Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

INS

Information Systems

INformation Systems

INS

Rule discovery: Tough, not meaningless

Z.R. Struzik

REPORT INS-R0304 APRIL 30, 2003

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

Rule Discovery: Tough, not Meaningless

Zbigniew R. Struzik
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
email: Zbigniew.Struzik@cwi.nl

ABSTRACT

'Model free' rule discovery from data has recently been subject to considerable criticism, which has cast a shadow over the emerging discipline of time series data mining. However, other than in data mining, rule discovery has long been the subject of research in statistical physics of complex phenomena. Drawing from the expertise acquired therein, we suggest explanations for the two mechanisms of the apparent 'meaninglessness' of rule recovery in the reference data mining approach.

One reflects the universal property of self-affinity of signals from real life complex phenomena. It further expands on the issue of scaling invariance and fractal geometry, explaining that for ideal scale invariant (fractal) signals, rule discovery requires more than just comparing two parts of the signal. Authentic rule discovery is likely to look for the possible 'structure' pertinent to the failure mechanism of the (position and/or resolution-wise) invariance of the time series analysed.

The other reflects the redundancy of the 'trivial' matches, which effectively smoothes out the rule which potentially could be discovered. Orthogonal scale space representations and appropriate redundancy suppression measures over autocorrelation operations performed during the matches are suggested as the methods of choice for rule discovery.

2000 Mathematics Subject Classification: 28A80, 68T10, 68P10
1998 ACM Computing Classification System: H1, I5, Jm, J2, E2
Keywords and Phrases: novelty detection, surprise detection, data mining, complex systems, multifractals, inverse problem, auto-correlation, Wiener Khinchin theorem
Note: Work carried out under the project INS1.1 "Data Mining and Knowledge Discovery"
A full colour version of this paper can be downloaded from www.cwi.nl/~zbyszek

1. Introduction: the big failure of 'rule discovery' *or* the most likely and correct result?

This article has been stimulated by the considerable criticism of the mainstream rule discovery algorithm in data mining [1]. The overwhelming conclusions of the article would be disastrous for the domain of research in questions if they lacked explanation and understanding.

The purpose of our paper is to propose a closer look at the possible and plausible causes for the devastating result reported in [1]. Additionally, a framework of established scientific research will be identified, which encompasses the problem definition and provides both theoretical and practical foundations which can reinstate model free rule discovery in data mining in the domain of sound science.

At a first glance, the preprint [1] by performing scrutiny testing suggests that the discussed algorithm [2] based rule discovery does not produce meaningful rules. In particular, the confidence of the rules recovered is not to be distinguished from the rules obtained from random noise.

The primary investigated example, coinciding with the example used by the primarily criticised paper by Das et al [2], is that of the S&P500 financial index. However, even to the casual observer, the main conclusion drawn from such an example must seem weak. The evidence for the lack of correlations in a financial time series like the S&P 500 index is so overwhelming that the 'meaninglessness' of any deterministic rule discovered may not seem surprising. Indeed, the authors of [1] suggest that there is no more confidence in the particular rule advocated in [2] than in any other deterministic rule. Thus any rule might do, which in actual fact means that such a rule is useless and irrelevant, holding at random, statistically meaningless instances. The mechanism of proving this conclusion has been devised by comparing the rule discovery algorithm from [2] on both the test time series (S&P500) and the surrogate time series (random walk).

The particular conclusion may not be that surprising when one consults the vast bulk of work in the 'econophysics' branch of statistical physics of complex phenomena [3, 4, 5, 6, 7, 8]. It belongs to the facts well established, the absence of correlations or 'memory' in S&P500 or such financial time series except for at very small time scales at the range of single minutes. The lack of correlations, thus the lack of apparent determinism and easy rule discovery in the raw time series, is not equivalent to Gaussian statistics. In fact, the analyses performed and still ongoing focus on the deviations from Gaussian behaviour [5, 6, 9, 10]. Such deviations, embedded in the distributions explaining data may in fact provide some statistical grounds for accommodating rare events like extreme crashes. They are, however, unlikely to provide deterministic rules of temporal market behaviour, being global and statistical in nature.

