**Fractals 11 new**

**Sample Entropy**

Sample entropy (SampEn) was introduced by Richman and Moorman [1], as a modification of Approximate entropy method [2]. It measures the rate of generation of new information by examining time series for similar epochs: more frequent and more similar epochs (more regularity in the time series) lead to lower values of SampEn.

SampEn(m, r, N ) is defined as the negative natural logarithm of the conditional probability that two sequences similar for m points remain similar at the next point, where self-matches are not included in calculating the probability.

Sample entropy algorithm proceeds in following steps [1]:

1. For a time series of length , , we form  vectors  where  is the vector of length .

1. The distance between vectors  and  is defined as a maximum difference of their corresponding scalar components:





1. Next we count the number  of vectors  such that where is a tolerance level of accepting matches,  and  to exclude self-matches (tolerance level : , - standard deviation of ).
2. We then define

 and 

where is the probability that two vectors will match for points.

1. We repeat steps i-iv for vectors of length : We form vectors  where  is the vector of length 



We count the number of vectors  which are within of  and again we exclude self-matches. We define

 and 

where is the probability that two vectors will match for points.

1. Sample entropy (SampEn) is defined as



which is estimated by the statistics



It can be shown that / = where *A* is the number of forward matches of length and *B* is the number of forward matches of length  *m*. The quantity  is precisely the conditional probability that two sequences within a tolerance  r for m points remain within r of each other at the next point.   can be expressed as  [1]. It is the negative natural logarithm of the conditional probability that two sequences similar for m points remain similar at the next point, where self-matches are not included in calculating the probability.

Sample entropy method was used in analyzing physiological processes [3,4], geophysical signals [5,6], climatic data [7], hydrological [8,9] and financial time series [10]. Recently Sample entropy was applied on data mining [11] and engineering problems [12].

**Cross-Sample Entropy**

Cross-Sample Entropy (Cross-SampEn) was introduced by Richman and Moorman [1], as a modification of Cross-Approximate Entropy method [13], to measure the degree of asynchrony between two related time series.

The implementation of Cross-SampEn method is similar to SampEn algorithm described above:

1. For two time series of length,  and we form  vectors  and  where  and are vectors of length .

1. Next we count the number of vectors  within tolerance  from : .



1. We then define

 and 

where is the probability that two vectors (from two different series) will match for points.

1. We repeat steps i-iii for vectors of length . For two time series of length , and we form  vectors and where  and are vectors of length .





We count the number  of vectors  which are within of , where . We then define

 and 

where is the probability that two vectors (from two different series) will match for points. Cross-Sample entropy (Cross-SampEn) is estimated by



It can be shown that  and  are the number of pairs of vectors of length m+1 and length m (respectively) from the two series that match within r , and Cross-sample entropy represents a conditional probability that sequences (from two different series) that are similar (within certain tolerance level) over m consecutive data points will remain similar after addition of one consecutive data point. Higher values of Cross-SampEn indicate less synchronization between analyzed temporal series [1]. Cross-SampEn was used in analyzing physiological [14, 15], geophysical [16], and financial data [17,18].

**Example** (Adapted from Ref.[19])

We examine alterations in daily rainfall and stream flow time series recorded in the Piracicaba river basin located at southeastern region of Brazil, using entropy analysis. We analyze daily rainfall and stream flow temporal series recorded in Atibaia and Jaguari sub-basins, where a series of reservoirs linked by tunnels (Cantareira system) were built to transfer an average 31m3s-1 from the Piracicaba basin to the metropolitan area of Sao Paulo city. The reservoir construction within Atibaia sub-basin started around 1968 and inter-basin transfer around 1975, while within Jaguari sub-basin the construction started around 1975 and inter-basin transfer around 1982. Earlier studies showed positive trend for precipitation for the entire basin, and negative trend for stream flow for some locations within the basin. We apply Sample entropy analysis on stream flow and rainfall data before and after reservoirs construction, to obtain more information about hydrological alterations caused by Cantareira system. We also apply Cross-Sample entropy between stream flow and rainfall to compare the level of synchronization between these variables, before and after reservoirs construction. Our results are shown on Table 1 and Table 2.

