Appendix 2

2.1 Symbolic dynamics and analysis blocks

Complex nonlinear dynamical systems can be modelled and studied in the context of coarse-graining, *i.e.*, they can viewed as information capacitors containing discrete series of symbolic messages.[1–5] Coarse-graining can be outlined through symbolisation, a process which allows partitioning of the full continuous phase space into finite number of cells.[1] In this manner, symbolisation provides a rigorous way of studying actual complex dynamics under finite precision. This operational procedure is often referred as symbolic dynamics.[1,4–7] One of the merits of symbolic dynamics is that it provides a strong link between dynamical systems and information theory.[1,4,5,8] According to symbolic dynamics, time-series are re-organised into new symbolic sequences, in which every symbol stands for a partition of the initial time-series. Every different symbol is represented by an alphabet letter and the whole procedure is called lettering.[1–3,6,7,9] The new subsequences of symbols are called words.[2,3,7] Reading of symbolic sub-sequences can be derived through the processes of lumping or gliding.[2,3,7] Lumping is the interpretation of symbolic words through independent sequential discrete portions of certain number of words, called blocks, opposed to gliding, where the portions are not independent.[2,3,7] Note that gliding is the standard convention in literature and is often referred also under the term sliding or moving-frame.[2,3,7] Block time-series symbolisation prerequisites selection of λ different letters from an alphabet and choice of the size, n, of blocks or words, *i.e.*, the number of sequential letters that will be treated as a whole. Depending on λ and n the maximum number, N of different words is determined in the selected alphabet. For example in a $\lambda = 2$ lettering, a threshold C may be considered. Each value above this threshold may be symbolised as 1 and each below, as 0.[2,3,7] Initial time-series, for instance, of length $L = 20$, may be transformed through symbolic dynamics to *e.g.* 11001010111000101010 in the $\lambda = 2$ lettering. Through lumping, the $\lambda = 2$-letter symbolic time-series may be organised in sets of $n = 2$ blocks as (11|00|10|10|11|10| 00|10|10|10|), in which each block is one of the $N = 2^2 = 4$ different words in this lettering, *i.e.*, (00, 01, 10,11). The same symbolic sequence through gliding-sliding will be treated as (11|10|00|01|10|01| 10|01|11|11|10|00| 00|01|10|01|10|01| 10|01|10), *i.e.*, the sequence will be of greater length. The $\lambda = 2$ letter sequence may be also organised in blocks of $n = 3$ letters with maximum of $N = N_{3,2} = n^{\lambda} = 3^2 = 9$ different words,*i.e.*, (000, 001, 010, 100, 110, 011, 010, 001, 111). Other sequences of words may be generated as well. In general,[1–3,6,7,9] through symbolic dynamics a L-length time-series is transformed to a symbolic time-series sequence, $[A_1, A_2 ... A_n ... A_L]$, composed by λ different letters, $[A^1, A^2 ... A^{\lambda}]$, from a λ-length alphabet. Symbolic time-series sequences are re-organised in n-sized words-blocks composed by letters of the alphabet $[A^1, A^2 ... A^{\lambda}]$. In linguistics the word size is unconfined and, hence, linguistic words contain some or, potentially, all letters. On the other hand, in symbolic dynamics, the words are of fixed lengths n, $n \geq \lambda$ and are chosen from $N_{max} = N_{n,\lambda} = n^{\lambda}$ different fixed-sized words in the $[A^1, A^2 ... A^{\lambda}]$ alphabet. In this manner, the symbolic time-series are re-organised as $...\underbrace{A_1 ... A_n}_{B_1}\underbrace{A_{n+1} ... A_{2n}}_{B_2} ... \underbrace{A_{in+1} ... A_{(i+1)n}}_{B_{i+1}} ...$ blocks, where i is the consecutive number of the block, *i.e.*, $i = 1 ...$ *Total number of blocks*. In lumping the n words-blocks are sequentially independent, on the contrary to the dependent sequential blocks in the gliding-sliding process. The total number of blocks of the symbolic time-series is greater for gliding-sliding and hence more computation is needed. The probability of occurrence of a block, $[A_1, A_2 ... A_n]$, of size n is calculated by

$$p^{(n)}(A_1, A_2 ... A_n) = \frac{Number\ of\ occurences\ of\ block\ [A_1, A_2 ... A_n]}{Total\ number\ of\ blocks} \tag{2.1}$$

