Chaos

Approximate entropy (ApEn) as a complexity measure

Cite as: Chaos **5**, 110 (1995); https://doi.org/10.1063/1.166092
Submitted: 06 May 1994 . Accepted: 02 August 1994 . Published Online: 14 August 1998

Steve Pincus


View Online

Export Citation

ARTICLES YOU MAY BE INTERESTED IN

Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series
Chaos: An Interdisciplinary Journal of Nonlinear Science **5**, 82 (1995); https://doi.org/10.1063/1.166141

Quantitative analysis of heart rate variability
Chaos: An Interdisciplinary Journal of Nonlinear Science **5**, 88 (1995); https://doi.org/10.1063/1.166090

Nonlinear time-series analysis revisited
Chaos: An Interdisciplinary Journal of Nonlinear Science **25**, 097610 (2015); https://doi.org/10.1063/1.4917289

AIP Publishing

© 1995 American Institute of Physics.

Approximate entropy (ApEn) as a complexity measure

Steve Pincus
990 Moose Hill Road, Guilford, Connecticut 06437

(Received 6 May 1994; accepted for publication 2 August 1994)

Approximate entropy (ApEn) is a recently developed statistic quantifying regularity and complexity, which appears to have potential application to a wide variety of relatively short (greater than 100 points) and noisy time-series data. The development of ApEn was motivated by data length constraints commonly encountered, e.g., in heart rate, EEG, and endocrine hormone secretion data sets. We describe ApEn implementation and interpretation, indicating its utility to distinguish correlated stochastic processes, and composite deterministic/ stochastic models. We discuss the key technical idea that motivates ApEn, that one need not fully reconstruct an attractor to discriminate in a statistically valid manner—marginal probability distributions often suffice for this purpose. Finally, we discuss why algorithms to compute, e.g., correlation dimension and the Kolmogorov–Sinai (KS) entropy, often work well for true dynamical systems, yet sometimes operationally confound for general models, with the aid of visual representations of reconstructed dynamics for two contrasting processes. © *1995 American Institute of Physics.*

I. INTRODUCTION

Approximate entropy (ApEn), has been recently introduced as a quantification of regularity in time-series data, motivated by applications to relatively short, noisy data sets.[1] Mathematically, ApEn is part of a general development as the *rate of entropy* for an approximating Markov chain to a process.[2] In applications to heart rate (e.g., Fig. 1), findings have discriminated groups of subjects via ApEn, in instances where classical [mean, standard deviation (SD)] statistics did not show clear group distinctions.[3–7] In applications to endocrine hormone secretion data (e.g., Fig. 2) based on as few as $N=72$ points, ApEn has provided vivid distinctions ($P<10^{-8}$) between actively diseased subjects and normals, with nearly 100% specificity and sensitivity.[8] The point of this article is to discuss ApEn, in part focusing on the technical point that motivates its definition—marginal probabilities often suffice to discriminate. We contrast reconstructed dynamics for two processes to understand why algorithms to compute correlation dimension[9] and the Kolmogorov–Sinai (KS) entropy[10] often work well for true dynamical systems, yet sometimes operationally confound for general models. This contrast indicates the need for a thematically faithful modification of a parameter such as the KS entropy for general applications so that visual intuition matches numerical results, for broad classes of stochastic processes as well as for dynamical systems.

To illustrate the distinctions we are trying to quantify, contrast Fig. 1(a) vs 1(b), Fig. 2(a) vs 2(b), and Fig. 3, taken from the MIX(p) process discussed below, for which we see time series that apparently become increasingly irregular as we proceed from (a) to (c). Historical context frames this effort. Complexity statistics developed for application to chaotic systems and relatively limited in scope recently have been commonly applied to finite, noisy and/or stochastically derived time-series, frequently with confounding and nonreplicable results. This caveat is particularly germane to biologic signals, especially those taken *in vivo,* as such signals likely represent the output of a complicated network with both stochastic and deterministic components.

II. QUANTIFICATION OF REGULARITY

Definition of ApEn: Two input parameters, m and r, must be fixed to compute ApEn—m is the "length" of compared runs, and r is effectively a filter. Given N data points $\{u(i)\}$, form vector sequences $x(1)$ through $x(N-m+1)$, defined by $x(i)=[u(i),\ldots,u(i+m-1)]$. These vectors represent m consecutive u values, commencing with the ith point. Define the distance $d[x(i),x(j)]$ between vectors $x(i)$ and $x(j)$ as the maximum difference in their respective scalar components. Use the sequence $x(1),x(2),\ldots,x(N-m+1)$ to construct, for each $i\leqslant N-m+1$, $C_i^m(r)=$(number of $j\leqslant N-m+1$ such that $d[x(i),x(j)]\leqslant r)/(N-m+1)$. The $C_i^m(r)$'s measure *within a tolerance* r the regularity, or frequency, of patterns similar to a given pattern *of window length* m. Define $\Phi^m(r)=(N-m+1)^{-1}\sum_{i=1}^{N-m+1}\ln C_i^m(r)$, where ln is the natural logarithm, then define the *parameter* $\mathrm{ApEn}(m,r)=\lim_{N\to\infty}[\Phi^m(r)-\Phi^{m+1}(r)]$.

