Permutation Entropy: A Natural Complexity Measure for Time Series

Christoph Bandt and Bernd Pompe

Institute of Mathematics and Institute of Physics, University of Greifswald, Greifswald, Germany
(Received 19 June 2001; revised manuscript received 20 December 2001; published 11 April 2002)

We introduce complexity parameters for time series based on comparison of neighboring values. The definition directly applies to arbitrary real-world data. For some well-known chaotic dynamical systems it is shown that our complexity behaves similar to Lyapunov exponents, and is particularly useful in the presence of dynamical or observational noise. The advantages of our method are its simplicity, extremely fast calculation, robustness, and invariance with respect to nonlinear monotonous transformations.

DOI: 10.1103/PhysRevLett.88.174102 PACS numbers: 05.45.Tp, 02.50.–r

I. Complexity measures for time series.—Various measures of complexity were developed to compare time series and distinguish regular (e.g., periodic), chaotic, and random behavior. Among others, it has been reported that complexity of heart and brain data can distinguish healthy and sick subjects and sometimes even predict heart attack or epileptic seizure [1]. The main types of complexity parameters are *entropies, fractal dimensions, and Lyapunov exponents*. They are all defined for typical orbits of presumably ergodic dynamical systems, and there are profound relations between these quantities [2,3].

Problem: The basic conceptual problem is that *these definitions are not made for an arbitrary series of observations* $\{x_1, x_2, \ldots\}$. As a consequence, there is also a computational problem. Many ingenious algorithms, tricks, and recipes have been developed during the last 20 years in order to estimate complexity measures from real-world time series [4–7]. They work wonderfully when the time series is simulated from a low-dimensional dynamical system, but most of them break down as soon as noise is added to the series. For real-world series, "noise elimination" requires careful preprocessing of the data and fine-tuning of parameters, and the *results cannot be reproduced without specifying details of the method*. One idea to overcome these problems is the application of Kolmogorov-Chaitin *algorithmic complexity* to orbits of dynamical systems [7–9].

Our approach: We go another way, defining simple complexity measures which are easily calculated for *any* type of time series, be it regular, chaotic, noisy, or reality based. Practical and theoretical examples were chosen to compare our complexity with established concepts. In Sec. III, we recognize voiced sounds in a speech signal. In Sec. IV, we show for the well-known family of logistic maps that our *permutation entropy* is very similar to Lyapunov exponents over the range of several thousand parameter values. Most importantly, it yields meaningful results in the presence of observational and dynamical noise (Sec. V).

Source entropy: Given a stationary time series $\{x_t\}$, most complexity parameters measure in one or the other way the multitude of different vectors $(x_{t+1}, \ldots, x_{t+n})$ for various n, passing then with n to ∞. If the x_t attain a finite number M of values, the classical *source entropy h of Shannon* measures the mean conditional uncertainty of the future x_{t+1} given the whole past $\ldots, x_{t-1}, x_t$. We have $0 \le h \le \log M$, with $h = 0$ if the series is perfectly predictable from the past and $h = \log M$ iff all values are independent and uniformly distributed. Large h indicates high complexity. Shannon's entropy can be generalized by Rényi's α entropies [2,4,10].

Partitions: If the x_t attain infinitely many values, it is common to replace them with a symbol sequence $\{s_t\}$ with finitely many symbols, and calculate source entropy for the s_t. We can use a partition $X = P_1 \cup \cdots \cup P_m$ of the set of values and define $s_t = i$ if x_t is in P_i. Then we can increase the number m of pieces and get a limit for fine resolution which is the *Kolmogorov-Sinai entropy* h_{K}, an isomorphism invariant of the dynamical system [2,3,11]. For special *generating* partitions, no such limit is needed. These partitions are difficult to find, however, even for two-dimensional examples like the Henon system [12]. For unimodal maps, the critical point defines a generating partition [13], but any misplacement of the partition point gives an entropy smaller than h_{K} [14]. Our viewpoint (cf. [15]) is that *the symbol sequence must come naturally from the x_t, without further model assumptions*. Thus we suggest to take partitions given by comparison of neighboring values x_t. For interval maps, this is similar to using generating partitions [16].

