Proceedings of the Workshop Net-Works 2008,
Pamplona, 9–11 June, 2008, pp. 1–6.

Generalized Statistical Complexity Measure: A New Tool for Dynamical Systems

O.A.Rosso[1,2], H.A.Larrondo[3], M.T.Martin[4] and A.Plastino[4]

[1] *Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine, School of Electrical Engineering and Computer Science, The University of Newcastle, Australia*

[2] *Instituto de Cálculo, Universidad de Buenos Aires, Argentina*

[3] *Facultad de Ingeniería, Universidad Nacional de Mar del Plata, Argentina*

[4] *Instituto de Física-IFLP-CCT La Plata-Conicet, Universidad Nacional de La Plata, Argentina*

Abstract

The generalized Statistical Complexity Measure (SCM) is a functional that characterizes the probability distribution P associated to the time series generated by a dynamical system under study. It quantifies not only randomness but also the presence of correlational structures. In this seminar several fundamental issues are reviewed: *a)* selection of the information measure $\mathcal{I}$; *b)* selection of the probability metric space and its corresponding distance $\mathcal{D}$; *c)* definition of the generalized disequilibrium $\mathcal{Q}$; *d)* selection of the probability distribution P associated to a dynamical system or time series under study, which in fact, is a basic problem. Here we show that improvements can be expected if the underlying probability distribution is "extracted" by appropriate consideration regarding causal effects in the system's dynamics. Several well-known model-generated time series, usually regarded as being of either stochastic or chaotic nature, are analyzed. The main achievement of this approach is the possibility of clearly distinguish between them in the Entropy-Complexity representation space, something that is rather difficult otherwise.
Keywords: complexity measure

An information measure $\mathcal{I}$ can primarily be viewed as a quantity that characterizes a given probability distribution. $\mathcal{I}[P]$ is regarded as the measure of the uncertainty associated to the physical processes described by the probability distribution $P = \{p_j, j = 1, \cdots, N\}$, with N is the number of possible states of the system under study. If $\mathcal{I}[P] = 0$ we are in a position to predict with certainty which of the N possible outcomes will actually take place. Our knowledge of the underlying process described by the probability distribution is in this instance maximal. On the other hand, our ignorance is maximal if $\mathcal{I}[P] = \mathcal{I}_{max}$. This two extreme circumstances of *(i)* maximum foreknowledge ("perfect order") and *(ii)* maximum ignorance (or maximum "randomness") can, in a sense, be regarded as "trivial" ones. We define for a given probability distribution P and its associate information measure $\mathcal{I}[P]$, an amount of "disorder" H in the fashion

$$\mathcal{H}[P] \ = \ \mathcal{I}[P]/\mathcal{I}_{max} \ , \tag{1}$$

where $\mathcal{I}_{max} = \mathcal{I}[P_e]$ and P_e (uniform probability distribution) is the probability distribution which maximize the information measure. Then $0 \leq \mathcal{H} \leq 1$. It follows that a definition of *Statistical Complexity Measure* (SCM) must not be made in terms of just "disorder" or "information". It might seen reasonable to propose an SCM by adopting some kind of distance $\mathcal{D}$ from the equilibrium distribution P_e of the accessible states of the system [1, 2, 3, 4, 5, 6]. We define the "disequilibrium"

$$\mathcal{Q}[P] = \mathcal{Q}_0 \cdot \mathcal{D}[P, P_e] \ , \tag{2}$$

where $\mathcal{Q}_0$ is a normalization constant and $0 \leq \mathcal{Q} \leq 1$. The disequilibrium $\mathcal{Q}$ would reflect on the systems "architecture", being different from zero if there exist "privileged", or "more likely" states among the accessible ones. Consequently, we will adopt the following functional product form for the SCM introduced originally by Lopez-Ruiz, Mancini and Calbet [1]

$$\mathcal{C}[P] \ = \ \mathcal{H}[P] \cdot \mathcal{Q}[P] \ . \tag{3}$$

This quantity reflects on the interplay between the amounts of information stored in the system and its disequilibrium.