2.2 Block entropy analysis

In the framework of complex signal analysis, specific entropy methodologies based on symbolic dynamics have been developed in the previous decade.[5,10–13] All these methodologies are referred as block entropies. Most common techniques rely on the extension of Shannon entropy[14]

$$H_S = -\sum p_i ln p_i \tag{2.2}$$

where p_i is the number of possible microscopic configurations. Note that equation (2.2) represents the classical Boltzman's entropy for the Gibbs canonical ensemble (B-Gentropy).[9] Combining equations (2.1) and (2.2), the Shannon block entropy, $H(n)$ of n-sized blocks is derived by (2.3):

$$H(n) = -\sum_{(A_1, A_2, ... A_n)} p^{(n)}(A_1, A_2, ... A_n) ln p^{(n)}(A_1, A_2, ... A_n) \tag{2.3}$$

Equation (2.3) calculates the entropy due to all possible words. It is a measure of uncertainty or disorder, *i.e.*, it measures organisation deficiency of a complex system. It also gives the average amount of information necessary to predict a sub-sequence of words or blocks of length n

From equation (2.3), the Shannon block entropy per letter may be derived by:

$$h^{(n)} = \frac{H(n)}{n} = \frac{-\sum_{(A_1, A_2, ... A_n)} p^{(n)}(A_1, A_2, ... A_n) ln p^{(n)}(A_1, A_2, ... A_n)}{n} \tag{2.4}$$

This entropy may be interpreted as the average uncertainty of a block of size n per letter.[6]

From the Shannon block entropy the conditional entropy may be derived by equation (2.5):

$$h_{(n)} = H(n+1) - H(n) \tag{2.5}$$

The conditional entropy $h_{(n)}$ measures the uncertainty of predicting a state one step into the future, provided a history of the preceding n states.[6]

For physical phenomena with long-range interactions or long-range memory effects, an important property observed is the violation of Boltzmann-Gibbs (B-G) statistics.[15] A generalised expression of the B-G statistics has been proposed based on multifractal concepts by Tsallis[16,17]

$$S_q = \frac{1}{q-1}(1 - \sum_{i=1}^{W} p_i^q) \tag{2.6}$$

where p_i denotes, in these references, the probabilities of a sequence and W their total number. q is a real nuber which is the measure of non-extensivity of the system.[16,17] Using $p_i^{(q-1)} = e^{(q-1)\ln(p_i)} \sim 1 + (q-1)\ln(p_i)$ in the limit $q \to 1$ the B-G entropy is derived.[6,9] The generalization of B-G expression, suggests the non-extensive statistical mechanics. The entropic index q characterises the degree of nonadditivity in the following pseudo-additivity rule[2,3,6,9,16]

$$S_q(A,B) = S_q(A) + S_q(B) + (1-q)S_q(A)S_q(B) \tag{2.7}$$

with $q > 1$ referring to sub-additivity and $q < 1$ to superadditivity. Systems that called non-extensive, have special probability correlations and extensivity may occur for S_q for specific value of index q.[18] Tsallis entropy has been used in terms of symbolic dynamics for electromagnetic time series prior to earthquakes.[2,3,6,9,16,19] By properly adjusting symbolisation in accordance to (2.3) and (2.4), the Tsallis entropy of a block $[A_1, A_2 ... A_n]$ of length n in a λ-letter alphabet can be calculated by[2,3,6,9,19]

$$S_q(n) = \frac{1}{q-1}(1 - \sum_{(A_1,A_2,...A_n)} [p^{(n)}(A_1, A_2...A_n)]^q) \tag{2.8}$$

where $p^{(n)}(A_1, A_2...A_n)$ is the probability of occurrence of block $[A_1, A_2 ... A_n]$. As already mentioned, high level of organisation is indicated when low values of Tsallis entropy are produced. Tsallis entropy has been explored in the field of earthquake time series analysis.[6,9,15,20–22] Recent work has also been published in biomedical imaging, with suggestions in replacing Shannon's theorem[23] and bioinformatics.[23] An interconnection between fractals and Tsallis entropy that has been introduced in previous decade could provide natural frame for studying fractally structured systems.[24] Moreover, a possible interconnection could exist between generalized Tsallis statistics and quantum groups.[17]

From (2.8) the normalised Tsallis entropy may be derived. The formula for the computation is[16]

$$\hat{S}_q(n) = \frac{\frac{1}{q-1}(1 - \sum_{(A_1,A_2,...A_n)} [p^{(n)}(A_1, A_2...A_n)]^q)}{\sum_{(A_1,A_2,...A_n)} [p^{(n)}(A_1, A_2...A_n)]^q} \tag{2.9}$$

where the symbolisation was adjusted in accordance to equations (2.3) and (2.4) following the approach proposed by other investigators.[6,9,19,20] In equation (2.9), $p^{(n)}(A_1, A_2...A_n)$ is the probability of occurrence of the block $[A_1, A_2...A_n]$ and q is the corresponding real number of equations (2.6),(2.8) and (2.9).