II. Basic definitions.—The following definitions apply to an arbitrary time series, with a weak stationarity assumption explained below. In the sequel we neglect equal values $x_{t^*} = x_t$, $t^* \neq t$, and consider only inequalities between the x_t. This is justified if the values x_t have a continuous distribution so that equal values are very rare. Otherwise, we can numerically break equalities by adding small random perturbations.

Our entropies are calculated for different embedding dimensions n, but we do not attempt to determine a limit for large n although this is an interesting theoretical problem [16]. For practical purposes, we recommend $n = 3, \ldots, 7$.

Example: Let us take a series with seven values:

$$x = (4, 7, 9, 10, 6, 11, 3).$$

We organize the six pairs of neighbors, according to their relative values, finding four pairs for which $x_t < x_{t+1}$ and two pairs for which $x_t > x_{t+1}$. So four of six pairs of values are represented by the permutation 01 ($x_t < x_{t+1}$) and two of six are represented by 10. We define the permutation entropy of order $n = 2$ as a measure of the probabilities of the permutations 01 and 10. So,

$$H(2) = -(4/6)\log(4/6) - (2/6)\log(2/6) \approx 0.918.$$

As usual, log is with base 2, thus H is given in bit. Next, we compare three consecutive values. $(4, 7, 9)$ and $(7, 9, 10)$ represent the permutation 012 since they are in increasing order. $(9, 10, 6)$ and $(6, 11, 3)$ correspond to the permutation 201 since $x_{t+2} < x_t < x_{t+1}$, while $(10, 6, 11)$ has the permutation type 102 with $x_{t+1} < x_t < x_{t+2}$. The permutation entropy of order $n = 3$ is $H(3) = -2(2/5)\log(2/5) - (1/5)\log(1/5) \approx 1.522$.

Definition: Consider a time series $\{x_t\}_{t=1,\ldots,T}$. We study all $n!$ permutations π of order n which are considered here as possible order types of n different numbers. For each π we determine the relative frequency (# means number)

$$p(\pi) = \frac{\#\{t \mid t \leq T - n, (x_{t+1}, \ldots, x_{t+n}) \text{ has type } \pi\}}{T - n + 1}.$$

This estimates the frequency of π as good as possible for a finite series of values. To determine $p(\pi)$ exactly, we have to assume an infinite time series $\{x_1, x_2, \ldots\}$ and take the limit for $T \to \infty$ in the above formula. This limit exists with probability 1 when the underlying stochastic process fulfills a very weak stationarity condition: for $k \leq n$, the probability for $x_t < x_{t+k}$ should not depend on t.

The permutation entropy of order $n \geq 2$ is defined as

$$H(n) = -\sum p(\pi)\log p(\pi),$$

where the sum runs over all $n!$ permutations π of order n. This is the information contained in comparing n consecutive values of the time series. It is clear that $0 \leq H(n) \leq \log n!$ where the lower bound is attained for an increasing or decreasing sequence of values, and the upper bound for a completely random system (i.i.d. sequence) where all $n!$ possible permutations appear with the same probability. The time series presents some sort of dynamics when $H(n) < \log n!$. Actually, in our experiments with chaotic time series, $H(n)$ did increase at most linearly with n; see Sec. IV and [16]. Thus we define the *permutation entropy per symbol* of order n, dividing by $n - 1$ since comparisons start with the second value:

$$h_n = H(n)/(n - 1).$$

Related entropies: We can also determine the information contained in sorting the nth value among the previous $n - 1$ when their order is already known: $d_n = H(n) - H(n - 1)$, $d_2 = H(2)$. This could be called *sorting entropy* of order n. Clearly, $h_n = (d_2 + \cdots + d_n)/(n - 1)$ which means that h_n is a stable parameter characterizing the overall behavior of the time series, while the d_n describe more special properties connected, for instance, with points of period n. In some cases, d_n can increase with n.