Within the same scope of econophysics work, however, the absolute jumps of the S&P index - the so-called volatility - have been shown to be substantially correlated. And thus they possess a meaningful range memory and are amenable to 'rule discovery'. Indeed, the volatility is governed by an entire multiscale 'cascade of information' flow, which can be detected in the volatility of financial time series [4].

Thus if the only purpose of Ref [1] were to show that blindly extracted rules from a fully developed financial index are meaningless, such conclusions would not be particularly innovatory. The article does, however, suggest that the same degree of meaninglessness is obtained no matter what input time series is used. The primary cause attributed to this failure is not in the clustering algorithm which is the only rule extraction mechanism investigated but in the pre-processing of the time series. In particular the 'moving window' overlapping selection of candidate time series intervals leads to so-called spurious matches, destroying the resolution of the clustering algorithm.

The purpose of this writing is to look closer at the likely cause for the inability of the algorithms discussed blindly to extract rules from real life time series. In particular, the issue of scale invariance will be addressed, which characterises not only an overwhelming range of real-life and artificial time series but can also be attributed to isolated singularities - often the building blocks of the real-life and artificial time series.

Additionally, scaling invariance will be linked to the rate of auto-correlation decay, which determines the impact of 'redundant' spurious matches on the blind clustering algorithms. While auto-correlation decay is considered an important diagnostic tool in the study of long range dependence (on a par with power spectrum decay to which it is equivalent), for the purpose of blind clustering only the extrema (maxima or minima) of the autocorrelation (or the local match) may need to be considered to provide rule extraction with sufficient resolution and sensitivity.

2. Redundant information → spurious matches

The primary cause of the meaninglessness of the rule discovery seems to be attributed by the authors of the critical work [1] to the shortcoming of the time series pre-processing algorithm and in particular to the redundant or trivial matches (the redundancy of the matching operation.) Such trivial matches of two time-series intervals shifted with respect to one another by a time lag will indeed in many cases show a slow rate of decorrelation. While the presence of long range correlations in the time series will obviously affect the selectivity of any rule recovery, this effect is relatively straightforward to explain and deal with.

Apart from the entire zoo of possible distance measures, the standard way of calculating the inner product of two time series is used for evaluating their 'correlation' level. For the time lag t shifted versions, the definition of the autocorrelation product/function $C(t)$ of a function $f(t)$ reads:

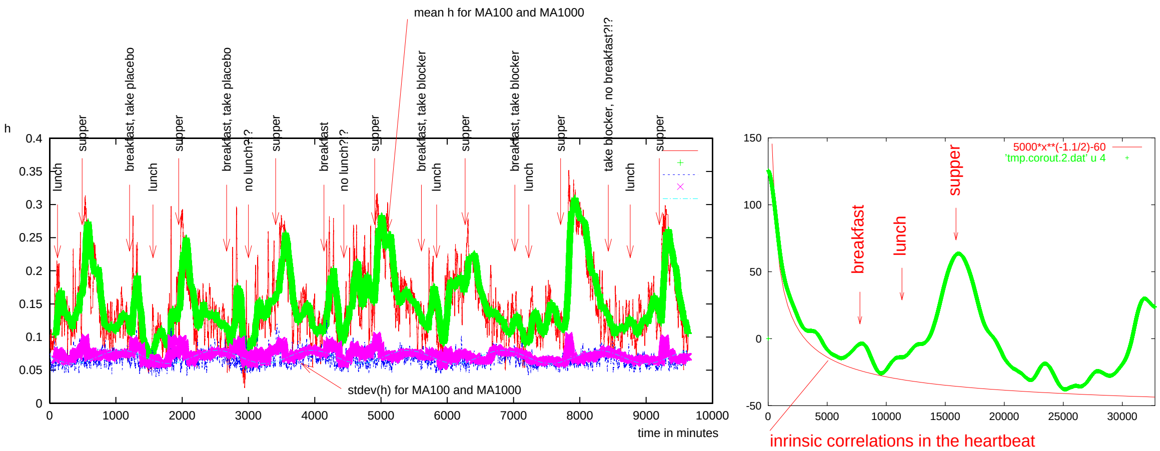

Figure 1: Left: the plot of the variability of human heartbeat from a seven day long experiment where the test persons were given placebo or beta-blocker. For the variability estimate, local roughness (local correlation) exponent h is used, smoothed with a moving average (MA) filter with 100 and 1000 long window. An interesting pattern of response to food is evident [11]. Right: autocorrelation function confirms the presence of an invariant, intra-day periodic structure. The autocorrelation plot is in fact the autocorrelation of the local correlation exponent (as described with the Hölder h exponent.)