Given N data points, we estimate this parameter by defining the *statistic* $\mathrm{ApEn}(m,r,N)=\Phi^m(r)-\Phi^{m+1}(r)$. ApEn measures the likelihood that runs of patterns that are close for m observations remain close on next incremental comparisons. Greater likelihood of remaining close, regularity, produces smaller ApEn values, and conversely. Upon unraveling definitions we deduce

$$\begin{aligned}-\mathrm{ApEn}&=\Phi^{m+1}(r)-\Phi^m(r)\\&=\text{average over } i \text{ of ln [conditional probability}\\&\quad\text{that } |u(j+m)-u(i+m)|\\&\leqslant r, \text{ given that } |u(j+k)-u(i+k)|\leqslant r \text{ for } k\\&=0,1,\ldots,m-1].\end{aligned}\tag{1}$$

To develop a more intuitive, physiological understanding of this definition, a multistep description of the algorithm with figures is developed in Pincus and Goldberger.[11]

 1054-1500/95/5(1)/110/8/$6.00 © 1995 American Institute of Physics

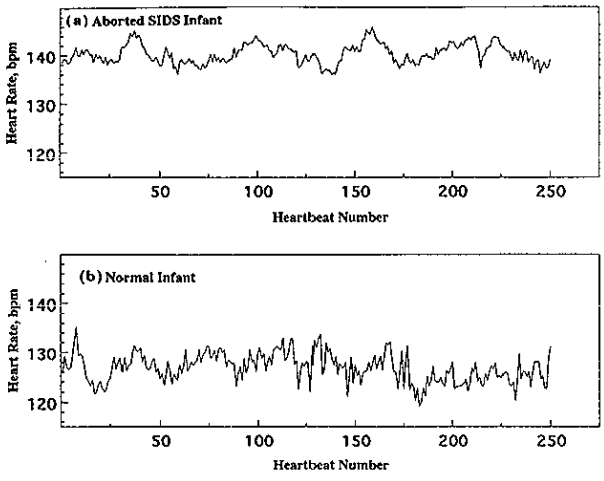

FIG. 1. Two infant quiet sleep heart rate tracings with similar variability, standard deviation (SD): (a) aborted SIDS infant, SD=2.49 beats per minute (bpm), ApEn(2, 0.15 SD, 1000)=0.826; (b) normal infant, SD=2.61 bpm, ApEn(2, 0.15 SD, 1000)=1.463. Tracing (a) appears to be more regular than tracing (b), confirmed by ApEn.

III. IMPLEMENTATION AND INTERPRETATION

The value of N, the number of input data points for ApEn computations is typically between 75 and 5000. Based on calculations that included both theoretical analysis[1,12–14] and clinical applications[3–8] we have concluded that for $m=1$ and 2, values of r between 0.1 to 0.25 SD of the $u(i)$ data produce good statistical validity of ApEn(m,r,N). For such r values, we demonstrated the theoretical utility of ApEn$(2,r,N)$ to distinguish data on the basis of regularity for both deterministic and random processes, and the clinical utility in the aforementioned applications.

ApEn is typically calculated via a short computer code (see Ref. 6, Appendix B for a FORTRAN listing). The form of ApEn provides for both *de facto* noise filtering, via choice of

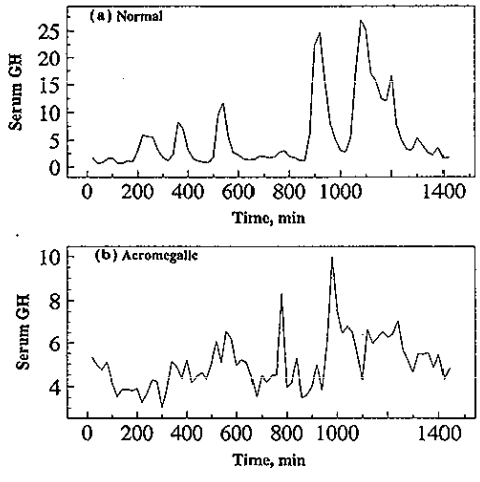

FIG. 2. Growth hormone (GH) serum concentrations, in milliunits/ml, measured at 20 minute intervals for 24 h ("fed" state) (a) normal subject, mean concentration =5.617, ApEn(1, 0.20 SD, 72)=0.783; (b) acromegalic subject, mean concentration =4.981, ApEn(1, 0.20 SD, 72)=1.420.

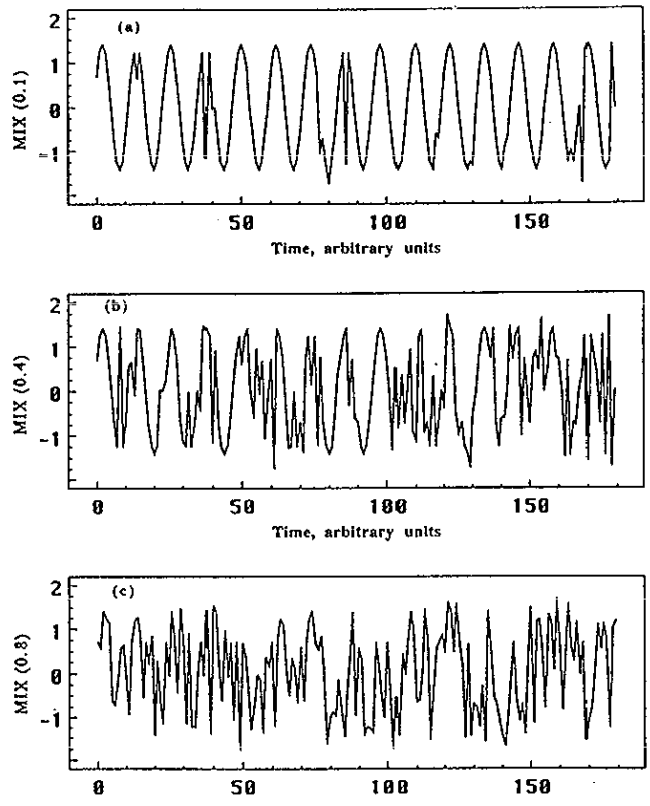

FIG. 3. MIX(p) model time-series output for three parameter values: (a) $p=0.1$; (b) $p=0.4$; (c) $p=0.8$. MIX(p) is a family of processes that samples a sine wave for $p=0$, samples independent uniform random variables for $p=1$, and intuitively becomes more irregular as p increases. ApEn quantifies the increasing irregularity with increasing p: for $m=2$, $r=0.18$, and $N=1000$, ApEn[MIX(0.1)]=0.436, ApEn [MIX(0.4)]=1.455, and ApEn[MIX(0.8)]=1.801.