Table1. SampEn and Cross-SampEn analysis for stream flow (3D-006,4D-009) and rain fall (D3-002, E3-017) data for Atibaia River basin

|  |  |  |  |
| --- | --- | --- | --- |
| **Station** | **Data** | **SampEn**  **(r=0,15)** | **CrossSampEn**  **(r=0.15)** |
| 3D-006 | 61- 67  75 - 81 | 0.79  0.52 | 0.33 \*  0.48 \*\* |
| 4D-009 | 61- 67  75 - 81 | 0.72  0.53 | 0.36 **§**  0.41 **§§** |
| D3-002 | 61- 67  75 - 81 | 0.46  0.49 | -  - |
| E3-017 | 61- 67  75 - 81 | 0.49  0.49 | -  - |

\*3D-006/D3-002 (61-67) §4D-009/E3-017 (61-67)

\*\* 3D-006/D3-002 (75-81) §§4D-009/E3-017 (75-81)

Table 2. SampEn and Cross-SampEn analysis for stream flow (3D-009, 4D-001) and rain fall (D3-018, D4-052) data for Jaguari River basin

|  |  |  |  |
| --- | --- | --- | --- |
| **Station** | **Data** | **SampEn**  **(r=0,15)** | **CrossSampEn**  **(r=0.15)** |
| 3D-009 | 67 - 75  85 - 93 | 0.59  0.76 | 0.35 \*  0.49 \*\* |
| 4D-001 | 67 - 75  85 - 93 | 0.56  0.59 | 0.35 **§**  0.39 **§§** |
| D3-018 | 67 - 75  85 - 93 | 0.50  0.53 | -  - |
| D4-052 | 67 - 75  85 - 93 | 0.49  0.50 | -  - |

\*3D-009/D3-018 (67-75) §4D-001/D4-052 (61-67)

\*\* 3D-009/D4-018 (85-93) §§4D-009/D4-052 (75-81)

It is seen from Table 1 that for Atibaia sub basin, the values of SampEn of daily streamflow temporal series decrease after construction of water reservoirs indicating the loss of complexity, with stronger effect for station 3D-006, which is closer to the reservoir. There is no change in complexity of corresponding daily precipitation time series, indicating the reservoir construction as cause of alteration in stream flow dynamics. The same conclusion holds for Jaguari sub basin, for which the values of SampEn increased for streamflow (with stronger effect for 3D-009 station which is closer to the reservoir) and didn’t change for precipitation data (Table 2). The values of Cross-SampEn, confirm that reservoir construction causes the alteration in complexity of discharge dynamics: the values of Cross-SampEn between discharge and precipitation time series at the same location increase after construction of reservoirs far all stations , indicating less synchrony between discharge and precipitation dynamics, and again this effect is stronger for stations that are closer to the reservoirs (3D-006and D3-002 in Atibaia sub basin and 3D-009 and D3-018 in Jaguari sub basin).

**Multiscale sample entropy (MSE)**

Multiscale sample entropy was introduced by Costa et al. [20], as a generalization of Sample entropy method [1]. Entropy-based measures such as Shannon entropy [21], Kolmogorov entropy [22], Approximate entropy [2] and its extension Sample entropy [1] grow monotonically with the degree of randomness and fail to quantify complexity as a “meaningful structural richness”, which exhibits higher regularity then random process. Both completely random (white noise) and completely regular (e.g. periodic) signals should exhibit less complexity then structurally “complex” process (e.g. 1/f noise) [20]. Multiscale entropy takes into account the multiple time scales by calculating sample entropy for consecutive coarse-grained time series  determined by the scale factor:  where  is original time series.

