - most of the example test time series we considered

were records of pure or correlated noise. Still such noise is a perfectly valid and valuable time series (in fact a record which in absence of an appropriate model seems to be just an uncorrelated noise is likely to contain perfectly coded information). Therefore, in addition to comparing two different noise records, it makes perfect sense to evaluate how one noise record is corrupted by other noise. We took three examples.
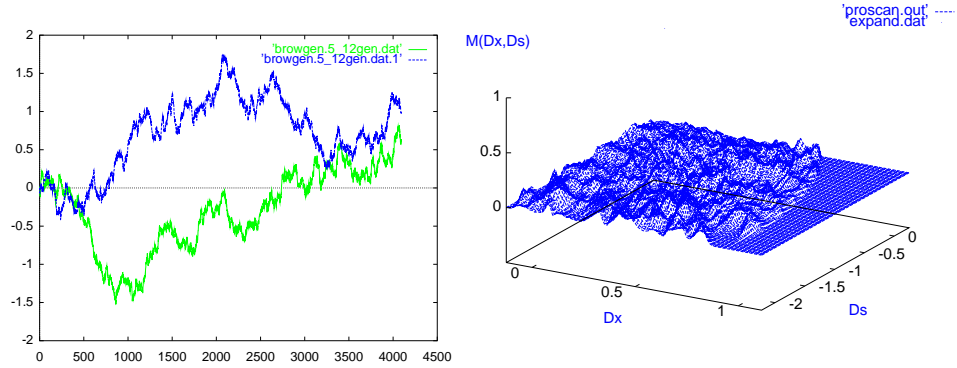

Fig. 8. Two independent random walks (left) investigated for the presence of similar parts with respect to scaling translation and linear bias show only slight local similarities at a residual level of measure fluctuations (right).

In the first, Figure 8, two independent random walks are scanned for similarities. The level of the measure remains low but significant within the searched range of scale and position shift. It can be considered as the fluctuations of the similarity measure reaching a significant level due to random occurrences of parts remotely looking like one another. These similarities are more likely to be assumed when going to higher $\Delta x, \Delta s$, due to the simple fact of considering the overlap of just a few bifurcations[11].
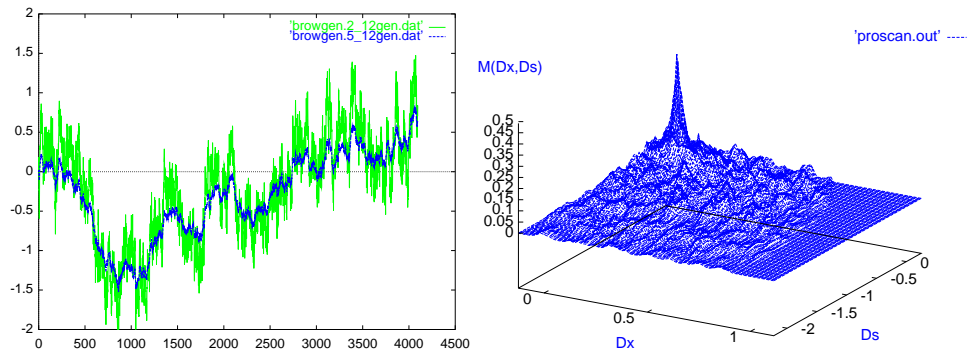

Fig. 9. Two noises with $H = 0.5$ and with $H = 0.2$ created with the same random sequence. Right, the similarity measure response of about 0.5 at no shift.