"r,"[1] and artifact insensitivity,[6] useful statistical properties for practical applications. Despite algorithmic similarities, it is important to note that ApEn(m,r,N) is not intended as an approximate value of KS entropy.[1,12] It is imperative to consider ApEn(m,r,N) as a *family* of statistics; for a given application, system comparisons are intended with fixed m and r. For a given system, there usually is significant variation in ApEn(m,r,N) over the range of m and r.[5,6,12]

A significant aspect of ApEn utility is that numerical calculation generally concurs with pictorial intuition, both for theoretical and clinically derived data. ApEn confirms differences that are visually "obviously distinct," as in the comparison of (a) vs (c) in Fig. 3, and provides information in subtler comparisons, distinguishing (b) from (c) in Fig. 3. Noted below, theoretical calculations for correlation dimension and KS entropy produce confounding results for the MIX(p) model of Fig. 3, and often for general stochastic and composite models. Clinically, illustrated in Fig. 3 of a fetal heart rate study,[7] abnormally low ApEn values not only indicate the visually apparent, regular heart rate tracing of an acidotic, distressed fetus measured 1 hr before delivery, but also the less visually apparent, modestly regular tracing mea-

sured early in labor. In the endocrine hormone secretion study,[8] ApEn discerned subjects in remission as intermediate between normal and actively diseased subjects ($P<0.001$). Both these clinical results suggest the potential of a regularity statistic to not only confirm obvious findings, but to provide new diagnostic, possibly clinically predictive capability.

To enact direct comparisons of ApEn results, it is best to fix not only m and r, as indicated above, but also N (data length) and choice of coordinates (scale), as would typically be done in a set protocol. The requirement to fix N is due to the statistical bias of ApEn (similar bias occurs for other complexity statistics); if one assumes a model form, corrections could typically be made to compare ApEn results for differing data lengths. ApEn is not invariant under coordinate transformations, hence scale must be fixed; as noted below, such noninvariance is also common to the differential entropy. The choice of "r" in ApEn ensures that it will be impervious to physical device perturbations and recalibrations, important properties for laboratory applications.

The physiologic *modeling* of many complex biological systems is often very difficult; one would expect accurate models of such systems to be complicated composites, with both deterministic and stochastic components, and interconnecting network features. ApEn is a broadly applicable *statistic* in that it can distinguish many classes of systems, and can be meaningfully applied to $N>100$ data points. Via extensive Monte Carlo calculations we established that the standard deviation (SD) of ApEn(2, 0.15 SD, 1000)<0.055 for a large class of candidate models.[11–13] It is this small SD ("error bars") of ApEn, applied to 1000 points from various models, that provides its utility to practical data analysis of moderate length time series.

It seems important to determine a unifying theme suggesting greater signal regularity in diverse complicated systems. Our mechanistic hypothesis is that in a variety of systems; greater regularity corresponds to greater component and subsystem autonomy. This hypothesis has been mathematically established via analysis of several very different, representational model forms, conferring a robustness to the model form of the hypothesis.[13,15]

IV. RELATIVE CONSISTENCY—A VARIATIONAL PRINCIPLE

For many processes, ApEn(m,r,N) grows with decreasing r like $\log(2r)$, thus exhibiting infinite variation with r.[12] We typically observe that for a given time series, ApEn(2, 0.1) is quite different from ApEn(4, 0.01), so the question arises of which parameter choices (m and r) to use. The guidelines above address this, but the most important requirement is consistency. For deterministic dynamical systems, we typically see that when KS entropy$(A)\leqslant$KS entropy(B), then ApEn$(m,r)(A)\leqslant$ApEn$(m,r)(B)$ and conversely, for a wide range of m and r. Furthermore, for both theoretically described systems[1,13] and those described by experimental data,[5,6] we have found that when ApEn$(m_1,r_1)(A)\leqslant$ApEn$(m_1,r_1)(B)$, then ApEn$(m_2r_2)(A)\leqslant$ApEn$(m_2,r_2)(B)$, and conversely. This latter property also generally holds for parametrized systems of stochastic processes, in which KS entropy is infinite. We call this ability of ApEn to preserve order a relative property. It is key to the general and clinical utility of ApEn.

From a theoretical perspective, the interplay between meshes [(m,r) pair specifications] need not be nice, in general, in ascertaining which of (two) processes is "more" random. In general, we might like to ask: Given no noise and an infinite amount of data, can we say that process A is more regular than process B? The *flip-flop pair* of processes implies that the answer to this question is "not necessarily": in general, comparison of relative process randomness *at a prescribed level* is the best one can do.[12] Process A may appear more random than process B on many choices of partitions, but not necessarily on all partitions of suitably small diameter (r).

Fortunately, for many processes A and B, we can assert more than relative regularity, even though both A and B will typically have infinite KS entropy. For such pairs of processes, denoted a *completely consistent pair*, whenever ApEn$(m,r)(A)<$ApEn$(m,r)(B)$ for any specific choice of m and r, then it follows that ApEn$(n,s)(A)<$ApEn$(n,s)(B)$ for *all* choices of n and s.[12] Visually, process B appears more random than A at any level of view. We indicate elsewhere a conjecture that should be straightforward to prove, providing a sufficient condition to ensure that A and B are a completely consistent pair, and indicating the relationship of the autocorrelation function.[11]

V. RELATIONSHIP TO OTHER APPROACHES

The development of mathematics to quantify regularity has centered around various entropy measures. However, there are numerous entropy formulations, and many entropy definitions can not be related to one another.[1] KS entropy, developed by Kolmogorov and expanded upon by Sinai, classifies *deterministic* dynamical systems by rates of information generation.[10] It is this form of entropy that Grassberger and Procaccia,[16] Eckmann and Ruelle,[17] and Takens[18] estimate.