The appropriate choice of the entropic index q has crucial meaning for the Tsallis and the normalised Tsallis entropy computation and requires further exploration for its proper use.[9] For every specific use of Tsallis entropy the ranges of the q values will result in significant discrimination.[17] Non-additive Tsallis entropy combined with Gutenberg-Richter law provided excellent fit to seismicities with q-values range from 1.4 to 1.85. The q-values are rooted in a rather solid physical background and describe the non-additivity of a seismic emmision in a correct manner.[9] Moreover, index q can be consider as bias parameter with $q < 1$ refer to rare events and $q > 1$ refer to prominent events.[22] For pre-eqrthquake electromagnetic disturbances, the q-values are restricted in the region $1 < q < 2$, and are consistent with several studies that suggest the upper limit to be equal to 2.[9] It is noteworthy that entropic index q is not a measure of complexity but measures the non-extensivity of the system.[6]

References

1 K. Karamanos and G. Nicolis, *Chaos Soliton. Frac.*, 1999, **10**, 1135–1150.

2 K. Karamanos, D. Dakopoulos, K. Aloupis, A. Peratzakis, L. Athanasopoulou, S. Nikolopoulos, P. K. P and K. Eftaxias, *Phys. Rev. E.*, 2006, **74**, 21–36.

3 K. Karamanos, *J Phys. A Math. Gen.*, 2001, **34**, 9231–9241.

4 J. Nicolis, *Chaos and Information Processing*, World Scientific, Singapore, 1991.

5 G. Nicolis and P. Gaspard, *Chaos Soliton. Fract.*, 1994, **4**, 41–57.

6 K. Eftaxias, G. Balasis, Y. Contoyiannis, C. Papadimitriou, M. Kalimeri, L. Athanasopoulou, S. Nikolopoulos, J. Kopanas, G. Antonopoulos and C. Nomicos, *Nat. Hazard Earth Sys.*, 2009, **9**, 1953–1971.

7 K. Karamanos, D. Dakopoulos, K. Aloupis, A. Peratzakis, L. Athanasopoulou, S. Nikolopoulos, P. K. P and K. Eftaxias, *Phys. Rev. E.*, 2006, **74**, 21–36.

8 A. Voss, J. Kurths, H. Kleiner, A. Witt, N. Wessel, N. S. ad K. Osterziel, R. Schurath and R. Dietz, *Cardiovasc. Res.*, 1996, **31**, 419–433.

9 M. Kalimeri, C. Papadimitriou, G. Balasis and K. Eftaxias, *Physica A*, 2008, **387**, 1161–1172.

10 W. Ebeling, *Physica D*, 1997, **109**, 42–52.

11 W. Ebeling and G. Nicolis, *Chaos Soliton Fract.*, 1992, **2**, 635–650.

12 W. Ebeling and G. Nicolis, *Europhys Lett.*, 1991, **14**, 191–196.

13 K. Karamanos, *Lect. Notes Phys.*, 2000, **550**, 357–371.

14 C. Shannon, *Tech. J.*, 1948, **27**, 379–423.

15 Y. Contoyiannis and K. Eftaxias, *Nonlin. Processes Geophys.*, 2008, **15**, 379–388.

16 H. Suyari, 2002.

17 C. Tsallis, *J. Stat. Phys.*, 1988, **52**, 479–487.

18 J. Boon and K. Tsallis, *Europhys. News*, 2005, **36**, 185–186.

19 K. Eftaxias, G. Balasis, Y. Contoyiannis, C. Papadimitriou, M. Kalimeri, L. Athanasopoulou, S. Nikolopoulos, J. Kopanas, G. Antonopoulos and C. Nomicos, *Nat. Hazard Earth Sys.*, 2010, **10**, 275–294.

20 K. Eftaxias, *Physica A*, 2010, **389**, 133–140.

21 V. Surkov, S. Uyeda, H. Tanaka and M. Hayakawa, *J. Geodyn.*, 2002, **33**, 477–487.

22 L. Zunino, D. Perez, A. Kowalski, M. Martin, M. Garavaglia, A. Plastino and O. Rosso, *Physica A*, 2008, **387**, 60057–6068.

23 J. Mohanalin, S. Beenamol, P. Kalra and N. Kumar, *Comput. Math. Appl.*, 2010, **60**, 2426–2432.

24 F. Lopes, E. de Oliveira and R. Cesar, *BMC Syst. Biol.*, 2011, **5**, 61–69.