Here we will define $\mathcal{I}$ in terms of entropies. $\mathcal{H}$ refers to different entropic functional forms: *(a)* Shannon, $\mathcal{H}_1^{(S)}$; *(b)* Tsallis, $\mathcal{H}_q^{(T)}$; *(c)* escort-Tsallis $\mathcal{H}_q^{(G)}$; *(d)* Renyi $\mathcal{H}_q^{(R)}$. For the corresponding entropics forms see [4, 5]. q is a parameter ($q = 1$ for Shannon), which in the $q \to 1$ limit all the expression coincides with Shannon's measure.

We face infinite different choices for the metric and its induced distance $\mathcal{D}$ entering in $\mathcal{Q}$ definition. We consider here: *a)* Euclidean $\mathcal{D}_E$ [1]; *b)* Wootters's $\mathcal{D}_W$ [2]; *c)* relative Kullback entropies $K_q^{(\kappa)}$; *d)* Jensen divergences $\mathcal{J}_q^{(\kappa)}$; with $\kappa =$ Shannon (S), Tsallis (T), escort-Tsallis (G) and Renyi (R)) [4, 5].

On the basis of the functional product form we obtain then a family of SCM for each of three distinct disorder measures and disequilibriums, namely,

$$\mathcal{C}_{\nu,q}^{(\kappa)}[P] \;=\; \mathcal{H}_q^{(\kappa)}[P] \cdot \mathcal{Q}_q^{(\nu)}[P] \; . \tag{4}$$

with $\kappa =$ S, T, G, R for a fixed q. In Shannon instance ($\nu =$ S) we have, of course, $q = 1$. The index $\nu = E, W, K_q^{(\kappa)}, \mathcal{J}_q^{(\kappa)}$ tell us that the disequilibrium is to be evaluated with the appropriate distance measures.

It is interesting to note that for $\nu = K_q^{(\kappa)}$ the SCM family becomes $\mathcal{C}_q^{(\kappa)}[P] = (1 - \mathcal{H}_q^{(\kappa)}[P]) \cdot \mathcal{H}_q^{(\kappa)}[P]\cdot$ One could raise the objection that in this case this complexity family is just a simple function of the entropy. As a consequence, it might not contain new information vis-a-vis the measure of order. All the remaining members of the family $\mathcal{C}_{\nu,q}^{(\kappa)}$ (with $\nu \neq K_q^{(\kappa)}$) are not just a function of the entropy. On the contrary, for a given $\mathcal{H}_q^{(\kappa)}$-value, an ample range of SCM is obtained, from a minimum one $\mathcal{C}_{min}$ up to a maximal value $\mathcal{C}_{max}$. Evaluation of $\mathcal{C}_{\nu,q}^{(\kappa)}$ yields, consequently new information, according to the peculiarities of the pertinent probability distribution. A general procedure to obtain the bounds $\mathcal{C}_{min}$ and $\mathcal{C}_{max}$ corresponding to the generalized $\mathcal{C} = \mathcal{H} \cdot \mathcal{Q}$ -family is given in [4].

In statistical mechanics one is usually interested in isolated systems characterized by an initial, arbitrary, and discrete probability distribution. Evolution towards equilibrium is to be described. At equilibrium, the distribution is the equiprobability distribution. In order to study the time evolution of the SCM, a diagram of $\mathcal{C}$ versus time t can then be used. But, as we know, the second law of thermodynamics states that entropy grows monotonically with time ($d\mathcal{H}/dt \geq 0$). This implies that $\mathcal{H}$ can be regarded as an arrow of time, so that an equivalent way to study the time evolution of the SCM is to plot $\mathcal{C}$ versus $\mathcal{H}$. In this way, the normalized entropy substitutes the time axis.

If $\kappa = S$ ($q = 1$) and $\nu = E$ we recover the SCM of Lopez-Ruiz, Mancini and Calbet (LMC) [1], $\mathcal{C}_{LMC} = \mathcal{C}_{E,1}^{(S)}$. It has been pointed out by Crutchfield and co-workers [7] (that the LMC measure is marred by some troublesome characteristics that we list below:

- it is neither an intensive nor an extensive quantity.
- it vanishes exponentially in the thermodynamic limit for all one-dimensional, finite-range systems.