However, the KS entropy was not developed for statistical applications, and has major debits in this regard. The original, primary motivation for this entropy was to determine when two Bernoulli shifts are isomorphic. In its proper context, this form of entropy is primarily applied by ergodic theorists to well-defined theoretical transformations, for which no noise and an infinite amount of "data" are standard mathematical assumptions. Ornstein proved the important, deep result, that two dynamical systems are isomorphic if and only if they have identical KS entropy.[19] Also, for dynamical systems, positive entropy implies chaos.[17] But attempts to utilize KS entropy for practical data analysis represent out-of-context application, which often generates serious difficulties, as it does here. KS entropy is badly compromised by steady, (even very) small amounts of noise, and it generally requires a vast amount of input data to achieve convergence,[14,20] and is usually infinite for stochastic processes. All these debits are key in the present context, since most biological time series likely comprise both stochastic and deterministic components.

ApEn was constructed along thematically similar lines to the KS entropy, though with a different focus: to provide a widely applicable, statistically valid formula to distinguish data sets.[1,6] The technical point motivating ApEn is that if joint probability measures for reconstructed dynamics describing each of two systems are different, then their marginal probability distributions on a fixed partition, given by conditional probabilities as in Eq. (1), are likely different.

There exists a large literature on reconstructed dynamics for chaotic systems. Correlation dimension,[9] KS entropy, and the Lyapunov spectrum have been much studied, as have techniques to utilize related algorithms in the presence of noise and limited data.[21–23] Even more recently, prediction techniques have been developed for chaotic systems.[24–26] Most of these methods employ embedding dimensions larger than $m=2$, as is typically employed with ApEn. Thus in the purely *deterministic dynamical system* setting, they are more powerful than ApEn in that they reconstruct the probability structure of the space with greater detail. However, in the general stochastic process setting, the statistical accuracy of the aforementioned parameters and methods appears to be poor, and the prediction techniques are not always defined.

Generally, changes in ApEn agree with changes in dimension and entropy algorithms for low-dimensional, deterministic systems. The essential points here, assuring broad utility, are that (i) ApEn can potentially distinguish a wide variety of systems: low-dimensional deterministic systems, periodic and multiply periodic systems, high-dimensional chaotic systems, stochastic and mixed (stochastic and deterministic) systems, and (ii) ApEn is applicable to noisy, medium-sized data sets.[1,13] Thus ApEn can be applied to settings for which the KS entropy and correlation dimension are either undefined or infinite, with good replicability properties as discussed below.

VI. MARGINAL PROBABILITIES AND FULL RECONSTRUCTION

As noted above, full attractor (invariant measure) reconstruction is often unnecessary to distinguish processes. A primary data question is: Are data $\{X_i\}$ "atypical" (abnormal)? This is a discrimination question; the question of accurately modeling the process underlying $\{X_i\}$ is often much harder, but one that we may well be able to avoid. We think of marginal probabilities as *partial* process characterization, given by the small m and relatively coarse r in ApEn(m,r). The rationale is that we typically need orders of magnitude fewer points to accurately estimate these marginal probabilities than to accurately reconstruct the "attractor" measure defining the process. We now indicate what marginal probabilities are, and how they arise in time-series reconstruction.

A *joint probability measure* is a means of assigning probability to a region of space. Consider, e.g., all points (x,y) such that $0\leq x,y\leq 1$ (unit square). Define the joint measure for the density $f(x,y)$: the probability of (a subset of the square) $\mathbf{A}=\int\int_{\mathbf{A}} f(x,y)dx\,dy$. Then define the *marginal density* of x, $f_X(x)=\int f(x,y)dy$ (analogously for y). An intuitive interpretation of marginal density is easiest in the finite state setting—given the joint distribution $p(x,y)$ of two variables x and y, define $p_X(x):=\Sigma_y p(x,y)$ [aggregation of $p(x,y)$ over all values of y]. It is essential to note that marginal probabilities of two processes may be equal while the joint probabilities are quite different. *However, if two measures have distinctly different marginal probabilities, then that alone is sufficient to discriminate the measures.*

Define the *conditional probability* of $X\in A$ given (‖) $Y\in B$ as the joint probability that $X\in A$ and $Y\in B$ divided by the marginal probability that $Y\in B$; i.e., $\int_{\mathbf{B}}\int_{\mathbf{A}} f(x,y)dx\,dy/\int_{\mathbf{B}} f_Y(y)dy$. Marginal and conditional probability definitions extend to general numbers of variables. For application to time series, we study the joint distribution for n-contiguous observations $\{X_m,\ldots,X_{m+n-1}\}$ as the n-variable measure. Complete steady-state description is given by expressions of the form conditional probability $[u(j+m)\in A_m\|u(j+1)\in A_1,\ldots,u(j+m-1)\in A_{m-1}]$, for all integers m and sets A_i. The question of full reconstruction (embedding dimension n) is subsumed by a broader question, is the process nth-order Markov? The general answer is not necessarily; there may be no fixed n for which the process is characterized entirely by conditioning on n previous observations. However, for fixed m, these *marginal conditional probabilities* contain a wealth of probabilistic detail about the underlying process, often allowing discrimination. Such probabilities form the building blocks of the KS entropy, correlation dimension, and ApEn, by the probabilities

$$\{|u(j+m)-u(i+m)|\leq r\|\,|u(j+k)-u(i+k)|\leq r \text{ for } k=0,1,\ldots,m-1\}. \quad (2)$$

In the KS entropy, Lyapunov spectra, and correlation dimension, $m\to\infty$ (or to full embedding) and $r\to 0$, whereas ApEn stops short via small m and coarse r, sacrificing an attempt to reconstruct the full process dynamics. The tradeoffs between the approaches to time-series reconstruction are statistical. Obviously bigger m and smaller r describe sharper parameter (probabilistic) detail. However, for each template vector $\{u(i),\ u(i+1),\ldots,u(i+m)\}$, we estimate Eq. (2) by A/B, where $A=$number of j such that $|u(j+k)-u(i+k)|\leq r$ for $k=0,1,\ldots,$ and m, $B=$number of j such that $|u(j+k)-u(i+k)|\leq r$ for $k=0,1,\ldots,$ and $m-1$. If either m is relatively large or r is too small, then both A and (importantly) B are small numbers, thus the estimate of Eq. (2) by A/B is unreliable.