Instead of $H(n)$, we also studied $H^0(n) = \log\#\{\pi \text{ of order } n \mid p(\pi) > 0\}$ and defined a corresponding $h_n^0 = H^0(n)/(n - 1)$. This definition is simpler, but h_n^0 is not robust under noise, is influenced by outliers, and needs more data for accurate estimates.

III. Results for a speech signal.—To demonstrate how permutation entropy applies to real data, we consider in Fig. 1a the utterance "permutation entropy measures complexity" spoken by a male speaker in a modal register. The signal was sampled at 11 kHz, using a low cost PC sound card with standard data preprocessing (antialiasing low pass), without special filtering. The signal was analyzed using a sliding window of $T_{\text{win}} = 512$ samples, that is, 46 ms, and a one sample window shift. The 43 000 windows of the 4 s signal were processed on a PC in less than a second, which means that permutation entropy can be used in real-time applications. Since the window contains less than 6! samples, we chose only low order $n = 3, 4, 5$ for Fig. 1b.

The signal starts and ends with noise, and the corresponding normalized entropies $H(n)/\log n!$ are close to 1, as well as for unvoiced sounds, while all voiced sounds can be recognized by the decrease of entropy.

A well-known complexity parameter for short-time speech analysis is the zero-crossing rate (ZCR) [17]:

$$\text{ZCR} = \#\{t \mid x(t)x(t + 1) < 0\}/(T_{\text{win}} - 1).$$

Assuming zero mean of the signal, ZCR is expected to be close to 0.5 for noisy data like pure fricatives (unvoiced sounds) or noisy breaks, and close to zero for voiced speech (Fig. 1c). Therefore, it is recommended for detection of voiced segments.

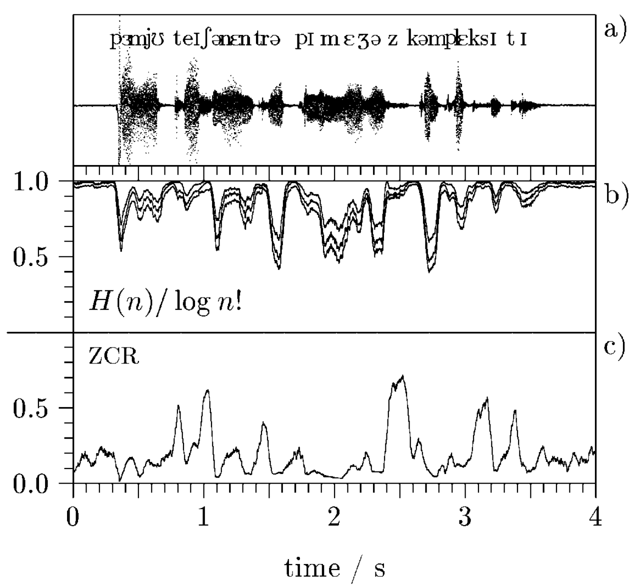

FIG. 1. Sliding window complexity analysis of the utterance "permutation entropy measures complexity." (a) Speech signal, (b) permutation entropies of order $n = 2, 3, 4$ (from top to bottom), (c) zero-crossing rate.

In Fig. 1, at first glance, ZCR and permutation entropy (PE) behave similarly. However, at some places PE seems to be a better indicator. During silence (e.g., $0,\ldots,0.3$ s) ZCR misleading indicates low complexity due to a tiny 100 Hz component (first harmonic of 50 Hz mains hum, not visible in the figure). PE is more robust against such perturbations. Moreover, PE sometimes better indicates transitions, e.g., at 1.6 s, from voiced segment to a break, or at 2.9 s where the two first syllables of "complexity" are not properly separated by ZCR. The "u" in "permutation" is also better detected by PE.

Various experiments have shown that PE, in contrast to ZCR, is robust with respect to the choice of window length (even 128 samples were sufficient), to the sampling frequency ($3,\ldots,11$ kHz gives very similar results) and to observational noise. The order $n = 3,\ldots,7$ of PE also had little influence on the results. The computational effort is only about $2n$ times the effort for ZCR.