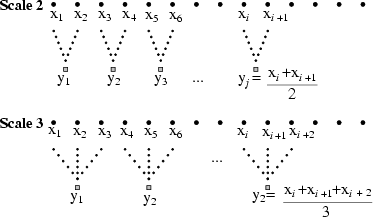


Figure 1. Schematic illustration of the coarse-graining procedure for scale 2 and 3 (Adapted from Ref. [20]).

By plotting MSE versus scale factor , we can analyze structural complexity of different components of underlying stochastic processes, which can serve to discriminate time series generated either by different systems or by the same system under different conditions[20]. It was showed by Costa et al., that uncorrelated random signals have, for larger scales, lower MSE values then correlated noise, making MSE more appropriate for quantifying complexity in short and noisy time series, then traditional entropy methods that evaluate pattern repetition on single temporal scale.

|  |
| --- |
| \begin{figure}\centerline{\epsfig{file=figures/noise,width=.7\linewidth}}\end{figure}  Figure 2. MSE analysis of *simulated white and 1/f noise* time series. Symbols represent mean values over 30 time series. Parameters to calculate sample entropy are: , , and  (Adapted from Ref. [20]). |

MSE method was used in analyzing physiological processes [23,24,25], financial time series [26,27], earthquake sequences [28], hydrological data [29,30] and traffic time series [31].

**Example** (Adapted from Ref. [20])

MSE method was applied to the cardiac Interbeat (RR) time series derived from 24h continuous electrocardiographic (ECG) Holter monitor recordings of healthy subjects, subjects with congestive heart failure, a life threatening condition, a subjects with atrial fibrillation , a major cardiac arrhythmia.

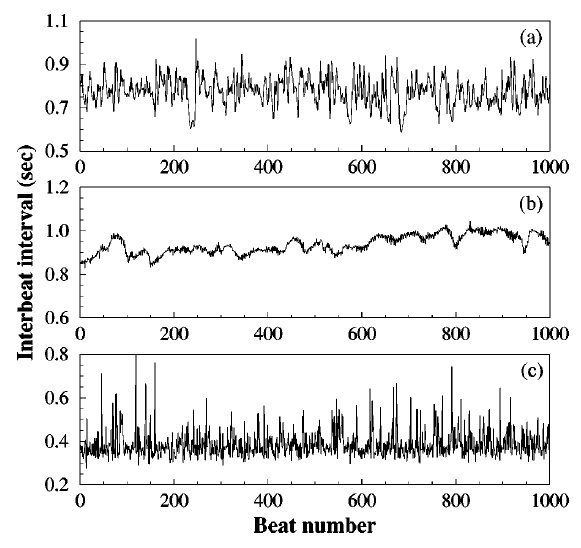
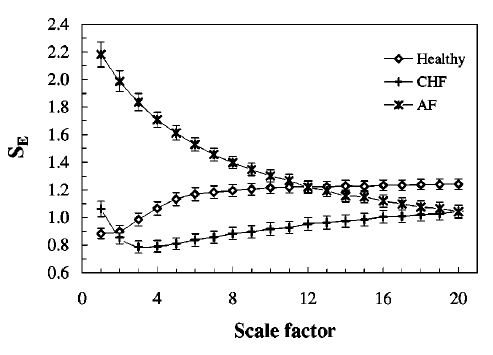


Figure 4. MSE analysis of RR time series derived from long-term ECG recordings of healthy subjects in normal sinus rhythm, those with congestive heart failure CHF in sinus rhythm, and those with atrial ﬁbrillation AF . Time series length is 2 104 beats. The SE values from healthy subjects are signiﬁcantly t-test, p 0.05 higher than from CHF and AF subjects for scales larger than scale 2 and scale 20, respectively [10].

Figure 3. Representative interbeat interval time series from (a) healthy individual sinus rhythm ,(b) subject with congestive heart failure, and (c) subject with atrial ﬁbrillation, a highly erratic cardiac arrhythmia [20].

**Bibliography**

Parte superior do formulário

[1] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, American Journal of Physiology 278, H2039-H2049, 2000.

