Parameter-free quantification of stochastic and chaotic signals

S.R. Lopes, T.L. Prado, G. Corso, G.Z. dos S. Lima, J. Kurths

1. Introduction

Two of the foremost characteristics of a stochastic signal are its possible temporal correlation, preserving memory for some interval of time [1] and distinct characteristics of its probability distribution functions (PDF), that show information about how frequent can be the presence of a single element or a set of elements in a signal. Although the literature does not provide a formal and unique definition for signal complexity [2], both mentioned characteristics are related to this concept, providing some quantification about the signal.

Here we show how the recurrence entropy concept [3] can be used as a new parameter-free tool to quantify the time correlation of a signal. Additionally, using similar quantification procedure, the same tool can be used to evaluate how distinct can be the entire set of elements of a particular signal. The former concept can be better understood comparing it to the quantity of letters used to compose an idiom. We can use few letters but, in this case, the words will be more similar to each other. In the same way, time correlations implicitly impose some rules that make a particular sequence of elements of a signal more similar to others, diminishing the quantity of distinct sequences in a signal.

A common approach to characterize stochastic signal properties is to use information entropy, describing the amount of data needed to identify characteristics of a system [4]. Additionally, information entropy is also a fundamental concept to understand chaotic dynamics [5] and can be related to the level of chaos or the chaoticity of the system, mainly measured by the Lyapunov exponent [3,5].

Consider the case of a class of stochastic signals following a power-law distribution $P(f) \propto 1/f^\alpha$, where $\alpha$ quantifies the time correlation [16]. Specific values of $\alpha$ are associated with colors e.g., $\alpha = 0$ for “white”, $\alpha = 1$ for “pink” or $\alpha = 2$ for “red”, the later also known as Brownian noise. Stochastic processes with $1/f^\alpha$ power spectra are ubiquitous in science finding applications in all its subareas like physics [7–10], engineering [11], biology [12–14], cognition [15], astrophysics [8,9], geophysics [8,16], economics [17], psychology [18], language and music [19]. A useful estimation method for long-time correlation based on finite time series is a key issue and, hitherto, an open question in time series analyses [20–22]. Many methods described until now in the literature are based on just time correlation quantification such as the computation of the Hurst exponent [20,21], detrended fluctuation analysis (DFA) [22] or range-scaled analysis [21]. Others are computed on
the frequency or wavelet domains like periodogram or Wavelet methods [23].

In general, finite time-interval of sampled signals makes the characterization of time-correlation via those traditional methods a sophisticated technique since a certain level of non-stationary is imposed by the long-range dependence (manly for $\alpha > 1$) of the signal. Often the analyses lead to parameter dependent results and, since empirical time series are always finite, long-range correlations are, unavoidably, partly suppressed. On the other hand, the local dynamics characteristics of small temporal windows tend to be overestimated.

The second relevant point is the quantification of special properties of the entire set of elements of a signal. Usually the methods employ the measure of entropies [4,24,25], but they do not evaluate time correlations or are parameter dependent.

In this context, the evaluation of the recurrence entropy [3] of a signal and our definition of $\text{max}(S)$ show to be a powerful parameter-free tool to examine time series correlations. Here, we show that the new approach can evaluate short and long time correlations, possesses a good agreement with traditional methods, but going further, provides information about characteristics of the entire set of points of a signal.

2. The recurrence entropy

A visual tool to display recurrences of a $K$ length time series is defined as a $K \times K$ binary matrix [26]

$$\textbf{R}_{ij} = \begin{cases} 1, & \text{if } |x_i - x_j| \leq \epsilon, \\ 0, & \text{if } |x_i - x_j| > \epsilon, \\ i, j = 1 \cdots K, \end{cases}$$