IV. Results for chaotic time series.—To get a feeling for our entropy, we processed not only single time series, but orbits of several parametric families of dynamical systems. For the present paper we took 5001 parameter values of the logistic map $x_{t+1} = rx_t(1 - x_t)$ with $3.5 \leq r \leq 4$ since the reader is familiar with all peculiarities of this family [13]. Some two-dimensional dynamical systems, like Hénon and dissipative standard maps, were also investigated. It always turned out that PE is very similar to the positive Lyapunov exponent. In order to compare with periodic windows, band-merging points, etc., Feigenbaum diagram and Lyapunov exponents are given in Figs. 1a and 1b. In a more theoretical paper [16], it was shown that for piecewise monotone interval maps, h_n converges to the Kolmogorov-Sinai entropy h_{K} with $n \to \infty$, but this convergence is rather slow.

Proper range for n and T: We used an extremely fast algorithm where each pair of values need to be compared only once, so time was not a problem even on a PC. For larger n, the number $n!$ of permutations which can appear in the time series causes memory restrictions. More seriously, T should be considerably larger than this number in order to estimate $H(n)$ accurately. We took $T = 10^6$ which gave accurate results for $n \leq 15$, except for values r close to 4.

Our numerical results show that the functions h_6 and h_{12} (Figs. 2c and 2d) are very similar. This justifies the use of low entropy order in applications to real-world data like Sec. III. It is also good to know that h_6 can be reliably estimated already from $T = 1000$ values (Fig. 3a).

Permutation entropy and Lyapunov exponent in Figs. 2c, 2d, and 2b clearly *have a very similar appearance over the whole chaotic regime*. Differences appearing within periodic windows will be explained now. Suppose x_t approaches a stable orbit of period m, and C is the product of derivatives of the generating map along this orbit. Then the Lyapunov exponent $\frac{1}{m}\log|C|$ is negative. Moreover, inside a window of a stable period, C runs from $+1$ through 0 to -1 ([13], cf. Figs. 2a and 2b). Now for $C > 0$ there are at most m permutations in the time series which leads to $h_n = (\log m)/(n - 1)$ for $n \geq m$. For $C < 0$ we have $2m$ permutations, and h_n increases by $1/(n - 1)$. The difference is $1/5$ in Fig. 2c and $1/11$ in Fig. 2d. Unfortunately, numerical rounding produces $2m$ permutations already for some r with $C > 0$ which leads to the impression of two lines on the left of Figs. 2c and 2d. For large n, the entropy h_n tends to zero in periodic windows. Generally, it seems that $\lambda \leq h_{12} \leq h_6$.