[2] S.M. Pincus, Approximate entropy as a measure of system complexity, PNAS 88, 2297-2301, 1991.

[3] D. E. Lake , J. S. Richman, M. P. Griffin, J. R. Moorman, Sample entropy analysis of neonatal heart rate variability. Am J Physiol Regul Integr Comp Physiol  *283,* R789–R797, 2002.

[4] [M. D. Costa](http://scitation.aip.org/content/contributor/AU1093554), [T. Henriques](http://scitation.aip.org/content/contributor/AU1093555), [M. N. Munshi](http://scitation.aip.org/search?value1=Medha+N.+Munshi&option1=author&option912=resultCategory&value912=ResearchPublicationContent), [A. R. Segal](http://scitation.aip.org/search?value1=Alissa+R.+Segal&option1=author&option912=resultCategory&value912=ResearchPublicationContent), [A. L. Goldberger](http://scitation.aip.org/content/contributor/AU1093556). Dynamical glucometry: Use of multiscale entropy analysis in diabetes. Chaos 24, 033139, 2014.

[5]G. Balasis, I.A. Daglis, C. Papadimitriou, M. Kalimeri, A. Anastasiadis, K. Eftaxias, Investigating dynamical complexity in the magnetosphere using various entropy measures, Journal of Geophysical Research 114, A00D06, 2009.

[6] L. Min, M. Guang, N. Sarkar,   Complexity analysis of 2010 Baja California earthquake based on entropy measurements. Vulnerability, Uncertainty, and Risk: pp. 1815-1822, 2014.

[7] L. Shuangcheng, Z. Qiaofu, W. Shohong, D. Erfu, Measurement of climate complexity using sample entropy, International Journal of Climatology 26, 2131 –2139, 2006.

[8] F. Huang, Z. Hia, N. Zhang, Y. Zhang, J. Li, Flow-complexity analysis of the upper reaches of the Yangtze River, China, Journal of Hydrological Engineering 16, 914-919, 2011.

[9] [C.-M. Chou](http://link.springer.com/search?facet-creator=%22Chien-Ming+Chou%22), Complexity analysis of rainfall and runoff time series based on sample entropy in different temporal scales, [Stochastic Environmental Research and Risk Assessment](http://link.springer.com/journal/477) 28, 1401-1408, 2014.

[10] Y.V. Reddy, A. Sebastian, A. Parameters for estimation of entropy to study price manipulation in stock markets, Decision 34, 149, 2007. **SOURCE**

[11] A. [Holzinger](http://link.springer.com/search?facet-author=%22Andreas+Holzinger%22), [M. Hörtenhuber](http://link.springer.com/search?facet-author=%22Matthias+H%C3%B6rtenhuber%22), [C. Mayer](http://link.springer.com/search?facet-author=%22Christopher+Mayer%22), [M. Bachler](http://link.springer.com/search?facet-author=%22Martin+Bachler%22), [S. Wassertheurer](http://link.springer.com/search?facet-author=%22Siegfried+Wassertheurer%22), [A. J. Pinho](http://link.springer.com/search?facet-author=%22Armando+J.+Pinho%22), [D. Koslicki](http://link.springer.com/search?facet-author=%22David+Koslicki%22), On entropy based data mining, Ch: [Interactive Knowledge Discovery and Data Mining in Biomedical Informatics](http://link.springer.com/book/10.1007/978-3-662-43968-5), [Lecture Notes in Computer Science](http://link.springer.com/bookseries/558)  8401,  209-226. 2014.

[12] A. Widodo, M.-C. Shim, W. Caesarendra, B.-S. Yang, Intelligent prognostics for battery health monitoring based on sample entropy, Expert Systems with Applications 38, 11763-11769,2011.

[13] S. Pincus, B.H. Singer, Randomness and degrees of irregularity, Proc. Natl. Acad. Sci. USA 93, 2083-2088, 1996.

[14]T. Zhang, Z. Yang, J.H. Coote, Cross-sample entropy statistics as a measure of complexity and regularity of renal sympathetic nerve activity in the rat, Experimental Physiology 92, 659-669, 2007.