These authors forcefully argue that a reasonable SCM should

- be able to distinguish among different degrees of periodicity;
- vanish only for periodicity unity.

Finally, and with reference to the ability of the LMC measure to adequately capture essential dynamical aspects, some difficulties have also been encountered. With the product functional form for the generalized SCM it is

impossible to overcome the second deficiency mentioned above. In previous works we have shown that, after performing some suitable changes in the definition of the disequilibrium (utilization of Wootters distance [2] and Jensen divergence [3]) one is in a position to obtain a generalized SCM that is:
(i) able to grasp essential details of the dynamics,
(ii) an intensive quantity, and
(iii) capable of discerning among different degrees of periodicity.

An important point in the evaluation of the generalized SCM, is the determination of the underlying probability distribution P (associated to a given dynamical system or time series). This is an often neglected point that indeed deserves detailed consideration. The probability distribution P and the sample space Ω are inextricably linked. Many schemes have been proposed for a proper selection of the probability space (Ω, P). We can mention, among others: *(a)* procedures based on amplitude statistics, *(b)* binary symbolic dynamics, *(c)* Fourier analysis and, *(d)* wavelet transform. Their applicability depends on particular characteristics of the data such as stationarity, length of the time series, variation of the parameters, level of noise contamination, etc. In all these cases the global aspects of the dynamics can be somehow captured, but the different approaches are not equivalent in their ability to discern all the relevant physical details. One must also acknowledge the fact that the above techniques are introduced in a *rather ad hoc fashion and are not directly derived from the system under study's dynamical properties themselves,* as can be achieved, for instance, by recourse to the Bandt-Pompe methodology [8]. This requires suitable partitions of a D-dimensional embedding space that will, it is hoped, reveal important details concerning the ordinal structure of a given one-dimensional time series.

The Bandt-Pompe method for evaluating the probability distribution P is based on the details of the system's attractors reconstruction procedure. *Causal information* is, consequently, properly incorporated into the construction process that yields (Ω, P). The Bandt-Pompe probability distribution is the only one among those in popular use that takes into account the temporal structure of the time series generated by the physical process under study. A notable result from the Bandt-Pompe approach is a notorious improvement in the performance of the information quantifiers, like entropy and statistical complexity measures, obtained using the probability distribution P generated by their algorithm. Of course, one must assume with the Bandt-Pompe methodology that the system is weakly stationary and that enough data are available for a correct attractor reconstruction.

Although being of a quite different physical origin, time series arising from chaotic systems share with those generated by stochastic processes several

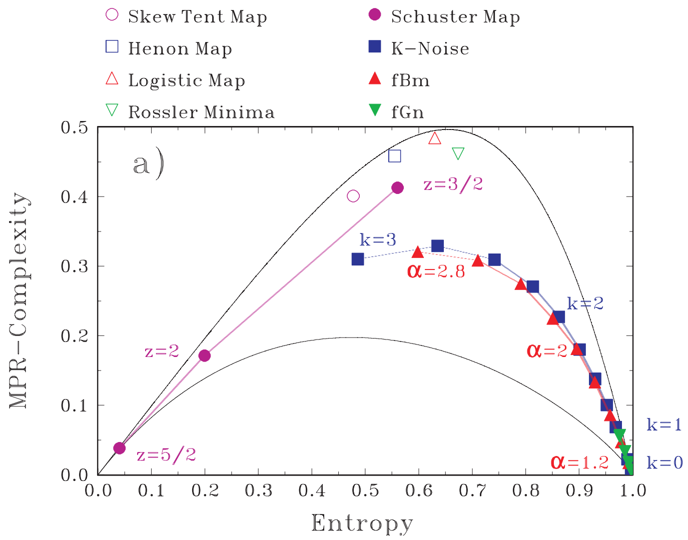

Figure 1: Localization of different chaotic systems and stochastic processes in the $\mathcal{C} \times \mathcal{H}$-plane. Continuous lines represent both minimum C_{min} and maximum C_{max} complexities. The area enclosed by them is the $\mathcal{C} \times \mathcal{H}$-plane. In the case of text-book models usually regarded as being of either stochastic or deterministic nature, our numerical results place them at clearly different planar locations.

