Eur. Phys. J. B **43**, 573–579 (2005)
DOI: 10.1140/epjb/e2005-00092-8

THE EUROPEAN PHYSICAL JOURNAL B

Identifying the ultrametricity of time series

F. Murtagh[a]

Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

Received 5 July 2004 / Received in final form 22 December 2004
Published online 30 March 2005 –

Abstract. High dimensional, sparsely populated data spaces have been characterized in terms of ultrametric topology. There are natural, not necessarily unique, tree or hierarchy structures defined by the ultrametric topology. Once such a structure is known, and can be defined, there are various implications including the feasibility of improved computational complexity for operations such as nearest neighbor searching. In this work, we consider the case where the data under investigation is temporal data, in the form of a time series. We develop an approach to characterizing how well time series data can be embedded in an ultrametric topology. Possible applications of this work include: (i) unique fingerprinting of a time series; (ii) discriminating between time series from various domains; and (iii) if data are inherently hierarchical, then using such hierarchies to model and predict.

PACS. 05.45.Tp Time series analysis – 05.45.-a Nonlinear dynamics and nonlinear dynamical systems – 02.50.Sk Multivariate analysis

1 Introduction

1.1 Relationship of ultrametricity with sparseness and dimensionality

The question has been raised [1,2] as to how ultrametric financial stocks are. However this analysis was through definition of a metric on a set of time series, followed by construction of a hierarchical clustering, which therefore yielded ultrametric properties by construction. We are instead interested in the inherent ultrametric properties of an individual time series.

To address this, firstly we embed the time series in a space of specified dimensionality, which we can endow with a metric, typically the Euclidean metric. This is a very traditional procedure. Secondly, inspired by the results of [3–5] we investigate data coding. We find that extent of ultrametricity is crucially dependent on coding of the data values in a given time series. In this article we are concerned with univariate time series, to simplify the description, and also to focus the experimental evaluations. Generalization to the case of multivariate time series is straightforward.

We use a "very coarse-grained description" of values in a time-varying signal, leading to a "symbolic analysis" [6,7]. Unlike work such as [6,7] which is based on ordinal properties of the data values, we are seeking persistent local hierarchical patterns that can be taken as defining a local ultrametric topology.

[a] e-mail: fmurtagh@acm.org

1.2 Ultrametricity

The triangular inequality holds for a metric space: $d(x,z) \leq d(x,y)+d(y,z)$ for any triplet of points x, y, z. In addition the properties of symmetry and positive definiteness are respected. The "strong triangular inequality" or ultrametric inequality is: $d(x,z) \leq \max\{d(x,y), d(y,z)\}$ for any triplet x, y, z. An ultrametric space implies respect for a range of stringent properties. For example, the triangle formed by any triplet is necessarily isosceles, with the two large sides equal; or is equilateral.

Ultrametricity has been shown to be a natural property of high-dimensional spaces [3,4]; and ultrametricity emerges as a consequence of randomness and of the law of large numbers. In previous work [5] we obtained further experimental results using Lerman's [8] approach to quantifying ultrametricity, and the new coefficient of ultrametricity described below, which supported the finding that high-dimensional sparse spaces are naturally ultrametric.

An ultrametric topology is associated with the p-adic numbers, and in clustering a bijection is defined between a rooted, binary, ranked, indexed tree, called a dendrogram, and a set of ultrametric distances. Using an agglomerative clustering algorithm, we can induce an ultrametric (i.e. respect for the ultrametric inequality, given any triplet of points) on any set of points endowed with a pairwise dissimilarity function. When a collection of points, in a space of any given dimensionality, is such that all triplets of points already satisfy the ultrametric inequality, then this collection of points has a natural hierarchical structure. It is not guaranteed that this hierarchy is unique.

1.3 Data coding

A time series can be easily embedded in a space of dimensionality m, by taking successive intervals of length m, or a delay embedding of order m. Thus we define points

$$\mathbf{x}_r = (x_{r-m+1}, x_{r-m+2}, \ldots, x_{r-1}, x_r)^t \in \mathbb{R}^m$$

where t denotes vector transpose. Based on previous results we expect that as the dimension m grows, then the point set in $\mathbb{R}^m$ becomes more ultrametric. This finding is borne out below.