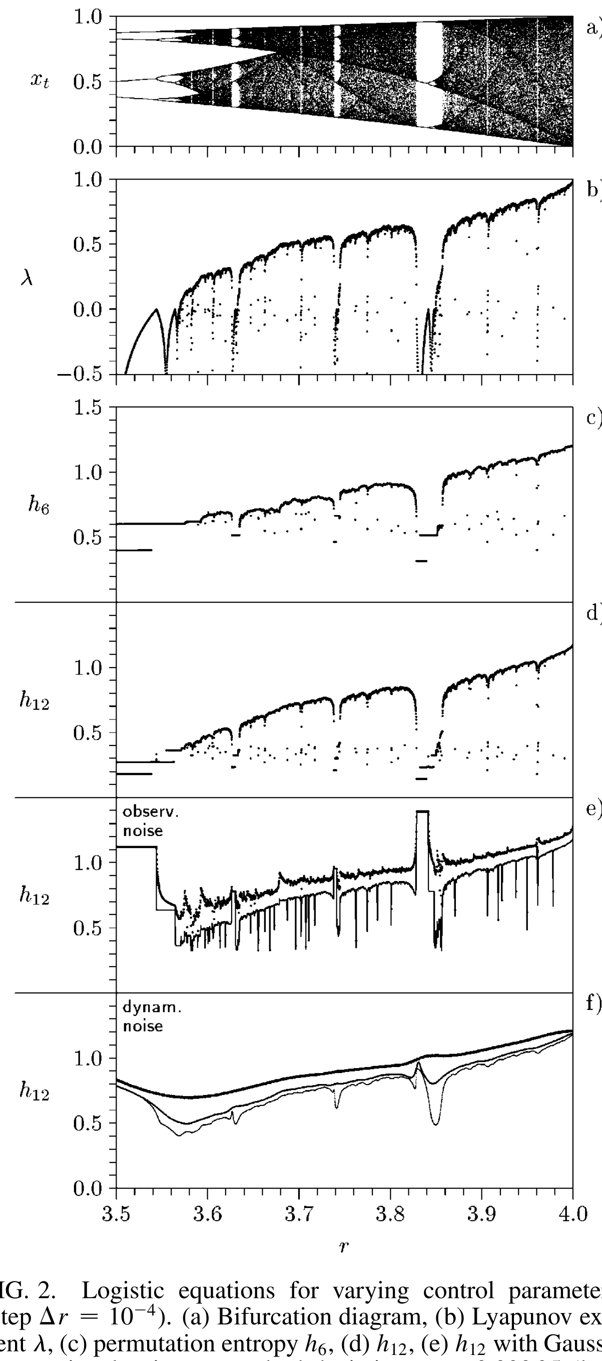

FIG. 2. Logistic equations for varying control parameter r (step $\Delta r = 10^{-4}$). (a) Bifurcation diagram, (b) Lyapunov exponent λ, (c) permutation entropy h_6, (d) h_{12}, (e) h_{12} with Gaussian observational noise at standard deviations $s = 0.000\,25$ (lower line), and $s = 0.004$ (upper line), (f) h_{12} with additive Gaussian dynamical noise at standard deviations $s = 0.000\,25$, $s = 0.001$, $s = 0.004$ (from thin to thick lines) ($T = 10^6$ data for each r value).

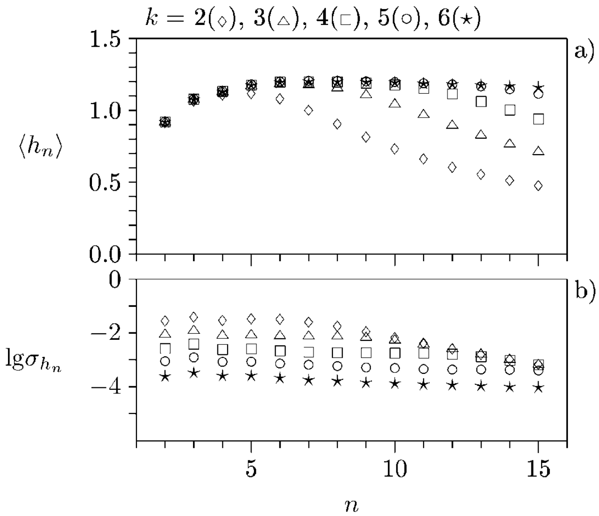

FIG. 3. Logistic map ($r = 4$). (a) Mean $\langle h_n \rangle$ of our estimator of order n permutation entropies h_n for varying length $T = 10^k$ of time series, (b) corresponding standard deviation σ of h_n.

The case $r = 4$, conjugate to the tent map [13], was studied for time series of various lengths. We took 1000 time series of size $T = 10^k$ with $k = 2, \dots, 6$. Results for $T = 10^6$ were confirmed by an exact study of space averages, giving differences less than 10^{-4} for $n \leq 10$ [16]. In Fig. 3a, we see that h_n increases up to $n = 7, 8$ and then decreases. The limit for $n \to \infty$ exists and is 1, the Kolmogorov-Sinai entropy [16]. The variance of the estimates of h_n (Fig. 3b) is rather small. There is a bias, however, when T is small compared to $n!$ which should be compensated by a correction term as in [18].