$$C(t) = \int_{-\infty}^{\infty} \bar{f}(\tau)\ f(t+\tau)\ d\tau \tag{2.1}$$

where $\bar{f}(\tau)$ is the complex conjugate of $f(t)$. Amazingly, the autocorrelation is simply given by the Fourier transform $\mathcal{F}$ of the absolute square of $f(t)$:

$$C(t) = \mathcal{F}(|f(t)|^2)\ , \tag{2.2}$$

and, of course, the Fourier transform of the second moment of the function is nothing else than its power spectrum $P(\omega) = \mathcal{F}(|f(t)|^2)$. This relationship is known by the name of the Wiener Khinchin theorem.

Thus, interestingly, Fourier power spectrum is also related to the likely cause of the inability of the rule discovery to be selective enough in its pre-processing phase (feeding the rule extraction algorithms.) The importance of this in the context of our discussion lies in the fact that it links the scaling properties of the Fourier power spectrum with the decay rate of the auto-correlation function. Thus any property of the scaling invariance as discussed above will reveal itself in the invariance of the auto-correlation function. In particular it will also determine the rate of decay of the auto-correlation function and will be inherited by the cross-correlation products of the time series with its sub-parts.

The sub-part matching operation is the key operation used in the rule discovery algorithms [2, 1] and it clearly inherits the self-similarity properties of the time series. The explanation of the 'trivial match' redundancy which contributes to the inability of the algorithm to select sound rules thus comes from the spectral properties of the time series. The same spectral properties which as we have shown above, describe self-similarity properties of the time series.

It is worth noting that the autocorrelation decay is an important diagnostic tool widely used in investigating long range dependence (correlations), see e.g Ref [4] in the context of S&P500 analysis. However, due to the equivalence shown above, power spectrum decay has been extensively investigated in the same context. A modern method which allows local multiscale or multi-resolution decomposition as opposed to the global Fourier approach is the recently introduced wavelet transform. It allows local location-wise (temporal) and scale-wise (frequency) extraction of required information, including moments of the decomposition measure and regularity (scaling) exponents.

3. Estimating regularity properties of rough time series

The advent of multi-scale techniques (like WT), capable of locally assessing the singular behaviour, greatly contributed to the advance of analysis of 'strange' signals, including (multi)fractal functions and distributions. The wavelet transform [12, 13, 14, 15] is a decomposition of the input time series into the discrete or continuous basis of localised (often compactly supported) functions - the wavelets. This decomposition is defined through an inner product of the time series with the appropriately rescaled and translated wavelet of a fixed shape. Wavelet decomposition schemes exist which allow decomposition optimisation through the choice from various wavelet bases [16, 17] or adaptive decomposition (notably the lifting scheme [18]).

In the continuous formulation, the wavelet transform can be seen as a convolution product of the signal with the scaled and translated kernel - the wavelet $\psi(x)$. The scaling and translation actions are performed by two parameters; the scale parameter s 'adapts' the width of the wavelet kernel to the resolution required and the location of the analysing wavelet is determined by the parameter b:

$$(Wf)(s,b) = \frac{1}{s} \int dx \; f(x) \; \psi(\frac{x-b}{s}) \tag{3.1}$$

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

For analysis purposes, one is not so much concerned with numerical or transmission efficiency or representation compactness, but rather with accuracy and adaptive properties of the analysing tool. Therefore, in analysis tasks, continuous wavelet decomposition is mostly used. The space of scale s and position b is then sampled semi-continuously, using the finest data resolution available. The numerical cost of evaluating the continuous wavelet decomposition is not as high as it may seem. Algorithms have been proposed which (per scale) have a complexity of the order n, the number of input samples, at a relatively low constant cost factor [19]. Additionally, computationally cheap, discretised, semi-continuous versions of the decomposition are possible [20, 21].