Given any $\mathbf{x}_r = (x_{r-m+1}, x_{r-m+2}, \ldots, x_{r-1}, x_r)^t \in \mathbb{R}^m$, let us consider the set of s such contiguous intervals determined from the time series of overall size n. For convenience we will take $s = \lfloor n/m \rfloor$ where $\lfloor . \rfloor$ is integer truncation. The contiguous intervals could be overlapping but for exhaustive or near-exhaustive coverage it is acceptable that they be non-overlapping. In the experiments below, the intervals were non-overlapping. We will study how ultrametric the point $\mathbf{x}_r$ is, through simply investigating the m points, $x_{r-m+1}, x_{r-m+2}, \ldots, x_{r-1}, x_r$ on which we will define a signal ultrametricity index value. The quantification of the ultrametricity of the overall time series is provided by the aggregate over s time intervals of the ultrametricity of each $\mathbf{x}_r$, $1 \leq r \leq s$. We will simplify notation by defining the values in any given "window" or data interval, $\mathbf{x}_r$, as $\mathbf{x}_r = \{x_{rj} | 1 \leq j \leq m\}$.

In this article we want to directly quantify the extent of ultrametricity in time series data. In [4,5] it was shown how increase in ambient spatial dimensionality leads to greater ultrametricity. However it is not satisfactory from a practical point of view to simply increase the embedding dimensionality m insofar as short memory relationships are of greater practical relevance (especially for prediction). The greatest possible value of $m > 1$ is the total length of the time series, n. Instead we will look for an ultrametricity measurement approach for given and limited sized dimensionalities m. Our experimental results below for real and for random data sets are for $m = 5, 10, \ldots, 105, 110$.

We will seek local ultrametricity, i.e. hierarchical structure, by studying the following: Euclidean distance squared, $d_{jj'} = (x_{rj} - x_{rj'})^2$ for all $1 \leq j, j' \leq m$ in each time window, x_r.

Analyzing the local structure of the time series in this way is not remarkable. There is a long tradition of carrying out local principal components analysis or local regression [9,10]. Insofar as we are using a simple mapping of pairwise distances, our approach can be seen as defining a texture measure related to the widely-used Haralick co-occurrence matrix [11]. Hence our measure is a simple texture measure for one-dimensional signals, and more advanced measures can be easily envisaged (entropy-based, etc.).

Using distance allows for immediate generalization to multivariate time series. Distance also detrends the signal. The distance squared can be considered as an energy measure (as is usual in signal processing). We investigated the distance (not squared) and found that it provides an equally feasible alternative.

We will enforce sparseness [3–5] on our given distance values, $\{d_{jj'}\}$. We do this by linearly approximating each value $d_{jj'}$, in the range $\max_{jj'} d_{jj'} - \min_{jj'} d_{jj'}$, by an integer in $1, 2, \ldots p$. Note that the range is chosen with reference to the currently considered time series window, $1 \leq j, j' \leq m$. Note too that the value of p must be specified. Thus far, the recoded value, $d'_{jj'}$ is not necessarily a distance. With the extra requirement that $d'_{jj'} \longrightarrow 0$ whenever $j = j'$ we will show below that $d'_{jj'}$ is a metric.

This coding of distances is similar to the ordinal coding used in [6,7,12]. In this work we study the case of $p = 2$. We investigated exactly the same index of ultrametricity for increasing values of p. We found decrease of ultrametricity index value for increasing p. We will not report further on this line of investigation here.

To summarize, in our coding, a small pairwise transition is mapped onto a value of 1; and an exceptionally large pairwise transition is mapped onto a value of 2. A pairwise transition is defined not just for data values that are successive in time but for any pair of data values in the window considered.

1.4 Recoding: Properties

Consider the case of $p = 2$. Given any two values in the window, we code 1 if there is no change or small change; and 2 if the change is large. The threshold between large versus small is defined as $\max_{jj'} d_{jj'}/2$ where values indexed by j, j' are in the window. The ultrametric inequality for such dissimilarities typifies an energy landscape [3,4] in that change in value between any j and j'' is given by the maximum of change between values indexed by j, j', and values indexed by j', j''. We are simply considering the appropriately discretized one-dimensional signal within the m-length window in terms of an "energy landscape". In our assessment of ultrametricity, we will see how ultrametric these pairwise recoded distances are.