ApEn is related to a parameter in information theory, *conditional entropy*.[27] Assume a finite state space, where the entropy of a random variable X, $\text{Prob}(X=a_j)=p_j$, is $H(X):=-\Sigma p_j\log p_j$, and the entropy of a block of random variables $X_1,\ldots,X_n=H(X_1,\ldots,X_n):=-\Sigma\Sigma\cdots\Sigma p^n(a_{j1},\ldots,a_{jn})\times\log p^n(a_{j1},\ldots,a_{jn})$. For two variables, the conditional entropy $H(Y\|X)=H(X,Y)-H(X)$; this extends naturally to n variables. Closely mimicking the proof of Theorem 3, Ref. 1, the following theorem is immediate: for $r<\min_{j\neq k}|a_j-a_k|$, $\text{ApEn}(m,r)=H(X_{m+1}\|X_1,\ldots,X_m)$; thus in this setting, ApEn is a conditional entropy. Observe that we do not assume that the process is mth order Markov, i.e., that we fully describe the process; we aggregate the mth-order marginal probabilities. The rate of entropy $=\lim_{n\to\infty}H(X_n\|X_1,\ldots,X_{n-1})$ is the discrete state analog of the KS entropy. However, we cannot go from discrete to

continuous state naturally as a limit; most calculations give ∞. As for differential entropy, there is no fundamental physical interpretation of conditional entropy (and no invariance, Ref. 27, p. 243) in continuous state.

VII. RECONSTRUCTED DYNAMICS—A COMPARISON OF TWO AUTOCORRELATED MAPS

General Issues: Analysis of the MIX(p) processes indicates some of the *theoretical* difficulties realized in applying correlation dimension and KS entropy statistics to general time series.[1] To define MIX(p), first fix $0 \leq p \leq 1$. For all j, define $X_j = \sqrt{2}\sin(2\pi j/12)$, Y_j=i.i.d. uniform random variables on $[-\sqrt{3}, \sqrt{3}]$, and Z_j=i.i.d. random variables, $Z_j = 1$ with probability p, $Z_j = 0$ with probability $1-p$. Then define MIX $(p)_j = (1-Z_j)X_j + Z_j Y_j$. In conjunction with intuition, ApEn monotonically increases with increasing p (Fig. 3). In contrast, correlation dimension of MIX(p) $=0$ for $p<1$, and correlation dimension of MIX(1) $=\infty$.[1,28] As well, KS entropy of MIX(p) $=\infty$ for $p>0$, and $=0$ for MIX(0). Thus both of these statistics fail to quantify the evolving complexity change. Similar "confounding" results can be established for Gaussian processes, ARMA models, and more generally for weak-dependence processes, via proofs similar to that given in Ref. 1.

Nonetheless, one might still consider applying an algorithm to compute, e.g., the correlation dimension or KS entropy, and should then query what the *operational* consequences are of application to a general correlated stochastic process. To understand the statistical limitations imposed by the choice of embedding dimension m and scaling parameter r, we can think of an (m,r) choice of input parameters as partitioning the state space into uniform width boxes of width r, from which we estimate mth-order conditional probabilities. For state space, e.g., $[-A/2, A/2]$, we would have $(A/r)^{m+1}$ conditional probabilities to estimate. Specifically, divide $[-A/2, A/2]$ into A/r cells; the ith cell $C(i)$ is the half-open interval $[x, x+r)$, where $x = -(A/2) + (i-1)r$. Then define the conditional probability $p_{i\text{vect},j}$ for all length m vectors of integers $i$vect and integers j, $i\text{vect} = (i_1, i_2, \ldots, i_m)$, $1 \leq i_k \leq A/r$ for all k, $1 \leq j \leq A/r$, by $p_{i\text{vect},j}$ = {conditional probability that $u(k) \in C(j)$, given that $u(k-1) \in C(i_1)$, $u(k-2) \in C(i_2)$, ..., and $u(k-m) \in C(i_m)$}. In the very general ergodic case, these conditional probabilities are given by limits of time averages.

For general stochastic processes, many of these conditional probabilities are nonzero, so we need to accommodate reasonable estimates of the $(A/r)^{m+1}$ conditional probabilities given N data points. If m is relatively large, or if r is too small, the number of probabilities to be estimated will be unwieldy, and statistical estimates will be poor for typical data set lengths.

Probabilistic Distinction: We can now operationally see the issues in analyzing reconstructed dynamics for general stochastic processes. Compare two processes, logis(3.6), the logistic map $f(x) = 3.6x(1-x)$ and MIX(0.4) [Figs. 4(a) and 4(b)], with embedding dimension $m=2$ and a moderately coarse mesh width r that subdivides the state spaces into 20 equal-width cells [thus $r = \sqrt{3}/10$ for MIX(0.4), while $r = 0.05$ for logis(3.6)]. We want to estimate the conditional probabilities $p_{i\text{vect},j}$ for all length 2 vectors $i\text{vect} = (i_1, i_2)$ and integers $1 \leq j \leq 20$ for each process. For logis(3.6), the ith cell $C(i) = [0.05(i-1),\ 0.05(i-1)+0.05)$. We calculate these probabilities as $\lim_{N\to\infty} A/B$ (estimated from $N = 2\,000\,000$ points), with A = number of $k \leq N$ when $u(k) \in C(j)$ and $u(k-1) \in C(i_1)$ and $u(k-2) \in C(i_2)$; B = number of $k \leq N$ when $u(k-1) \in C(i_1)$ and $u(k-2) \in C(i_2)$. In Figs. 4(c) and 4(d), we visualize these $\{p_{i\text{vect},j}\}$ by associating the triple $\{i\text{vect},j\} = \{i_1, i_2, j\}$ with the three-dimensional subcube $[C(i_2) \times C(i_1) \times C(j)]$. We shade each subcube to illustrate its conditional probability $p_{i\text{vect},j}$. The extremely different character in reconstructed dynamics between these two maps is visually apparent—logis(3.6) is extremely sparse, with only 27 nonzero subcubes of a possible 8000, whereas MIX(0.4) displays "diffuse" dynamics, with all 8000 subcubes nonzero. Furthermore, in general the conditional probabilities are not rare (nearly 0) for most nonzero cells for logis(3.6), compared to MIX(0.4), an important added distinction when considering the effects of the bias indicated below. This sparseness marks the *strong dependence*[29] of dynamical systems, and it should not be surprising that algorithms that work well for such strong-dependence processes behave entirely different for typically encountered weak-dependence (e.g., Gaussian) processes.