properties that make them almost undistinguishable: (1) a wide-band power spectrum, (2) a delta-like autocorrelation function, (3) an irregular behavior of the measured signals, etc. In fact this similitude has made it possible to replace stochastic processes by chaotic systems in many practical applications. Note also, that chaotic systems always produce time series with a physical structure. Recently, by recourse of the $\mathcal{C} \times \mathcal{H}$-plane we have illustrated on the possibility of clearly distinguishing between them in our representation space, something that is rather difficult otherwise [6].

For chaotic systems we consider: *(a)* The logistic map ($r = 4$); *(b)* The Skew tent map ($\omega = 0.1847$); *(d)* The Henon's map ($a = 1.4$, $b = 0.3$); *(d)* The Lorenz map of Rossler's oscillator ($a = b = 0.2$, $c = 5.7$); *(e)* Shuster maps (intermittent signals with chaotic bursts that also display $1/|f|^z$ power spectrum). We consider for stochastic processes: *(f)* Noises with $1/|f|^k$ power spectrum; *(g)* Fractional Brownian Motion (fBm) and fractional Gausian Noise (fGn) (for these process is possible define a generalized power spectrum of the form $1/|f|^\alpha$, with $1 < \alpha < 3$ for fBm and $-1 < \alpha < 1$ for fGn). For all the cases we studied 10 time series of 2^{15} *data* each were analyzed, each series starting at a different initial condition. The concomitant mean values of both $\mathcal{H}_1^{(S)}$ and $\mathcal{C}_{\mathcal{J},1}^{(S)}$ are plotted in Fig. 1. These two functionals are evaluated using the Bandt-Pompe recipe (with embedding dimension $D = 6$) to assign a probability distribution function to the time series generated by the system.

Some results are displayed in Fig. 1 (for a detailed discussion see Ref. [6]). They show that our representation plane accommodates noise and chaos at clearly different planar locations. Such property could be useful when dealing with real data (that always have a stochastic component due to omnipresent dynamical noise) so as to classify different degrees of stochasticity. Our representation also distinguishes *(a)* Gaussian from non-Gaussian process and, *(b)* amongst different correlation degrees (see the colored noises). Consequently, this representation plane is an effective tool for revealing the sometimes subtle difference between noise and chaos.

References

[1] R. Lopez-Ruiz, H. L. Mancini and X. Calbet, "A statistical measure of complexity", *Phys. Lett. A.* **209** (1995) 321–326.

[2] M. T. Martin, A. Plastino and O. A. Rosso, "Statistical Complexity and disequilibrium", *Phys. Lett. A.* **311** (2003) 126–132.

[3] P. W. Lamberti, M. T. Martin, A. Plastino and O. A. Rosso, "Intensive entropic non-triviality measure", *Physica A.* **334** (2004) 119–131.

[4] M. T. Martin, A. Plastino and O. A. Rosso, "Generalized statistical complexity measures: Geometrical and analytical properties", *Physica A.* **369** (2006) 439–462.

[5] O. A. Rosso, M. T. Martin, A. Figliola, K. Keller and A. Plastino, "EEG analysis using wavelet-based information tools", *J. Neurosci. Meth.* **153** (2006) 163–182.

[6] O. A. Rosso,H. L. Larrondo, M. T. Martin, A. Plastino and M. A. Fuentes, "Distinguishing noise from chaos", *Phys. Rev. Lett.* **99** (2007) 154102.

[7] D. P. Feldman and J. P. Crutchfield, "Measures of statistical complexity: why?", *Phys. Lett. A.* **238** (1998) 244–252.

[8] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series", *Phys. Rev. Lett.* **88** (2002) 174102.