Consider the case of the following mapping: $d_{jj'} \longrightarrow 0$ if $d_{jj'} \leq \max_{jj'} d_{jj'}/2$; and $d_{jj'} \longrightarrow 1$ if $d_{jj'} > \max_{jj'} d_{jj'}/2$. It is easily verified that d is a metric and furthermore an ultrametric.

Our ordinal coding requires: $d_{jj'} \longrightarrow 1$ if $d_{jj'} \leq \max_{jj'} d_{jj'}/2$; $d_{jj'} \longrightarrow 2$ if $d_{jj'} > \max_{jj'} d_{jj'}/2$; and $d_{jj'} \longrightarrow 0$ if $j = j'$. We have that $d_{jj'} = d_{j'j} \geq 0$, and $d_{jj'} = 0 \Longleftrightarrow j = j'$. For the triangular inequality, consider distinct j, j', j''. Given $d_{jj''} \leq d_{jj'} + d_{j'j''}$, if $d_{jj''} = 2$, then this inequality always holds, potentially as an equality. If $d_{jj''} = 1$ then also this inequality must hold, as a strict inequality. If now we consider non-distinct j, j', j'', the triangular inequality is still found to hold. We have therefore proved that our ordinal coding results in a metric. There is not a guarantee however that our ordinal coding is an ultrametric since we cannot guarantee that $d_{jj''} \leq \max \{d_{jj'}, d_{j'j''}\}$. Consider e.g. $d_{jj''} = 2$, and $d_{jj'}, d_{j'j''} = 1$.

2 Approaches to quantifying ultrametricity

2.1 Past approaches

The Rammal ultrametricity index [3,4] is given by $\sum_{x,y}(d(x,y) - d_c(x,y))/\sum_{x,y} d(x,y)$ where d is the metric distance being assessed, and d_c is the subdominant ultrametric. The Rammal index is bounded by 0 (= ultrametric) and 1. This index suffers from "the chaining effect and from sensitivity to fluctuations" [3,4]. The single link hierarchical clustering method, yielding the subdominant ultrametric, is, as is well known, subject to such difficulties.

Lerman's H-classifiability measure [8] considers all possible triplets of points, and then considers how isosceles the associated triangles are. So as to avoid influence of distribution of the distance values, Lerman's measure is based on ranks (of these distances) only. We investigated Lerman's measure extensively in [5], and found that the ordinal approach unduly downgrades the importance of the equilateral triangle case, which is important for us. With Lerman's index, ultrametricity is associated with $H = 0$ but ultrametricity is not bounded.

Treves [13] considers triplets of points giving rise to minimal, median and maximal distances. In the plot of $d_{\min}/d_{\max}$ against $d_{\text{med}}/d_{\max}$, the triangular inequality, the ultrametric inequality, and the "trivial limit" of equilateral triangles, occupy definable regions.

Hartmann [14] considers $d_{\max} - d_{\text{med}}$. To give (translation, scale, etc.) invariance to the sensitivity (i.e., instability, lack of robustness) of distances, Hartmann fixes the remaining distance $d_{\min}$.

2.2 A new measure of ultrametricity based on angles

We seek to avoid, as far as possible, lack of invariance due to use of distances. We seek to quantify both isosceles with small base configurations, as well as equilateral configurations. Finally, we seek a measure of ultrametricity bounded by 0 and 1. We will therefore use a coefficient of ultrametricity – we will term it γ – which is specified algorithmically as follows.

1. All triplets of points are considered, with a distance (by default, Euclidean) defined on these points. For n points (data values, vector or scalar: in this paper we deal with scalar values) the number of triplets is $n(n-1)(n-2)/6$.
2. We check for possible alignments (implying degenerate triangles) and exclude such cases.
3. Next we select the smallest angle as less than or equal to 60°. (We use the well-known definition of the cosine of the angle facing side of length x as: $(y^2 + z^2 - xy)/2yz$.) This is our first necessary property for being a strictly isosceles (<60°) or equilateral (=60°) ultrametric triangle.
4. For the two other angles subtended at the triangle base, we seek an angular difference of strictly less than 2° (0.03490656 radians). This condition is an approximation to the ultrametric configuration. This condition is targeting a configuration that may not be exactly ultrametric but nonetheless is very close to ultrametric.
5. Among all triplets (1) satisfying our exact properties (2, 3) and close approximation property (4), we define our ultrametricity coefficient as the relative proportion of these triplets. Approximately ultrametric data will yield a value of 1. On the other hand, data that is non-ultrametric in the sense of not respecting conditions 3 and 4 will yield a low value, potentially reaching 0.