where $\epsilon$ is the vicinity parameter. $\textbf{R}$ summarizes visually, in a binary pattern, the information about how many recognizable subsets are embedded in a $K$ sequence of data showing how distinct will be the recurrence pattern (sequences of zeros and ones) of $K$ consecutive points. The most explored subsets of $\textbf{R}$ are diagonal lines of "ones" representing the mutual recurrences of a sequence of points. However, other structures of $\textbf{R}$ also have dynamical interpretations: the vertical/horizontal lines are associated to stationary points and the abundance of isolated points is an indicative of chaotic or stochastic dynamics [26]. We generalize these concepts defining recurrence microstates $A(\epsilon)$ as all possible cross-recurrence states among two randomly selected short sequences of $N$ consecutive points in a $K (K \gg N)$ length time series (we use $N = 2, 3, 4$), namely $A(\epsilon)$ are $N \times N$ small binary matrices. For example, supposing a time series of $K$ elements, $\{a_1, a_2, \ldots, a_K\}$, and using $N = 2$, we randomly select two sequences of two elements, say $\{a_{01}, a_{11}\}$ and $\{a_{02}, a_{12}\}$. In the case of $N = 2$ our microstates $A(\epsilon)$ will be binary numbers composed of four elements, a $2 \times 2$ binary matrix expressing the cross-recurrences among $a_{10}$ and $a_{11}$, $a_{10}$ and $a_{12}$, $a_{11}$ and $a_{12}$, and finally, $a_{01}$, $a_{02}$. For a large enough randomly selected number of samples $M$, the recurrence entropy $S$ is adequately computed by [3]

$$S(A) = -\sum_A P_A \ln P_A,$$

where $P_A$ measures the probability of occurrence of a specific state $A(\epsilon)$ considering $M$ randomly samples. Usually, $\epsilon$ is an adjustable parameter as Eq. 1 and $A(\epsilon)$ suggest, but this dependence is eliminated observing that $S$ is null when computed for sufficient large or small $\epsilon$, due to the absence of diversity of $A(\epsilon)$ for both cases. So, we impose a natural condition of a maximum for $S(\epsilon)$ as a function of $\epsilon$ [27] turning $\max(S(\epsilon)) \equiv \max(S)$ and $A(\epsilon) \equiv A$ in parameter-free quantities.

At first sight, $M$ should be larger than the quantity of all possible microstates $2^{MN}$, but as observed in [3] the number of microstates effectively populated is small and the convergence of Eq. (2) is fast. So a much smaller number of randomly select microstates $M \sim 10,000$ is enough for good results in a large variety of cases and, in special, for all cases treated here, turning the method fast even for microstate sizes $N = 4$.

3. Time correlated stochastic signals analyses

Firstly, we consider time series of Gaussian distributed stochastic signals [28], characterized by a power spectrum $P(f) \propto 1/f^n$ for $\alpha \in [0,2]$. Examples of the mean power spectrum obtained from 500 time series are plotted in Fig. 1(a) for 5 distinct values of $\alpha$. Corresponding individual time series examples are plotted in Figs. 1(b-f).

All properties of the stochastic signal are kept constant in the following analyses, but values of $\alpha \geq 1.0$ impose a finite level of non-stationarity due to long-term correlations associated to finite time sampling, as observed in Figs. 1(e, f). For such cases, correlation-based methods overestimate (underestimate) short/long-term correlations.

Fig. 2 (a) depicts the results of $\max(S)$ computed for distinct colored stochastic signals ($0 \leq \alpha \leq 2$) for 3 values of $N$ and 3 time series lengths. In general, $\max(S)$ displays a smooth logistic shaped curve as a function of $\alpha$. For vanishing values of $\alpha$, $\max(S)$ asymptotes its maximum theoretical values $N^2 \ln 2$, obtained for uncorrelated stochastic signals and infinite time series lengths. For the interval $0 \leq \alpha \leq 2$, similar results for distinct $N$ show that the variability of $\max(S)$ as a function of $\alpha$ is measurable even for the smallest possible value of the microstate matrix size $N = 2$. Another important conclusion is that for a fixed $N$, longer time series lead to smaller values of $\max(S)$ since longer time series provide a better evaluation of long-term time correlations. An error bar analysis specially for $N = 4$ indicates that smaller time series associated to larger microstate size and $\alpha$ values result in larger dispersion of $\max(S)$. This behavior reveals the natural dispersion expected for a quantification of long-term correlations when just finite time series are used. The results for $N = 2$ and $N = 3$ are less sensitive to the natural dispersion since the number of possible microstates are also smaller, such that tiny changes of the time correlation are not captured. All these features explored at the same time bring useful results when unknown source signals are analyzed. Fig. 2(b) displays all curves depicted in Fig. 1(a) but normalized by its respective maximum. The data collapse reveals that the shape of $\max(S)$ for all time series lengths and all microstate sizes are equivalent, despite the small differences and details discussed above showing that the results are independent of $N$.

![Fig. 1](image_url)  
**Fig. 1.** (a) Power spectral densities as a function of frequency $(P(f) \propto 1/f^n)$, for distinct $\alpha$ values. (b)-(f) Corresponding time series of the stochastic signals for all five values of $\alpha$. 