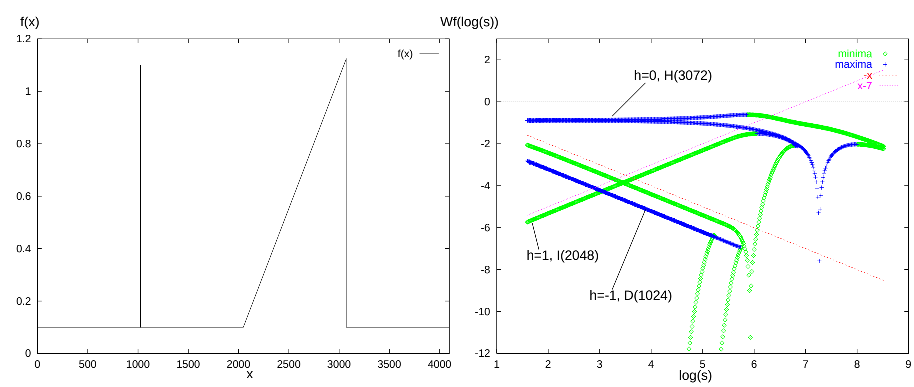

Figure 2: The local scaling exponent h can be estimated for isolated singularities - and it shows persistent scaling invariance. Left: the test signal consisting of the Dirac pulse $D(1024)$, the change in slope - integrated Heaviside step I(2048), and the Heaviside step H(3072). Right: the log-log plot of the maxima corresponding to all three singularities: $D(1024)$, I(2048) and H(3072). Theoretical slopes are also indicated; these are -1 for $D(1024)$, 1 for I(2048) and 0 for H(3072). The wavelet used is the Mexican hat.

Suppose, we can *locally* approximate the time series (function f) with some polynomial P_n, but the approximation fails for P_{n+1}. One can think of this kind of approximation as the Taylor series decomposition:

$$\begin{aligned} f(x)_{x_0} &= c_0 + c_1(x-x_0) + \cdots + c_n(x-x_0)^n + C|x-x_0|^{h(x_0)} = \\ &= P_n(x-x_0) + C|x-x_0|^{h(x_0)} \; . \end{aligned}$$

The exponent $h(x_0)$ characterises such local singular behaviour by capturing what 'remains' after approximating with P_n and what does not yet 'fit' into an approximation with P_{n+1}. Thus, our function or time series $f(x)$ is locally described by the polynomial component P_n and the so-called Hölder exponent $h(x_0)$.
It can be shown that we have the following power law proportionality for the wavelet transform (WT) of the (Hölder) singularity of $f(x_0)$:

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

From the functional form of the equation, one can attempt to extract the value of the local Hölder exponent from the scaling of the wavelet transform coefficients in the vicinity of the singular point x_0. A common approach to tracing such singularities and to revealing the scaling of the corresponding wavelet coefficients is to follow the so-called maxima lines of the WT, converging towards the analysed singularity.

In figure 2 three trivial structures have been shown containing localised and isolated singularities of Hölder type. (Note, that singularities exist which are not localised - e.g. oscillating singularities [12, 22].) Such simple objects are often used as building blocks for artificial signals. The singularities involved have been analysed with the wavelet transform and their decay rate of the power spectrum reveals perfect scaling invariance below some threshold resolution of the characteristic frequency. Such scaling is indistinguishable from that of more complex structures unless more advanced tests are used; e.g Heaviside related exponent $h = 0$ is very close to the mean H for the human heartbeat.