3 Quantitative assessments

Table 1 lists 31 real data sets used. In Table 1 we looked at subsets of the financial futures and Mississippi data in order to assess how processing part of the data differed from processing all. The EEG channel data (two channels used) is from `http://www.cs.colostate.edu/eeg/index.html`. Other EEG data is from `ftp://sigftp.cs.tut.fi/pub/eeg-data`. In the present work we treated each channel as an individual signal. The irregular epilepsy is described as between petit mal epilepsy and a tic symptom. The quadratic map is chaotic for nearly all $x_0 \in [0,1]$.

Time series used in this work which are not otherwise easily available are at the following address, as is also the program used (written in C) for most of the experimental appraisal: `http://astro.u-strasbg.fr/~fmurtagh/time-series`. (Figures in this paper were produced with R and/or S-Plus. The program for principal components analysis, used for summarizing the results, is available on our site at: `http://astro.u-strasbg.fr/~fmurtagh/mda-sw`).

Then in addition to the 31 time series from Table 1, in order to baseline the results, we also used uniformly random data with the same numbers of data values: 1326, 1169, 2739, 1374, 3080, 6160, 1471, 20000, 43829, 34726, 2500, 999, 1999. We investigated differing initializations for the random number generation: this had little effect and we do not report on these results here. In all therefore 44 time series were analyzed.

Summarizing the methodology, we begin by considering pairwise squared Euclidean signal distance in each time window: $(x_{rj} - x_{rj'})^2$ for all $1 \leq j, j' \leq m$ for each time window, $\mathbf{x}_r$. We map each value of squared distance between distinct time series data values into an integer, $1, 2$, depending on the threshold given by the average squared distance. We determine the ultrametricity measure, γ, for varying m ($m = 5, 10, \ldots, 105, 110$).

Typical results are presented in Table 2. The baseline random results do not differ much for varying n (number of data values in the time series). In all cases, there is generally an increase in γ for increasing window size m, which equates to increasing embedding dimensionality. To see this, consider the fact that we are defining pairwise distances between all data values in the window. Then

Table 1. Real data sets used in computational assessments. Simulated data sets (see text for details) had sequence numbers 32 through 44.

Seq. No.	Abbreviated Name	No. Values and Description
1	FTSE	1326
	FTSE – Financial Tmes Stock Exchange index	
2	USD/EUR	1169
	USD/EUR daily foreign exchange rates	
3	Sunspot	2739
	Monthly index values of sunspot solar physics activity	
4	Stock	1374
	Stock price, unknown origin	
5	Futures-3080	3080
	First 3080 values of futures	
6	Futures	6160
	Futures, daily highs	
7	Eyegaze	1471
	One coordinate of eyegaze position from eye tracker	
8	Mississippi-20000	20,000
	First 20,000 values of Mississippi data	
9	Mississippi	43,829
	Mississippi River daily water levels	
10	WWW traffic	34,726
	Bytes transferred per hour by a web server	
11	EEG-chan4	2500
	EEG channel p4, sampled at 250 Hz for 10 seconds	
12	EEG-chan5	2500
	EEG channel o1, sampled at 250 Hz for 10 seconds	
13	Quadratic map 1	2500
	$x_{t+1} = 4x_t(1 - x_t)$, $x_0 = 0.2$	
14	Quadratic map 2	2500
	$x_{t+1} = 4x_t(1 - x_t)$, $x_0 = 0.37777$	
15	Quadratic map 3	2500
	$x_{t+1} = 4x_t(1 - x_t)$, $x_0 = 0.451$	
16	Sleep EEG chan. 1	999
	EEG, sleep recording, normal, male, 28 years	
17	Sleep EEG chan. 2	999
	EEG, sleep recording, normal, male, 28 years	
18	Sleep EEG chan. 3	999
	EEG, sleep recording, normal, male, 28 years	
19	Sleep EEG chan. 4	999
	EEG, sleep recording, normal, male, 28 years	
20	Sleep EEG chan. 5	999
	EEG, sleep recording, normal, male, 28 years	
21	Sleep EEG chan. 6	999
	EEG, sleep recording, normal, male, 28 years	
22	Sleep EEG chan. 7	999
	EEG, sleep recording, normal, male, 28 years	
23	Sleep EEG chan. 8	999
	EEG, sleep recording, normal, male, 28 years	
24	Petit mal EEG chan. 1	1999
	EEG, petit mal epilepsy, male, 13 years	
25	Petit mal EEG chan. 2	1999
	EEG, petit mal epilepsy, male, 13 years	
26	Petit mal EEG chan. 3	1999
	EEG, petit mal epilepsy, male, 13 years	
27	Petit mal EEG chan. 4	1999
	EEG, petit mal epilepsy, male, 13 years	
28	Irreg. epil. EEG chan. 1	1999
	EEG, irregular type epilepsy, female, 10 years	
29	Irreg. epil. EEG chan. 2	1999
	EEG, irregular type epilepsy, female, 10 years	
30	Irreg. epil. EEG chan. 3	1999
	EEG, irregular type epilepsy, female, 10 years	
31	Irreg. epil. EEG chan. 4	1999
	EEG, irregular type epilepsy, female, 10 years	