V. The effects of noise.—Permutation entropies have a practically important *invariance property.* If $y_t = f(x_t)$, where f is an arbitrary strictly increasing (or decreasing) real function, then h_n is the same for x_t and y_t. Such nonlinear functions f occur, for example, when measuring physiological data with different equipment. However, the invariance of h_n also implies its *discontinuity near the constant time series* $x_t = c$ where $h_n = 0$. If this series is disturbed by an i.i.d. noise, no matter how small, then $h_n = (\log n!)/(n - 1)$, the largest possible value.

Observational noise: For time series of period m and $n = km$ the disturbed time series admits m times $k!^m$ permutations so that $h_n = (\log m + m \cdot \log k!)/(n - 1)$ as long as the noise preserves the order of the periodic orbit. In Fig. 2e, Gaussian noise was added to the deterministic time series. Here we get for the period-4 window (on the left of Fig. 2a) $m = 4$ with $h_{12} \approx 1.12$, and for the period-3 window (in the middle of Fig. 2a) $m = 3$ with value $h_{12} \approx 1.39$. Outside the low-period windows, however, the observational noise causes only a small increase of entropy. For the higher noise level, there are some peaks exactly at band-merging points, for example, $r_c \approx 3.5748$, 3.5925, 3.6785 (compare Figs. 2a and 2e). For $r < r_c$ the noise anticipates the band-merging point by creating lots of new permutations as in the periodic windows. When the band structure disappears, the influence of noise becomes weaker and the entropy is smaller.

Dynamical noise, added to x_t during each step of the iteration, gives still better results. The entropy function $h_n(r)$ becomes smooth, approximating the entropy of the undisturbed time series for small noise level (Fig. 2f). Actually, Fig. 3a does not change at all for $s \leq 0.004$.

Near the period 3 window in Fig. 2f there are examples of noise-induced order where larger noise gives smaller h_n—but in general h_n increases with noise level. The effects of low periods disappear for larger noise level where the increase of entropy with growing r becomes apparent.

VI. Conclusion.—Permutation entropy is an appropriate complexity measure for chaotic time series, in particular in the presence of dynamical and observational noise. In contrast with all known complexity parameters, a small noise does not essentially change the complexity of a chaotic signal. Permutation entropies can be calculated for arbitrary real-world time series. Since the method is extremely fast and robust, it seems preferable when there are huge data sets and no time for preprocessing and fine-tuning of parameters.

[1] *Nonlinear Analysis of Physiological Data,* edited by H. Kantz *et al.* (Springer, Berlin, 1996); *Chaos in Brain?,* edited by K. Lehnertz *et al.* (World Scientific, Singapore, 1999).

[2] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[3] Ya. B. Pesin, *Dimension Theory in Dynamical Systems* (University of Chicago Press, Chicago, 1998).

[4] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).

[5] M. Ding, C. Grebogi, E. Ott, T. Sauer, and J. A. Yorke, Phys. Rev. Lett. **70**, 3872 (1993).

[6] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1996).

[7] G. Bofetta, M. Cencini, M. Falcioni, and A. Vulpiani (to be published).

[8] A. A. Brudno, Trans. Moscow Math. Soc. **1983**, No. 2, 127 (1983).

[9] S. Galatolo, Discr. Cont. Dyn. Systems **7**, 477 (2001).

[10] L. L. Campbell, Z. Wahrsch. verw. Geb. **6**, 113 (1966).

[11] K. Petersen, *Ergodic Theory* (Cambridge University Press, Cambridge, U.K., 1983).

[12] P. Grassberger and H. Kantz, Phys. Lett. **113A**, 235 (1985).

[13] P. Collet and J.-P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhaeuser, Basel, 1980).

[14] E. M. Bollt, T. Stanford, Y.-C. Lai, and K. Zyczowski, Phys. Rev. Lett. **85**, 3524 (2000).

[15] C. Bandt and B. Pompe, J. Stat. Phys. **70**, 967 (1993).

[16] C. Bandt, G. Keller, and B. Pompe, www.math-inf.uni-greifswald.de/~bandt/pub.html.

[17] L. R. Rabiner and R. W. Schaefer, *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ, 1978).

[18] P. Grassberger, Phys. Lett. A **128**, 369 (1988).