Primary Statistical Issues: We contrast the probabilistic distinction between logis(3.6) and MIX(0.4) with its statistical consequences, considering a "clinically sized" data set length $N = 1000$. Figs. 4(e) and 4(f) indicate the shaded subcubes of Figs. 4(c) and 4(d) for which statistical estimates based on 1000 points might be adequate. Defining A and B as in the previous paragraph, we only display subcubes for which $B > 10$; otherwise, A/B is a poor estimate of the true conditional probability as given as $N \to \infty$. The percentage of nonzero probability cells from Figs. 4(c) and 4(d) displayed in Figs. 4(e) and 4(f) thus provide a measure of how reasonably estimated the conditional probabilities (and associated statistics) are, based on $N = 1000$. As an additional indicator, in Figs. 4(e) and 4(f) darker shading corresponds to larger B, hence better estimation of the true conditional probability of the cell, given by more conditioning vectors. Note that the conditional probabilities for logis(3.6) are reasonably estimated—25/27 (93%) of the nonzero subcubes seen in Fig. 4(c) satisfy $B > 10$. In contrast, the conditional probabilities for MIX(0.4) are poorly estimated—260/8000 (3%) of nonzero subcubes satisfy $B > 10$. Furthermore, generally there are greater numbers of conditioning vectors (darker cells) for logis(3.6) than for MIX(0.4), reinforcing the superior estimation based on $N = 1000$. For example, for logis(3.6) we find that there are 108 values of $k \leq 1000$ for which pairs of contiguous points $(u(k-1), u(k-2))$ satisfy $\{u(k-1) \in C(7)$ and $u(k-2) \in C(18)\}$. Of these 108 conditioning vectors, 61 3-tuples of points $(u(k-1), u(k-1), u(k-2))$ satisfy $\{u(k) \in C(17)$, $u(k-1) \in C(7)$, and $u(k-2) \in C(18)\}$; we then estimate $p_{(7,18),17} = 61/108$. A second estimate (different initial condition) based on a subsequent 1000 point sequence would likely produce a similar estimate for $p_{(7,18),17}$. It is precisely the increasing and ex-