Table 2. A sample of results: values in the table are the γ ultrametricity index for different data sets and for varying m. FTSE financial index, and EEG channel 1 data with petit mal epilepsy discharges, were used. Random data sets were used in both cases to provide a baseline for these results. Sliding window size (embedding dimensionality) m.

m	FTSE ($n = 1326$)	Random ($n = 1326$)	Petit mal 1 ($n = 1999$)	Random ($n = 1999$)
5	0.666792	0.668679	0.652130	0.670927
10	0.797854	0.745581	0.768844	0.747822
15	0.857168	0.778197	0.810113	0.774981
20	0.889421	0.787267	0.834574	0.787684
25	0.911665	0.798138	0.847419	0.796769
30	0.929086	0.807412	0.869242	0.805583
35	0.932682	0.811166	0.884085	0.811213
40	0.939093	0.818525	0.892667	0.817591
45	0.946903	0.819083	0.907404	0.818962
50	0.957005	0.819533	0.907352	0.820175
55	0.957574	0.821033	0.915256	0.821706
60	0.966657	0.825426	0.915082	0.824871
65	0.962117	0.824178	0.929823	0.825450
70	0.968155	0.826235	0.927563	0.826589
75	0.972261	0.824427	0.934610	0.827345
80	0.971924	0.827670	0.936480	0.828694
85	0.974162	0.829166	0.941335	0.830901
90	0.977248	0.832391	0.944002	0.833421
95	0.979769	0.831003	0.944998	0.832070
100	0.981258	0.832870	0.951889	0.834850
105	0.981922	0.831642	0.950013	0.831216
110	0.984535	0.830991	0.952017	0.833330

m points can be embedded in a metric space of dimensionality at most $m - 1$.

Figures 1, 2, and 3 give an overall view of the γ ultrametricity index. We used principal components analysis based on variances and covariances, i.e. based on centering to zero mean in attribute space. Our attributes were the 22 window sizes ($m = 5, 10, \ldots, 105, 110$), and our observations were the 44 time series.

Figure 1 shows the near-linear ordering (on the curve) of the window sizes, m. Figure 2 shows a wide range of findings: the random data sets (numbered 32 to 44) are all highly clustered; the chaotic data sets (numbered 13, 14, 15) are well-separated; the partial and complete data sets are close, viz. futures numbered 5 and 6, and Mississippi water levels numbered 8 and 9; the eyegaze trace data, numbered 7, is remarkably high in ultrametricity γ index value, which may be due to extreme values that were not subject to preprocessing; all EEG data sets are close together, viz. numbers 11 and 12; numbers 16 through 23 for the normal sleep subject, and numbers 24 through 27, and 28 through 31, for the epilepsy cases.