Additionally to the quantification of the time correlation of the stochastic signal in a unique number, another important questions is how to evaluate properties of the entire set of points of the time series. To do so, we destroy the time correlation making use of surrogate data analysis [29]. In particular, we shuffle the data, keeping only the collective properties due only to the entire set of points. The same analysis using recurrence entropy is then carried out to characterize the distribution of elements of the time series. Traditional methods like Hurst exponents and detrended fluctuation analysis only quantify the time correlation [1,2,23] and are not suitable for surrogate data, but as observed in Fig. 3, results of $\max(S)$ as a function of $\alpha$ still bring useful results. Fig. 3 depicts values of $\max(S)$ applied to the same data used in Fig. 2 for $K = 2^8$ and 3 values of $N$, but now shuffled in a random sequence (Fisher-Yates algorithm [30]). The results of $\max(S)$ reveal a new question: even when the sequence of points in the time series is randomly organized, distinct stochastic signals lead to distinct values of $\max(S)$. So the behavior of $\max(S)$, in this case, is due only to properties of the set of points of the time series. The results point out for a clear distinction between all our shuffled time series and makes us to conclude that long-term correlations imposed by larger $\alpha$ results in a smaller value of $\max(S)$ reflecting a more restrictive and (in same sense) more organized time series, due to restrictions imposed originally by the time correlation. So, we say that time correlations impose an upper limit to all possible sequence of subsets in the time series and some combinations will not be allowed, in a similar way to the situation of an idiom built using just few letters, where the words will display similarity due to the scarcity of letters. Quantification of randomly shuffled data can also be done using traditional recurrence analysis, like the recurrence rate, and/or determinism [26], however as far as we have tested, results using recurrence entropy are more sensitive to small changes in the signals.

To make this point clear, we analyze the results of $\max(S)$ obtained from the normalized time series built from the rule

$$x(t) = 5(\sin(t)) + \sigma \mathcal{F}(t),$$

where $\mathcal{F}$ intends for a shuffled process of the sine function following the rule described in [30]. $\mathcal{F}$ is an uncorrelated Gaussian noise, and $\sigma$ measures its amplitude level that is superposed to the shuffled harmonic signal. Thus, the stochasticity of this example comes from 2 sources: the shuffled process in the sine signal and the random noise generator. Despite to be a stochastic time series due to the shuffling of the sine domain of points, it will present a very low entropy due to the very limited sine domain. Corroborating this analysis, Fig. 4 depicts results of $\max(S)$ as a function of $\sigma$ for 3 values of $N$. For $\sigma = 0$, the signal is just an uncorrelated set of points obtained from the shuffled domain of the $\sin(t)$. So, in this case, $\max(S)$ is consistently smaller than those ones expected for uncorrelated stochastic signal that are plotted as black lines in all panels. For $0 \leq \sigma \leq 2$, $\max(S)$ grows monotonically, pointing out for an increasing number of distinct recurrence entropy microstates of the data set since the stochastic perturbation amplitude is being increased. For $\sigma > 2$, the uncorrelated stochasticity is large enough to turn the time series into an (practically) uncorrelated stochastic time series. Consequently, $\max(S)$ asymptotically reaches the expected value for uncorrelated noise. So, $\max(S)$ is capable to capture the progressive increase of the number of distinct microstates imposed by the uncorrelated stochastic signal summed to the original time series produced by the shuffled sin function domain.

### 4. Deterministic signals analyses

To prove the ability of $\max(S)$ to capture distinct characteristics of even more complex distributions of points in a time series, we analyze time series obtained by the generalized Bernoulli chaotic map