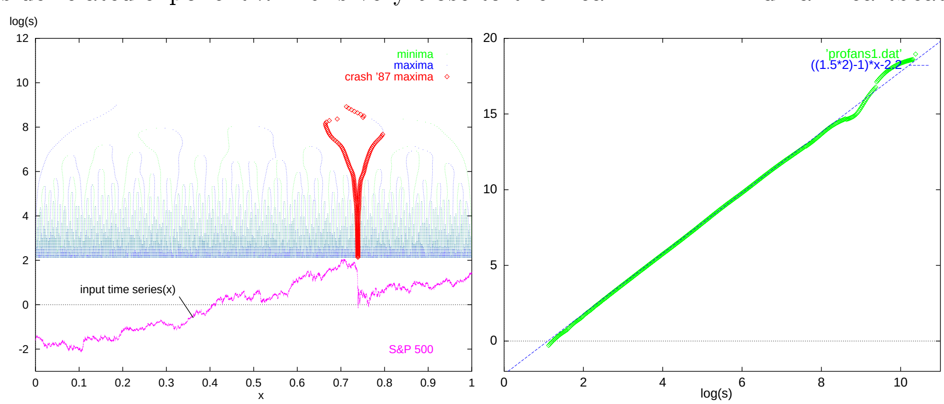

Figure 3: Left: the L1 normalised S&P500 index time series with the derived WT maxima tree above it in the same figure. The strongest maxima correspond to the crash of '87. Right: the second moment of the partition function over the entire CWT (thus not only the maxima lines) see Eq. 4.1, shows consistent scaling invariance with the exponent $H + 1 = 1.5$. This corresponds with the Brownian walk scaling invariance exponent at $H = 0.5$.

In figure 3, we plot the input time series which is a part of the S&P index containing the crash of '87. In the same figure, we plot corresponding maxima derived from the WT decomposition with the Mexican hat wavelet. The maxima converging to the strongest singularity - the '97 crash have been highlighted in the top view.

4. Multiscaling or multifractality

We have chosen the WTMM method [23] for its many merits. In addition to local regularity estimation, local and scale-wise feature extraction (maxima lines, see fig 3), the WTMM tree lends itself very well to defining the partition function based multifractal analysis [23]. The moments q of the measure distributed on the WTMM tree are taken to obtain the dependency of the scaling function $\tau(q)$ on the moments q:

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

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

$$\mathcal{Z}(s,q) = \sum_{\Omega(s)} (Wf\omega_i(s))^q \ , \tag{4.1}$$

with $\Omega(s) = \{\omega_i(s)\}$ is the set of maxima $\omega_i(s)$ at the scale s of the continuous wavelet transform $Wf(s,t)$ of the function $f(t)$. Non-linear dependency of the scaling exponent $\tau(q)$ on the moments q is the hallmark of statistical multiscaling (multifractality), while linearity corresponds with the monofractal (monoscaling) character of the analyzed process.

There is an ultimate link between the global scaling exponent: the Hurst exponent H (compare figure 4), and the Fourier power spectral exponent. The power spectrum of the input signal and the corresponding scaling exponent γ can be directly evaluated from the second moment of the partition function ($q = 2$) in Eq. 4.1. Indeed the wavelet transform decomposes the signal into scale (and thus frequency) [1] dependent components (scale and position localised wavelets), comparably to frequency localised sines and cosines based Fourier decomposition, but with added position localisation. Scaling of the second moment of the decomposition coefficients provides γ, the power spectrum scaling, through $\gamma = 2H + 1$, the relation which links the spectral exponent γ with the Hurst exponent H.

5. Conclusion: rules within rules or the principle of (statistical) self-similarity

Contrary to the long and widely accepted (Euclidean) view, real world time series/ signals are not smooth. On the contrary, they are often non-differentiable and densely packed with singularities. The often adopted view that these signals consist of some smooth information carrying a component with superimposed noise is also very often inaccurate. Real life records are not necessarily contaminated by 'noise'. Instead, in the case of the lack of a better model, they often intrinsically consist exclusively of noise - indeed, they are 'noise' themselves.