In the second example, figure 9, we consider two noises created with the same random sequence. One is uncorrelated with $H = 0.5$ and the other anti-correlated with $H = 0.2$. Both are created with the 'random midpoint displacement method', see e.g. [2], using the same random sequence, meaning the sign of displacement is consistent in both samples. Indeed, the similarity measure gives quite a good response, reaching about 0.5 at no shift.
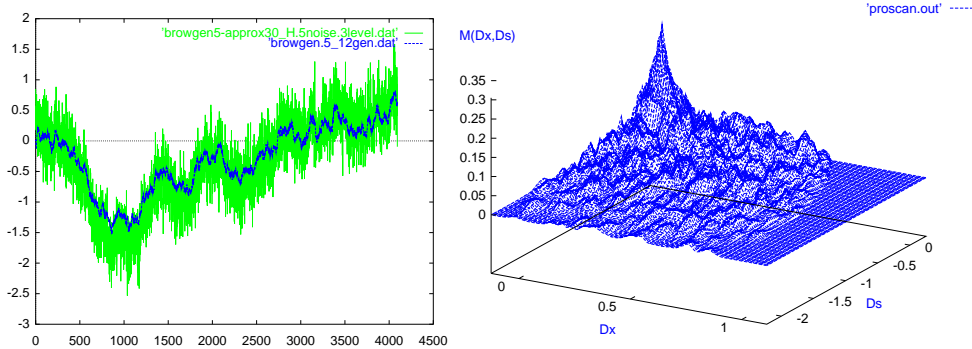

Fig. 10. Example random walk with the addition of random noise at $-10dB$ level. Right the resulting measure.

The third example, figure 10, demonstrates the influence of the additive random noise at a level of about $-10dB$ (amplitude factor 0.3). The result is a drop in the measure associated with a noticeable widening of the maximum cone. The resulting measure is at a level of one third of the measure norm for two identical time series.

5.4 The Real Life Sample within a Sample Example

As the last example, we took a real life time series. Two records of a financial index were scanned for the presence of similar parts with respect to scaling, translation and linear bias.

While from a visual inspection it is rather difficult to establish the degree of similarity between the two, the similarity measure reveals a high degree of similarity at $\Delta x = 0.3$ and $\Delta s = -1.4$. (This corresponds to a shift by $0.3 \cdot 4096 = 1229$ samples and rescaling by a factor $e^{-1.4} = 0.25$) [3]. At almost maximum level $C = 0.87$, it is far stronger than the rest of the measure. Applying the shift parameters to the second plot confirms a close fit, but only after the linear trend is restored! (Indeed, the second time series is just a part of the first with added linear bias.)

---

[3] We added a relatively strong bias in order to illustrate our purpose. The method will, of course, also work for smaller levels of bias for which a visual check of the similarity of both time series will be possible. Still, even in such cases, techniques without the ability to filter polynomial trends are likely to fail.
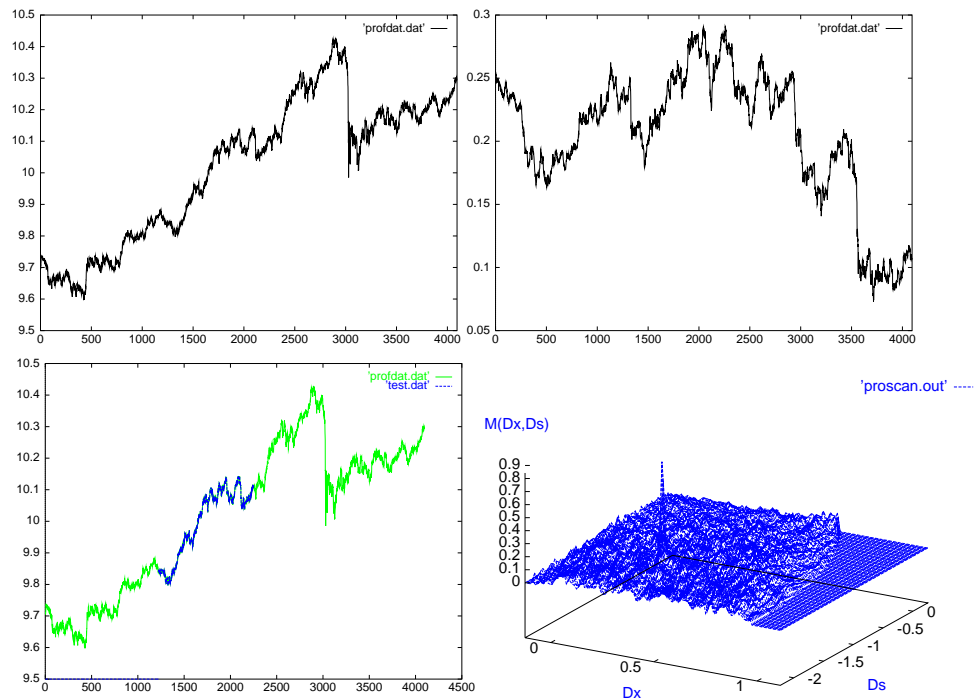
Fig. 11. Above, two example real life time series subject to investigation for the presence of similar parts with respect to scaling, translation and linear bias. Left below, the match revealed in the maximum of the measure shown at right below.