$$x = \beta x \quad (\text{mod } 1).$$
For $\beta > 1$ the level of chaoticity is evaluated by the Lyapunov exponent $\lambda = \ln \beta$ [31]. Entropy measures are expected to be related to $\lambda$ but not necessarily directly proportional since the entropy is also a function of the distribution of elements of the set of points of the attractor (the invariant measure) $\rho(x)$. The quantity $\rho$ generated from Eq. 4 is homogeneous for integer $\beta$, but becomes inhomogeneous for non-integer values [32], due to the discontinuities observed in the PDF, result of an inhomogeneous measure of the attractor and a corresponding more complex signal. Figs. 5(a-d) display $\rho(x)$ for Eq. 4 map, depicting more complex PDFs for non-integer values of $\beta$ (a-c), but collapsing in a homogene- ous one for integer $\beta$ (d). The entropy $\max(S)$ as a function of $\beta$ will be dependent of two factors: namely the continuous growing chaotic associated to parameter $\beta$ superposed by an also varying complex behavior of the PDF. Panel (e) shows the behavior of $\max(S)$ in the interval $2 < \beta < 4$ (blue curve). The dashed tone of blue is representative of the standard deviation of $\max(S)$ due to 100 initial conditions for each $\beta$. To evaluate just the effect of changes in the PDF, the black curve in panel (e) depicts $\max(S)$ computed for shuffled time series (out of y-scale magnification also shown for better visibility). Again, the dashed black tone indicates the standard deviation over 100 initial conditions. In this case, $\max(S)$ depicts a complex oscillatory pattern due to the behavior of the PDF as discussed before. In general $\max(S)$ grows as a function of $\beta$. However, the growth rate is faster for values of $\beta$ departing from integers, diminishing as $\beta$ approximates from the next subsequent integer. Such behavior can be understood since for values immediately larger than each integer, the simultaneous increases of chaoticity and complexity of the invariant map measure lead to a (local) maximum growth rate of $\max(S)$.

As $\beta$ approaches to an integer, the map invariant measure turns to be flatter and consequently the complexity of the map invariant measure decreases, but at the same time, the chaoticity level increases since the Lyapunov exponent is proportional to $\beta$. Due to both dependencies of $\max(S)$, its growth rate is smaller.

For large values of $\beta$ as the interval $\beta \leq 4$ the growth rate can be even negative since a progressive less complex invariant map measure can overtake the effect of the increasing of chaoticity. The value of $\max(S)$ also reflects specific more dramatic changes of the map invariant measure as the example highlighted by the arrow around $\beta \approx 2.3$, where a clear change in the invariant measure revealed by the shuffled time series analysis (black curve) leads to local small changes in the growth rate of $\max(S)$ (blue curve). In resume the complex change behaviors of the invariant map measure lead to a rich fine structure in $\max(S)$ computed over the shuffled time series.

5. Conclusions

In conclusion, we have shown that $\max(S)$ is a parameter-free tool that quantifies time correlation of stochastic and chaotic signals. Additionally it goes further, evaluating subtle properties of the entire set of points of a signal, that can be computed using simple shuffling of the points. When time correlation is evaluated, $\max(S)$ brings similar results to those obtained for traditional, but parameter-dependent, quantifiers, such as the Hurst exponent. However, the use of $\max(S)$ makes clear a more complex inter- relation about properties of the set of points and the complexity of the time series, bringing new perspective for stochastic and chaotic data analyses. Our results can be useful in the analysis of experimental noisy data such as seismic, paleontology, economic problems where the possibility to evaluate properties of the entire data set data associated with the quantification of time correlation are important.

Analyzing stochastic and deterministic signals we conclude that the new method identify and quantify a new cause/effect relation, where changes occurring in the time series set of points can be related directly to variations of short and long time correlations. It is worth to mention that due to its computation methodology [3], recurrence entropy is fast evaluated for arbitrary long real-world time series, leading to robust parameter-free way to process data.

An important question consists on the fact that $\max(S)$ can be used to distinguish non-correlated stochastic signals from deterministic ones. However, for time correlated stochastic signals only the value of $\max(S)$ is not enough to make the distinction. For these cases the complete set of microstates of the recurrence entropy must be evaluated.

A last point to mention is the fact that we have focused our results in the interval $0 < \alpha < 2$ due to its main interest in many areas, but $\max(S)$ can also be used in other intervals. In fact, recently, results from $\max(S)$ have been used to confirm a quite old interpretation (backs to1980) that chaotic defects exhibited by coupled map lattices with nearest-neighbor coupling display random behavior and diffuse along the lattice with a Gaussian signature [33]. The use of $\max(S)$ has helped to show that the motion of chaotic defects is well-represented by a stochastic time series with a power-law spectrum $1/f^\alpha$ with $2.3 < \alpha < 2.4$, i.e., a correlated Brownian motion. Specifically for the interval $-1 < \alpha < 0$ the distinction among near values of $\alpha$ using small size microstate values can turn the distinction still possible but harder, demanding larger temporal time series.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
CRediT authorship contribution statement

**S.R. Lopes:** Conceptualization, Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review & editing.

**T.L. Prado:** Formal analysis, Investigation, Methodology, Writing - review & editing.

**G. Corso:** Methodology, Writing - review & editing.

**G.Z. dos S. Lima:** Methodology, Writing - review & editing.

**J. Kurths:** Formal analysis, Writing - review & editing.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001 and trough project number 88881.119252/2016-01, Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq - Brazil, grant number 302785/2017-5, and Financiadora de Estudos e Projetos (FINEP).