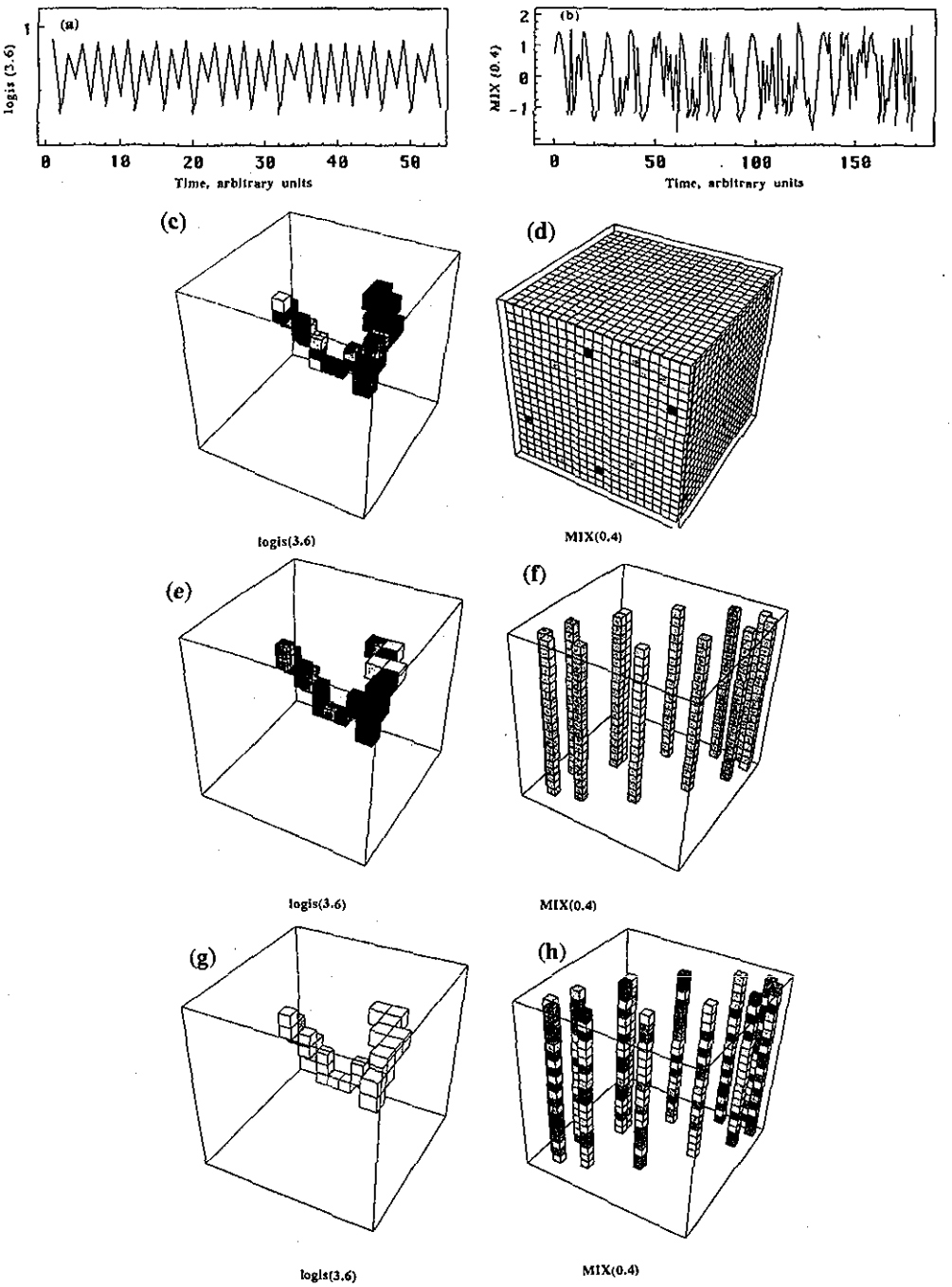

FIG. 4. Comparison of logis(3.6), the logistic map $f(x)=3.6x(1-x)$, and MIX(0.4) for embedding dimension $m=2$, mesh width $r=0.05x$ support of state space. (a), (b) Time-series realizations; (c), (d) probabilistic contrast. Visualize conditional probabilities $\{p_{ivect,j}\}$ for all length 2 vectors $ivect=(i_1,i_2)$, integers $1\leq j\leq 20$ by associating triple $\{ivect,j\}=\{i_1,i_2,j\}$ with three-dimensional subcube $[C(i_2)\times C(i_1)\times C(j)]$, shading each subcube per weight $p_{ivect,j}$, with black indicating a conditional probability of 1, white a conditional probability of 0. Contrast the extreme sparseness of logis(3.6), with 27 nonzero subcubes, and the "diffuse" dynamics of MIX(0.4), with all 8000 subcubes nonzero. (e), (f) Statistical contrast, $N=1000$. Shading by subcube of adequacy of conditional probability estimates. We only display subcubes for which $B>10$; otherwise, A/B is a poor estimate of the true conditional probability, with darker shading corresponding to larger B (and better estimation, with more conditioning vectors). Conditional probabilities for logis(3.6) are well estimated—25/27 (93%) of nonzero subcubes in (c) satisfy $B>10$; conditional probabilities for MIX(0.4) are poorly estimated—260/8000 (3%) of nonzero subcubes satisfy $B>10$. Additionally, logis(3.6) cells are generally darker (better estimated), e.g., for logis(3.6), estimate $p_{(7,18),17}=61/108$, while for MIX(0.4), estimate $p_{(6,11),13}=1/27$. (g), (h) Bias effects contrast. We compare percentage error of log(conditional probability estimate, $N=1000$) vs log(true conditional probability), for each subcube with $B>10$, with darker shading indicating larger error. Average percentage error/subcube, given $B>10$ is 9.5% for logis(3.6), 40.5% for MIX(0.4). Virtually all "A's" for MIX(0.4) in estimates A/B were 0, 1, or 2.

treme sparseness of the set of $m+1$-tuples for which the true conditional probabilities $p_{ivect,j}$ are nonzero *for true dynamical systems,* as a function of embedding dimension m and mesh width (scaling range) r, that affords large m and small r in conditional probability statistical estimation for these processes.

Bias: There is a secondary, but oftentimes very important statistical issue in estimation of conditional probabilities

such as those given by Eq. (2) by a ratio of A/B as above. The template vector $x(i)$ itself counts in the aggregation of vectors close to $x(i)$ in ApEn to ensure that calculations involving logarithms (which arise, e.g., in "correlation integrals") remain finite. This procedure ensures that ApEn is a *biased* statistic (though asymptotically unbiased for many processes), as are correlation dimension and KS entropy algorithms for similar reasons.[12,30] Operationally, the inclusion of the template vector changes a conditional probability estimate from A/B to $(A+1)/(B+1)$, adding 1 to both numerator and denominator counts.

This point is trivial, effectively irrelevant in calculations for dynamical systems—both A, B are typically much larger than 1, so that $A/B \approx (A+1)/(B+1)$. However, we aggregate expressions of the form $\log(A/B)$ in the algorithms under consideration, with $\log(A/B)$ a large negative number if A is small. In particular, $\log(A/B)$ is typically much different than $\log[(A+1)/(B+1)]$ for A small. For "diffuse" reconstructed dynamics settings [i.e., weak-dependence processes such as MIX(0.4)] with moderate data length N, frequently a nontrivial percentage of conditional probability estimates are of rare events, with $A \leqslant 3$, for which this issue becomes important.

We contrast the bias effects for logis(3.6) and MIX(0.4) in Figs. 4(g) and 4(h), comparing the percentage error of log(conditional probability *estimate*, $N=1000$) vs log(true conditional probability), for each subcube with $B>10$. The average percent error/subcube, given $B>10$ is 9.5% for logis(3.6) and 40.5% for MIX(0.4). For $N=1000$, virtually all "A's" for MIX(0.4) in conditional probability estimates A/B were 0, 1, or 2. This effect is exacerbated as m grows larger and as r shrinks.

Messages, Comparison of 2 Maps: For true dynamical (often chaotic) systems, there is extreme sparseness in reconstructed dynamics, producing well-estimated conditional probabilities for a range of embedding dimensions m and scaling ranges r, with small influence from statistical bias. Here exploit the structure by choosing large m, small r in algorithms, to estimate a fine-graded measure description. For general autocorrelated maps, one typically sees diffuse reconstructed dynamics, with many poorly estimated conditional probabilities for $m \geqslant 3$ and small r; further, many *rare* events produce added bias. Here, be cautious—choose small m (e.g., $m=2$) and moderate width r to ensure replicability of a partial (dynamical) measure description.