Their 'noisy' components are often distributed at various resolution and length scales - in other words each sub-part of the record is equally noisy and statistically similar (after affine rescaling of x,y coordinates with some factors β_x,β_y) to the entire record (or any other subpart). This kind of similarity can often be characterised by one single exponent $h = \log(\beta_y)/\log(\beta_x)$ for a range of β rescaling factors. This is the concept of local scaling which has been explored with the wavelet transform local scaling estimates in section 3. Additionally, the local scaling is often isotropic and the same one unique exponent can characterise both global and local ratio of the similarity rescaling. E.g. this is the case for 1-dim Brownian walk - the integral of white noise for which $h = 0.5$, and equals global $H = 0.5$, the so-called Hurst exponent. Such global scaling rules have been addressed through the partition function multifractal fromalism in section 4.

Indeed the very essence of scaling, ie. scale invariance, has the consequence that statistically similar patterns may occur at any resolution or scale length. This property characterising many real life signals

[1] The working scale of the wavelet s is inversely proportional to the (Fourier) frequency $f \sim 1/s$ and the continuous wavelet used is the second derivative of the Gaussian curve (*Mexican hat*).

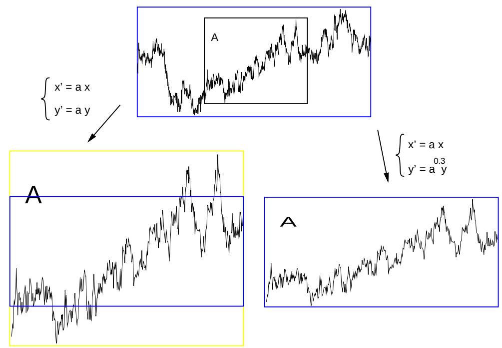

Figure 4: The principle of self-affinity; *Similar* rescaling in the bottom left figure versus *affine* rescaling, bottom right, of the fractional Brownian motion of $H = 0.3$. The rescaling factor used for the affine rescaling of the (x, y) axis is $(a, a^{0.3})$, while for a similar case both axes were rescaled using the a factor.

may be behind the limited ability to extract meaningful deterministic 'rules' from such records [1], although it does permit statistical rule discovery [4, 24]. Such can be used for distance evaluation for detecting rule violation for whole time series or streaming diagnosis etc, or streaming time series novelty assessment through departure from the model 'rule'.

Apparently such signals would seem ultimately pathologic and not lend themselves to analysis or novelty or 'rule' discovery. This does not have to be true [25]. Rough and wildly changing records are ever-present in nature [26, 27, 28, 29, 30] and have long been analysed with Fourier spectral methods. However, only recent progress in multiscale modelling (and analysis) methods based on fractal and scaling (or multiscaling) concepts, aided by analysis methods such as wavelets, has achieved substantial progress in understanding such time series and the underlying phenomena. In particular, local analysis of rough time series has become possible through modern methods of regularisation, notably the wavelet transform [12, 13, 14, 15].

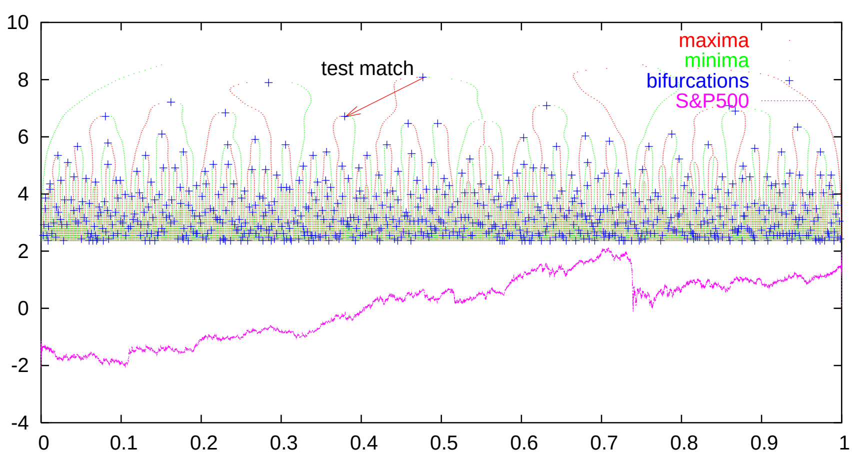

Figure 5: Investigating possible rules may be done by investigating structure pertinent to the invariant, orthogonal multiscale representation like the structure of the bifurcation of the maxima. Testing matches between sub-trees of the maxima representation of the wavelet transform of the S&P500.