[15] J. S. Chang, S. D. Lee, G. Ju, J.-W. Kim, K. Ha, I.-Y. Yoon, Enhanced cardiorespiratory coupling in patients with obstructive sleep apnea following continuous positive airway pressure treatment, Sleep Medicine 14, 1132-1138,2013.

[16]R. Hernández-Pérez, L. Guzmán-Vargas, A. Ramírez-Rojas, F. Angulo-Brown, Pattern synchrony in electrical signals related to earthquake activity, Physica A 389, 1239-1252, 2010.

[17]L.-Z. Liu, X.-Y. Qian, H.-Y. Lu, Cross-sample entropy of foreign exchange time series, Physica A 389, 4785-4792,2010.

[18] . [W. Shi](http://link.springer.com/search?facet-author=%22Wenbin+Shi%22), [P. Shang](http://link.springer.com/search?facet-author=%22Pengjian+Shang%22), Cross-sample entropy statistic as a measure of synchronism and cross-correlation of stock market, [Nonlinear Dynamics](http://link.springer.com/journal/11071) 71, 539-554, 2013.

[19]  L.S de Araujo, J.R.S. Santos, M. Cunha Filho, B.D. Stosic, T. Stosic, Avaliação da influência do sistema Cantareira na bacia do Rio Piracicaba através da análise de complexidade (Evaluation of the influence of Cantareira system on the Piracicaba River basin, using complexity analysis), Revista Brasileira de Biometria,**São Paulo, v. 33, n.3, p.403-413, jul.-set. 2015.**

[20] M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals, Physical Review E 71, 021906, 2005.

[21] C.E. Shannon, A mathematical theory of communication, The Bell System Technical Journal 27,379-423, 1948.

[22] P. Grassberger, I. Procaccia, Estimation of the Kolmogorov entropy from a chaotic signals, Physical Review A 28, 2591-2593, 1983.

[23] M. Costa, C.-K. Peng, A.L. Goldberg, J. M. Hausdorff, Multiscale entropy of human gait dynamics, Physica A 330, 53-60, 2003.

[24]A. Catarino, O. Churches, S. Baron-Cohen, A. Andrade, H. Ring, Atypical EEG complexity in autism spectrum conditions: A multiscale entropy analysis, Clinical Neurophysiology 122,2375-2383, 2011.

[25] W.-Y. Pan, M.-C. Su, H.-T. Wu, M.-C. Lin, I.-T. Tsai, C.-K. Sun, Multiscale entropy analysis of heart rate variability for assessing the severity of sleep disordered breathing, Entropy 17, 231-243, 2015.

[26] E. Martina, E. Rodriguez, R. Escarela-Perez, J. Alvarez-Ramirez, Multiscale entropy analysis of crude oil price dynamics, Energy Economics 33, 936-947, 2011.

[27] J. Xia, P. Shang, J. Wang, W. Shi, Classifying of financial time series based on multiscale entropy and multiscale time irreversibility, Physica A 400,151-158,2014.

[28] L. Guzman-Vargas, A. Ramirez-Rojas, F. Ângulo-Brown, Multiscale entropy analysis of electroseismic time series, Natural Hazards and Earth System Sciences 8, 855-860, 2008.

[29] Z. Li, Y.-K. Zhang. Multi-scale entropy analysis of Mississippi River flow. Stochastic Environmental Research and Risk Assessment 22, 507-512, 2008.

[30] Y. Zhou, Q. Zhang, K. Li, X. Chen. Hydrological effects on water reservoirs on hydrological processes in the East River (China) basin: Complexity evaluations based on the multi-scale entropy analysis . Hydrological Processes 26, 3253-3262, 2012.

[31] J. Wang, P. Shang, X. Zhao, J. Shia. Multiscale entropy analysis of traffic time series.

*Int. J. Mod. Phys. C* 24, 1350006, 2013.