The principal components analysis presents an overall view of the relationships between time series. Figure 3 takes (arbitrarily but indicatively) two of the time windows, $m = 10$ and $m = 110$, and plots values for the 44 time series. We see immediately that greater embedding dimensionality, which corresponds (not necessarily monotonically) to greater m, results in higher γ

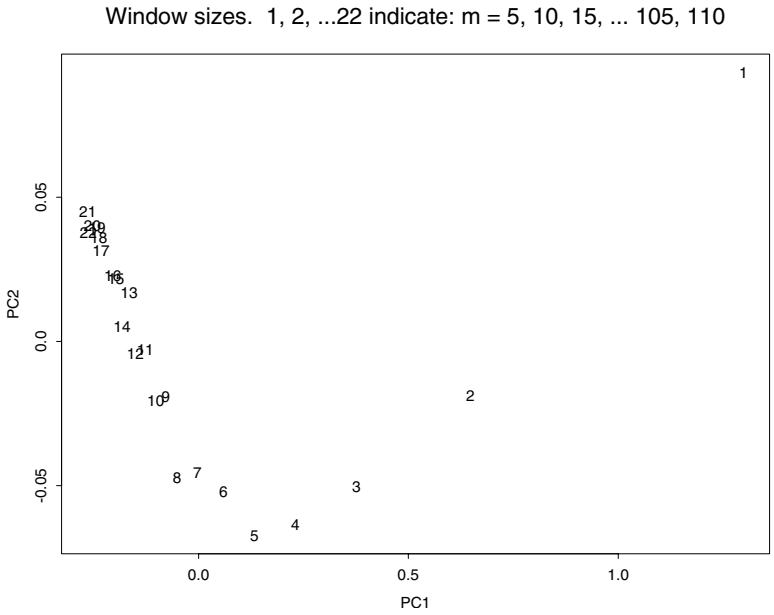

Fig. 1. Principal components analysis best planar projection. Successive "window" or embedding dimensionality sizes of length $m = 5, 10, 15, \ldots, 105, 110$ are used. These are labeled, respectively, $1, 2, \ldots, 22$.

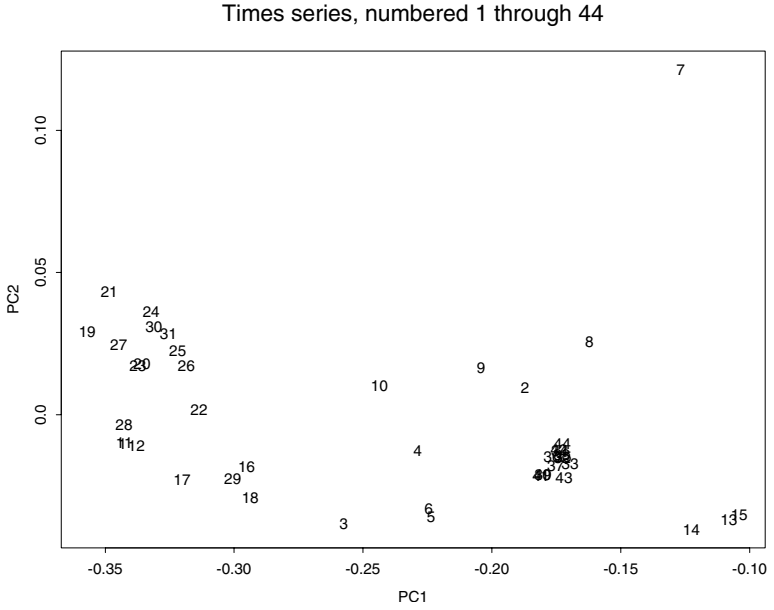

Fig. 2. Principal components analysis best planar projection. 44 time series data sets, labeled $1, 2, \ldots, 44$.

ultrametricity index value (upper curve of values, compared to lower curve). Secondly we see that the greatest departure from ultrametricity is to be found for time series 13, 14, and 15, i.e. the quadratic map chaotic series. On the other hand, the random data valued time series numbered 32 and greater are remarkably similar in γ value. The eyegaze trace here (time series 7) is not unduly different from other time series. The relatively short EEG time series raise some interesting issues (e.g., differences between waking, sleeping, or epileptic behaviors) which we will not pursue further here.

Again we find the futures (initial half of the time series and the full time series used, numbered 5 and 6) to be very similar. The Mississippi water level (initial half, and full, time series, numbered 8 and 9) are also very similar. This shows how γ is not influenced by the length of the time series data.