In conclusion, there is nothing really surprising in the result of [1] discussion of the algorithm of [2], however important it is for revealing the inherent problems which have been left unnoticed to date. The pressing question of course remains, what then is the meaningful methodology/strategy for dealing with signals inherently tough for rule discovery. The answer resides, in our opinion as outlined, in the spectral and auto-correlation properties of the time series. The rules which can be detected are instances of invariance violation, This can be manifested in the non-stationarity of spectral characteristics, be it a short-time power spectrum or multifractal spectrum. Or alternatively and simultaneously, in the breakdown of the scaling invariance - the structure which potentially emerges from the tree of the wavelet transform maxima [25, 31, 32, 33, 34], see e.g figure 5 right, or possibly the structure emerging from the conduct of the self-adapting mechanism in multiscale/multiresolution decomposition or approximation bases [16, 17, 18, 35].

References

1. J. Lin, E. Keogh, W. Truppel, When is Time Series Clustering Meaningful?, *preprint* Workshop on Clustering High Dimensional Data and its Applications, SDM 2003.
   will appear on the workshop site: `www.cs.utexas.edu/users/inderjit/sdm03.html`
2. G. Das, K. Lin, H. Mannila, G. Renganathan, P. Smyth, Rule Discovery from Time Series, in proceedings of the ..th Intl. Conference on Knowledge Discovery and Data Mining, New York, NY,Aug 27-31, 1998, pp 16-22, (1998).
3. R.N. Mantegna and H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* Cambridge, England: Cambridge University Press, (2000).
4. A. Arneodo, J.F. Muzy, D. Sornette, Eur. Phys J. B, **2**, 277 (1998). `\\http://xxx.lanl.gov/ps/cond-mat/9708012`
5. A. Johansen, D. Sornette, Stock Market Crashes are Outliers, Eur.Phys.J. B 1, pp. 141-143 (1998). A. Johansen, D. Sornette, Large Stock Market Price Drawdowns Are Outliers arXiv:cond-mat/0010050, 3 Oct 2000, rev. 25 Jul 2001.
6. B. Podobnik, P.Ch. Ivanov, Y. Lee, and H.E. Stanley. “Scale-invariant Truncated Lévy Process”. Europhysics Letters, **52** pp 491-497, (2000).
7. Z. R. Struzik. Wavelet Methods in (Financial) Time-series Processing. Physica A: Statistical Mechanics and its Applications, 296(1-2):307-319, June 2001.
8. D. Sornette, Y. Malevergne, J.F. Muzy, Volatility Fingerprints of Large Shocks: Endogeneous Versus Exogeneous, arXiv:cond-mat/0204626, (2002).
9. K. Matia, Y. Ashkenazy, H.E. Stanley, Multifractal Properties of Price Fluctuations of Stocks and Commodities, Europhysics Letters,**61**, (3), pp. 422-428 (2003).
10. X. Gabaix, P. Gopikrishnan, V. Plerou, H.E. Stanley, Understanding Large Movements in Stock Market Activity,
    X. Gabaix, P. Gopikrishnan, V. Plerou, H.E. Stanley, A Simple Theory of the 'cubic' Laws of Financial Fluctuations
11. Z. R. Struzik. Revealing Local Variablity Properties of Human Heartbeat Intervals with the Local Effective Hölder Exponent. Fractals **9**, No 1, 77-93 (2001).
12. S. Jaffard, Multifractal Formalism for Functions: I. Results Valid for all Functions, II. Self-Similar Functions, *SIAM J. Math. Anal.*, 28(4): 944-998, (1997).

13. I. Daubechies, *Ten Lectures on Wavelets*, (S.I.A.M., 1992).

14. M. Holschneider, *Wavelets - An Analysis Tool*, (Oxford Science Publications, 1995).

15. S.G. Mallat and W.L. Hwang, Singularity Detection and Processing with Wavelets. *IEEE Trans. on Information Theory* **38**, 617 (1992).
S.G. Mallat and S. Zhong Complete Signal Representation with Multiscale Edges. *IEEE Trans. PAMI* **14**, 710 (1992).