6 Final Remarks and Conclusions

Determining similarity of time series is an important problem in data-mining. We presented a powerful framework allowing construction of specific similarity criteria, in the presence of scaling, translation and polynomial bias. Two main classes of similarity evaluation measures were distinguished and the appropriate measures were proposed.

The global, statistical similarity was estimated with the Wavelet Transform derived Hurst exponent. It classifies time series according to their global scaling properties. The local, detail oriented measure, used the scale-position bifurcation representation of the wavelet modulus maxima transform of the time series. It makes it possible to obtain good matches of (the parts of) the time series with respect to scaling translation and polynomial bias. The degree of the polynomial bias filtered can be affected as well as the range of the translation and scaling parameters. The measure used for matching the two bifurcation representations of the time series can also be adapted to the specific user requirements.

This flexibility represents an important difference from existing methods like those presented, for example, in [6, 7, 5]. Let us, in closing, once again stress that the framework presented does not impose a particular model, but facil-

itates implementing models to suit one's needs. We demonstrated this in the two example similarity measures. Since they represent two extremes, it is, of course possible, that both the global and the local measure can be used together with appropriate weighting factors in order to satisfy an intermediate objective. However, the capabilities of the presented framework extend considerably beyond just modifying or merging the example applications presented. Due to the ability to represent the analysed time series in a unique set of position-space localised features (we presented bifurcations but using other features is possible), one can design a variety of (hierarchical) matching algorithms. The features used in such algorithms are dependent on the wavelet used and can inherit all of the advantages presented, such as selective sensitivity to polynomial trends. This is also the direction of current research about which we will be communicating shortly. In addition to this, we will consider issues of increasing the accuracy of the representations with compactly supported wavelets and optimization of the algorithms for speed.

Apart from developing the methodology, our work proceeds towards the particular goal of exploring time series from data-mining applications as described in the introduction.

References

1. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, Eds., *Advances in Knowledge Discovery and Data Mining*, AAAI Press/MIT Press, 1996.
2. K. Falconer, *Fractal Geometry - Mathematical Foundations and Applications*, John Wiley (1990).
3. A. Arneodo, E. Bacry, J.F. Muzy, The Thermodynamics of Fractals Revisited with Wavelets, *Physica A*, **213**, 232-275, (1995).
4. Z.R. Struzik, The Wavelet Transform in The Solution to the Inverse Fractal Problem, *Fractals* **3** No.2, 329-350 (1995).
5. R. Agrawal, K-I. Lin, H.S. Sawhney, K, Shim, Fast Similarity Search in the Presence of Noise, Scaling and Translation in Time-Series Databases, in *Proceedings of the 21 VLDB Conference*, Zürich, 1995.
6. G. Das, D. Gunopulos, H. Mannila, Finding Similar Time Series, In *Principles of Data Mining and Knowledge Discovery*, Lecture Notes in Artificial Intelligence 1263, Springer, 1997.
7. R. Agrawal, C. Faloutsos, A. Swami. Efficient Similarity Search in Sequence Databases, in *Proc. of the Fourth International Conference on Foundations of Data Organization and Algorithms*, Chicago, 1993.
8. I. Daubechies, *Ten Lectures on Wavelets*, S.I.A.M. (1992).
9. S.G. Mallat, S. Zhong, Complete Signal Representation with Multiscale Edges, *IEEE Trans. PAMI* **14**, 710-732 (1992).
10. C.J.G. Evertsz, Fractal Geometry of Financial Time Series, *Fractals* **3** No.3, 609-616 (1995).
11. Z.R. Struzik, A.Siebes, Wavelet Transform in Similarity Paradigm I - CWI Technical Report, *CWI Reports* INS-R 9802, (1998).
12. M.F. Barnsley, *Fractals Everywhere*, Academic Press, NY, (1988).