4 Discussion

4.1 Data embedding

Consider the case of each data value in our given time series as being associated with the node of a binary tree. For the present, it is not important how this binary tree is

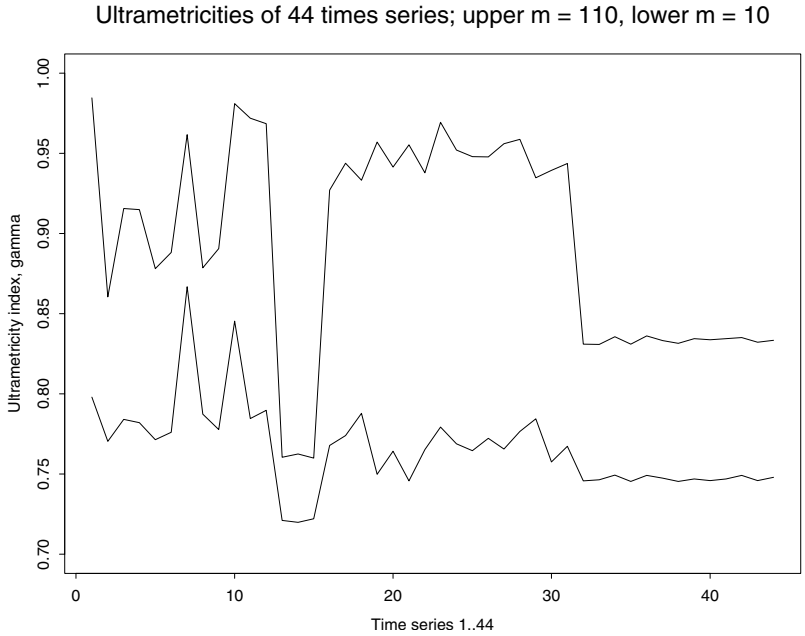

Fig. 3. Investigation of two of the windows (embedding dimensions), $m = 10$ and $m = 110$. Results for 44 time series are shown, with window size $m = 110$ on top and $m = 10$ on bottom. In both cases, an ultrametricity γ value is plotted for each time series. Portraying the γ values as a continuous curve for all data sets is done for visualization.

determined. We may also start with the time series data values being the terminal nodes of a binary tree, and then convert this into a tree defined by an in-order tree traversal (see [15]). With any such binary tree there is a bijection with an indexed binary tree, and in turn a bijection with a set of pairwise ultrametric distances [16].

A succession of m values in a time series will be considered as a set of randomly sampled (arbitrary distribution) nodes from a binary tree. Between any triplet of data values, an ultrametric relationship implies that the associated triangle is either equilateral or isosceles with small base. In our assessment of ultrametricity, we examine all possible triangles for these properties.

Choosing a fixed value of m has an evident computational advantage, and additionally, from the point of view of practical implications for modeling and prediction (for example), local (small m) ultrametricity is better than very large m.

As already noted [3–5] we must seek ultrametricity in a sparse and hence relatively high dimensional space. In our approach, we therefore map our given time series data into a sparse space by data coding.

4.2 Inferring hierarchies

We recall that our coding maps distances between data values into 1 or 2 (with strictly > 0 distance assumed for distinct data values). If the time series, thus coded, is found to be ultrametric (with a value of our ultrametricity index $\gamma = 1$), then the ultrametric inequality can be invoked to show that any triplet of distances has one of the following set of codes: 1, 1, 1; 1, 2, 2; 2, 2, 1; 2, 1, 2; and 2, 2, 2. The following set of codes is excluded: 1, 1, 2; 1, 2, 1; and 2, 1, 1.

Consider a window length of $m = 5$, with data values given by $(x_{r-4}, x_{r-3}, x_{r-2}, x_{r-1}, x_r)^t$. Following our data coding, let us asssume that values x_{r-4}, x_{r-3}, x_r are all mutually 1 apart; and that values x_{r-2}, r_{r-1} are 2 apart, and all of the latter are 2 apart from the former. This is a legitimate albeit contrived situation. We use it simply to have the following two-level hierarchy: $((x_{r-4}, x_{r-3}, x_r),\ x_{r-2}, r_{r-1})$.

Such a pattern results from the ultrametric inequality. The question as to whether the same pattern repeats in all $m-$length windows remains an open one.