References

[1] Beran J. Statistics for long-memory processes. Routledge; 2017.
[2] Ziemelis K. Nature insight (2001) complex systems. Nature 2001;410:241.
[3] Corso G, Prado TDL, Lima GZdS, Kurths J, Lopes SR. Quantifying entropy using recurrence matrix microstates. Chaos 2018;28:083108.
[4] Shannon CE. Bell system tech. j. 27 (1948) 379; ce shannon. Bell System Tech J 1948;27:218.
[5] Ott E. Chaos in dynamical systems. Cambridge university press; 2002.
[6] Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974;19:716–23.
[7] Bak P, Tang C, Wiesenfeld K. Self-organized criticality: an explanation of the 1/f noise. Phys Rev Lett 1987;59:381.
[8] Weissman M. 1/f noise and other slow, nonexponential kinetics in condensed matter. Rev Mod Phys 1988;60:537.
[9] Press WH. Flicker noises in astronomy and elsewhere. Comments Astrophys 1978;7:103–19.
[10] dos Santos Lima G, Corrêa M, Sommer R, Bohn F. Multifractality in domain wall dynamics of a ferromagnetic film. Phys Rev E 2012;86:066117.
[11] Hooge F, Kleippenning T, Vandamme L. Experimental studies on 1/f noise. Rep Prog Phys 1981;44:479.
[12] Glass L. Synchronization and rhythmic processes in physiology. Nature 2001;410:277.
[13] West BJ, Shlesinger M. The noise in natural phenomena. Am Sci 1990;78:40–5.
[14] dos Santos Lima G, Lobao-Saares B, Nascimento Gd, Franca AS, Muratori L, Ribeiro S, Corso G. Mouse activity across time scales: fractal scenarios. Plos One 2014;9:e105092.
[15] Gilden DL, Thornton T, Mallon MW. 1/f noise in human cognition. Science 1995;267:1837.
[16] Matthaeus W, Goldstein M. Low-frequency 1/f noise in the interplanetary magnetic field. Phys Rev Lett 1986;57:495.
[17] Granger CW, Ding Z. Varieties of long memory models. J Econom 1996;73:61–77.
[18] Gilden DL. Cognitive emissions of 1/f noise. Psychol Rev 2001;108:33.
[19] Voss RF, Clarke J. 1/f noise in music and speech. Nature 1975;258:317–18.
[20] Carbone A. Algorithm to estimate the hurst exponent of high-dimensional fractals. Physical Review E 2007;76:056703.
[21] Weron R. Estimating long-range dependence: finite sample properties and confidence intervals. Physica A 2002:312:285–99.
[22] Podobnik B, Stanley HE. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. Phys Rev Lett 2008;100:084102.
[23] Simonsen I, Hansen A, Nes OM. Determination of the hurst exponent by use of wavelet transforms. Phys Rev E 1998;58:2779.
[24] Bandt C, Pompe B. Permutation entropy: a natural complexity measure for time series. Phys Rev Lett 2002;88:174102.
[25] Eroglu D, Peron TKD, Marwan N, Rodrigues FA, Costa LD, Sebek M, Kiss IZ, Kurths J. Entropy of weighted recurrence plots. Physical Review E 2014;90:042919.
[26] Marwan N, Romano MC, Thiel M, Kurths J. Recurrence plots for the analysis of complex systems. Phys Rev 2007;438:237–329.
[27] Jaynes ET. Information theory and statistical mechanics. Phys Rev 1957;106:620.
[28] Kasdin NJ. Discrete simulation of colored noise and stochastic processes and 1/falpha/spl alpha/power law noise generation. Proc IEEE 1995;83:802–27.
[29] Hair Jr JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 7th Pearson; 2010.
[30] Fisher RA, Yates F. Statistical tables for biological, agricultural, and medical research. 6th ed, rev and enlarged. Edinburgh Oliver and Boyd; 1963.
[31] Allgood KT, Sauer TD, Yorke JA. Chaos. Springer; 1996.
[32] Góra P. Invariant densities for piecewise linear maps of the unit interval. Ergodic Theory Dyn Syst 2006;29:1549–83.
[33] Silva STD, Prado TL, Lopes SR, Viana RL. Correlated brownian motion and diffusion of defects in spatially extended chaotic systems. Chaos 2019;29:071104.