VIII. SUMMARY AND CONCLUSION

The principal focus of this article has been the description of a recently introduced regularity statistic, ApEn. Several properties of ApEn facilitate its utility for general time-series analysis: (i) ApEn is nearly unaffected by noise of magnitude below a *de facto* specified filter level; (ii) ApEn is robust to outliers; (iii) ApEn can be applied to time series of 100 or more points, with good confidence (established by standard deviation calculations); (iv) ApEn is finite for stochastic, noisy deterministic and composite (mixed) processes, these last of which are, e.g., likely models for complicated biological systems; (v) increasing ApEn corresponds to intuitively increasing process complexity in the settings of (iv). It thus appears that ApEn has potential widespread utility to practical data analysis, based on these five properties, and the aforementioned clinical applications.

[1] S. M. Pincus, "Approximate entropy as a measure of system complexity," Proc. Natl. Acad. Sci. USA **88**, 2297–2301 (1991).
[2] S. M. Pincus, "Approximating Markov chains," Proc. Natl. Acad. Sci. USA **89**, 4432–4436 (1992).
[3] L. A. Fleisher, S. M. Pincus, and S. H. Rosenbaum, "Approximate entropy of heart rate as a correlate of postoperative ventricular dysfunction," Anesthesiology **78**, 683–692 (1993).
[4] D. T. Kaplan, M. I. Furman, S. M. Pincus, S. M. Ryan, L. A. Lipsitz, and A. L. Goldberger, "Aging and the complexity of cardiovascular dynamics," Biophys. J. **59**, 945–949 (1991).
[5] S. M. Pincus, T. R. Cummins, and G. G. Haddad, "Heart rate control in normal and aborted SIDS infants," Am. J. Physiol. **264** (Regulartory & Integrative **33**), R638–R646 (1993).
[6] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis," J. Clin. Monit. **7**, 335–345 (1991).
[7] S. M. Pincus and R. R. Viscarello, "Approximate entropy: A regularity measure for fetal heart rate analysis," Obstet. Gynecol. **79**, 249–255 (1992).
[8] M. L. Hartman, S. M. Pincus, M. L. Johnson, D. H. Matthews, L. M. Faunt, M. L. Vance, M. O. Thorner, and J. D. Veldhuis, "Enhanced basal and disorderly growth hormone (GH) secretion distinguish acromegalic from normal pulsatile GH release," J. Clin. Invest. **94**, 1277–1288 (1994).
[9] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," Physica D **9**, 189–208 (1983).
[10] A. N. Kolmogorov, "A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces," Dokl. Akad. Nauk. SSSR **119**, 861–864 (1958).
[11] S. M. Pincus and A. L. Goldberger, "Physiological time-series analysis: what does regularity quantify?," Am. J. Physiol. **266** (Heart Circ. Physiol. **35**), H1643–H1656 (1994).
[12] S. M. Pincus and W. M. Huang, "Approximate entropy: Statistical properties and applications," Commun. Statist. Theory Meth. **21**, 3061–3077 (1992).
[13] S. M. Pincus and D. L. Keefe, "Quantification of hormone pulsatility via an approximate entropy algorithm," Am. J. Physiol. **262** (Endocrinol. Metab. **25**), E741–E754 (1992).
[14] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time-series," Physica D **16**, 285–317 (1985).
[15] S. M. Pincus, "Greater signal regularity may indicate increased system isolation," Math. Biosci. **122**, 161–181 (1994).
[16] P. Grassberger and I. Procaccia, "Estimation of the Kolmogorov entropy from a chaotic signal," Phys. Rev. A **28**, 2591–2593 (1983).
[17] J. P. Eckmann and D. Reulle, "Ergodic theory of chaos and strange attractors," Rev. Mod. Phys. **57**, 617–656 (1985).
[18] F. Takens, "Invariants related to dimension and entropy," in *Atas do 13.Col.Brasiliero de Matematicas* (Rio de Janeiro, 1983).
[19] D. S. Ornstein, "Bernoulli shifts with the same entropy are isomorphic," Adv. Math. **4**, 337–352 (1970).
[20] D. S. Ornstein and B. Weiss, "How sampling reveals a process," Ann. Prob. **18**, 905–930 (1990).
[21] D. S. Broomhead and G. P. King, "Extracting qualitative dynamics from experimental data," Physica D **20**, 217–236 (1986).
[22] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Phys. Rev. A **33**, 1134–1140 (1986).
[23] G. Mayer-Kress, F. E. Yates, L. Benton, M. Keidel, W. Tirsch, S. J. Poppl, and K. Geist, "Dimensional analysis of nonlinear oscillations in brain, heart, and muscle," Math. Biosci. **90**, 155–182 (1988).
[24] M. Casdagli, "Nonlinear prediction of chaotic time series," Physica D **35**, 335–356 (1989).
[25] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," Phys. Rev. Lett. **59**, 845–848 (1987).
[26] G. Sugihara and R. M. May, "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," Nature **344**, 734–741 (1990).
[27] R. E. Blahut, *Principles and Practice of Information Theory* (Addison–Wesley, Reading, MA, 1987), pp. 55–64.
[28] The reason that correlation dimension is 0, not 1 for $p<1$ is that the sampling frequency of the sine function is commensurate with π, so that the phase space realization from the sine function is a discrete point set. If

we choose a sampling frequency incommensurate with π in the sine component of the MIX definition, the correlation dimension $=1$ almost surely for $p<1$, and the difficulty remains—the correlation dimension fails to discriminate members of this process from one another.

[29] A. N. Kolmogorov and Y. A. Rozanov, "On strong mixing conditions for stationary Gaussian processes," Theory Prob. Appl. **5**, 204–208 (1960).

[30] A family of ϵ estimators for ApEn(m,r) is proposed in Ref. 11 to achieve bias reduction.