4.3 Relationship with self-similarity

Self-similarity (for various works, including use for signal modeling and prediction, see [17–19]) points to the need to take long memories into account in modeling. From the literature on web traffic flows, EEG signals, eyegaze trace signals, etc., it is very likely that data sets used by us have self-similarity properties. However our methodology here is not directly related to self-similarity. Instead we are looking for strong hierarchical relationships that are consistently (and persistently over time) present in the m-length sliding windows.

4.4 Relationship with ordinal coding

In [6,7,12], ordinal and hence permutation coding of the data is used. There is a clear relationship between a permutation and a hierarchical or tree structuring of a set of

data values (again we note, in this article univariate, but with trivial extension to the multivariate case). There is an isomorphism between a class of hierarchic structures, termed unlabeled, ranked, binary, rooted trees, and a class of permutations. The latter are the alternating permutations where $p(i) < p(i+1) > p(i+2)$ for $1 \leq i \leq n-2$, which is an up-down or down-up permutation depending on the state of $i = 1$. Such a permutation can be read off the unlabeled, ranked, binary, rooted tree (or dendrogram) through in-order tree traversal. The number of complementary up-down or down-up permutations are counted by the André numbers (for $n = 4, 5, 6, 7, 8$, respectively 2, 5, 16, 61, 272). See [15,20,21]. We can conclude that our work differs from [6,7,12] in that while both are in terms of permutation- or ordinal-based coarse-grained symbolic description, our work uses a more restrictive coding.

4.5 Relationship with hierarchies in graphs

As opposed to our approach where the embedding used is fixed (window of length m), in considering the degree of hierarchical structure in a graph or network, Trusina et al. [22] use interconnecting paths, and therefore variable-sized or adaptive embeddings. Based on graph vertex weights, vertex interconnection paths are characterized in terms of up and down. There are intriguingly close links between this work and ours, for the discovery of hierarchical patterns in, respectively, graphs and time series, which we will return to in future work.

References

1. R.N. Mantegna, Eur. Phys. J. B **11**, 193 (1999)
2. R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics* (Cambridge, Cambridge University Press, 2000)
3. R. Rammal, J.C. Angles d'Auriac, B. Doucot, J. Phys. Lett. **46**, L-945 (1985)
4. R. Rammal, G. Toulouse, M.A. Virasoro, Rev. Mod. Phys. **58**, 765 (1986)
5. F. Murtagh, J. Classification **21**, 167 (2004)
6. K. Keller, H. Lauffer, Int. J. Bifurcation Chaos **13**, 2657 (2003)
7. K. Keller, K. Wittfeld, Int. J. Bifurcation Chaos **14**, 693 (2004)
8. I.C. Lerman, *Classification et Analyse Ordinale des Données* (Paris, Dunod, 1981)
9. Zhi-Yong Liu, Lei Xu, Neurocomputing **55**, 739 (2003)
10. H. Späth, *Cluster Dissection and Analysis* (Ellis Horwood, Chichester, 1985)
11. R.M. Haralick, Proc. IEEE **67**, 786 (1979)
12. C. Bandt, B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002)
13. A. Treves, BioSystems **40**, 189 (1997)
14. A.K. Hartmann, Europhys. Lett. **44**, 249 (1998)
15. F. Murtagh, Discrete Applied Math. **7**, 191 (1984)
16. J.P. Benzécri, *La Taxinomie*, 2nd edn. (Paris, Dunod, 1979)
17. A. Aussem, F. Murtagh, Int. J. Intelligent Syst. **16**, 215 (2001)
18. A.M. Battro, *La temperature de la mirada*, Laboratorio de Investigaciones Sensoriales, Conicet, Buenos Aires (1992). See also A.M. Battro, *A fractal story*, `http://www.byd.com.ar/fractalstory.htm`
19. T.C. Ferree, R.C. Hwa, Neurocomputing **52–54**, 755 (2003)
20. L. Comtet, *Advanced Combinatorics* (Reidel, Dordrecht, 1974)
21. R. Donaghey, J. Combin. Theory (A) **18**, 141 (1975)
22. A. Trusina, S. Maslov, P. Minnhagen, K. Sneppen, Phys. Rev. Lett. **92**, 178702 (2004)