16. S. Mallat, Z. Zhang, Matching pursuit in a time-frequency disctionary, *IEEE Transactions on Signal Processing*, **41** pp. 3397-3415, (1993).

17. R.R. Coifmann M.V. Wickerhauser, Entropy-based Algorithm for Best-basis Selection. *IEEE Transactions on Information Theory*, **38**, pp. 713-718, (1992).

18. W. Sweldens, The Lifting Scheme: Construction of Second Generation Wavelets, *SIAM J. Math. Anal.* **29**, (2), pp 511-546, (1997).

19. A. Muñoz Barrutia, R. Ertlé, M. Unser, Continuous Wavelet Transform with Arbitrary Scales and O(N) Complexity, Signal Processing, vol 82, no. 5, pp. 749-757, May 2002

20. M. Unser, A. Aldroubi, S.J. Schiff, Fast Implementation of the Continuous Wavelet Transform with Integer Scales, IEEE Transactions on Signal Processing, vol. 42, no. 12, pp. 3519-3523, December 1994.

21. Z. R. Struzik, Oversampling the Haar Wavelet Transform. Technical Report INS-R0102, CWI, Amsterdam, The Netherlands, March 2001.

22. A. Arneodo, E. Bacry, J.F. Muzy, Oscillating Singularities in Locally Self-Similar Functions, *PRL*, **74**, No 24, 4823-4826, (1995).

23. 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).

24. A.C.-C.Yang, S.-S. Hseu, H.-W. Yien, A.L. Goldberger, C.-K. Peng, Linguistic Analysis of the Human Heartbeat using Frequency and Rank Order Statistics, *PRL*, in press, (2003).

25. Z.R. Struzik, Taming Surprises, in proceedings of the *New Trends in Intelligent Information Processing and Web Mining* conference, Zakopane June 2-5, (2003).
Z.R. Struzik, Looking for Surprises in a Rough World, *preprint*, (2003).

26. K. Falconer, Fractal Geometry: Mathematical Foundations and Applications, John Wiley, 1990; paperback 1997.

27. A. Arneodo, E. Bacry, J.F. Muzy, Wavelets and Multifractal Formalism for Singular Signals: Application to Turbulence Data, *PRL*, **67**, No 25, 3515-3518, (1991).

28. H.E. Stanley, P. Meakin, Multifractal Phenomena in Physics and Chemistry, Nature, vol 335, 405-409, (1988)

29. P.Ch. Ivanov, M.G. Rosenblum, L.A. Nunes Amaral, Z.R. Struzik, S. Havlin, A.L. Goldberger and H.E. Stanley, Multifractality in Human Heartbeat Dynamics, *Nature* **399**, 461-465, (1999).

30. A. Bunde, J. Kropp, H.J. Schellnhuber, (Eds), The Science of Disasters, Climate Disruptions, Heart Attacks, and Market Crashes, Springer, (2002).

31. A. Arneodo, E. Bacry and J.F. Muzy, Solving the Inverse Fractal Problem from Wavelet Analysis, *Europhysics Letters*, **25**, No 7, 479–484, (1994).

32. A. Arneodo, A. Argoul, J.F. Muzy, M. Tabard and E. Bacry, Beyond Classical Multifractal Analysis using Wavelets: Uncovering a Multiplicative Process Hidden in the Geometrical Complexity of Diffusion Limited Aggregates. *Fractals* **1**, 629 (1995).

33. Z.R. Struzik The Wavelet Transform in the Solution to the Inverse Fractal Problem. *Fractals* **3** No.2, 329 (1995).

34. Z. R. Struzik, A. P. J. M. Siebes. Wavelet Transform in Similarity Paradigm. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Volume 1394 of Lecture Notes in Artificial Intelligence, pp 295-309, Melbourne, Australia, April 1998.

35. A. Smola, B.Schl̈kopf, A Tutorial on Support Vector Regression, NeuroCOLT2 technical report NC-TR-1998-030, (1998).