CONTROL ENTROPY:
A COMPLEXITY MEASURE FOR NONSTATIONARY SIGNALS

ERIK M. BOLLT AND JOSEPH D. SKUFCA

Clarkson University, P.O. Box 5815, Potsdam, NY 13699-5815, USA

STEPHEN J. MCGREGOR

Health Prom and Human Perf, 318 Porter Bldg, Ypsilanti, MI 48197, USA

(Communicated by Ying-Cheng Lai)

ABSTRACT. We propose an entropy statistic designed to assess the behavior of slowly varying parameters of real systems. Based on correlation entropy, the method uses symbol dynamics and analysis of increments to achieve sufficient recurrence in a short time series to enable entropy measurements on small data sets. We analyze entropy along a moving window of a time series, the entropy statistic tracking the behavior of slow variables of the data series. We employ the technique against several physiological time series to illustrate its utility in characterizing the constraints on a physiological time series. We propose that changes in the entropy of measured physiological signal (e.g. power output) during dynamic exercise will indicate changes in underlying constraint of the system of interest. This is compelling because CE may serve as a non-invasive, objective means of determining physiological stress under non-steady state conditions such as competition or acute clinical pathologies. If so, CE could serve as a valuable tool for dynamically monitoring health status in a wide range of non-stationary systems.

1. **Introduction.** Information theoretic analysis of dynamical data has become a popular approach to leverage thermodynamic concepts, information theory, and statistical mechanics in an attempt to quantify the so-called "complexity" of the system is responsible for generating the data. Particularly for the biological and physiological data-sets , quantifying disorder of the system has become popular as an intense area of promising recent research; see for example cardiac variability analysis [19, 16, 9, 13], gait analysis [15, 8], circadian rhythm [1], postural control [6, 7], and other important physiological experimental time series have been analyzed by such methods. At the heart of such analysis, is the concept of quantifying the information evolution of transitions associated with probabilities assigned to each state, with a goal of providing single value (an entropy) to describe this information content. For example, with an appropriate finite partition of labeled states, $i = 1, 2, ..., n$, and a probability measure p_i on that partition, the Shannon entropy of a random variable is defined [30, 10] by

$$SE = - \sum p_i \ln p_i. \tag{1}$$

However, given a finite data set (for instance, a portion of a time series), the issue of how to partition and how to best estimate these quantities p_i is not trivial. When the data comes from a dynamical system, one may proceed as Grassberger and Proccia, to generalize Shannon entropy by constructing the the Renyi spectrum of entropies, [14, 26]. From a time series of measurements, $\{x(t_i)\}_{i=1}^N$, after constructing delay vectors, $\mathbf{X}(t_i) = (x(t_i), x(t_i + \tau), x(t_i + (m-1)\tau), i = 1, ..., n, n = N - (m-1)\tau$, to create the m-dimensional Taken's embedding [32, 29, 12]. Then in terms of the m-dimensional partition of uniformly sized hypercubes of side r-hypercubes, a relative probability p_i may be associated to be the relative occupancy of the orbit in the i-th hyper-cube. Then Renyi-entropies are defined,

$$K_q = \lim_{r \to 0} \lim_{m \to \infty} \frac{1}{1-q} \ln I_q(r), \text{ where, } I_q(r) = \sum (p_j)^q, q \ge 0, \qquad (2)$$

although, more generally one would need to take the supremum over all possible partition refininements, rather than simply refining the partition. As a special case, Komolgorov-Sinai entropy [31, 12], h_{KS} -also called measure theoretic entropy, associated with K_1, provides an ergodic quantifier of information theoretic complexity of a dynamical system. It defines a rate at which a dynamical system loses initial precision, and amplifies external noise, where this rate is emphasized according to the system's natural measure [22]. However, h_{KS} is generally not easy to accurately estimate from data, with difficulty in actually quantifying the appropriate probabilities. Another special case is the correlation entropy (K_2) [12, 17], defined by (2) with $q = 2$. In addition to being a lower bound of KS-entropy as discussed in [17], due to convexity [35], it serves in its own right as a suitable quantifier of diversity of a data set. Most important to application is that K_2 can be quickly and accurately computed from a correlation integral, for which there are excellent algorithms for estimation from finite discrete data sets, such as in [33].

A great deal of attention has been paid lately toward a particular statistical estimator of an entropy of a given time series, called Approximate Entropy (ApEn) by Pincus [23, 25]. Recently, an "unbiased" estimator of an entropy called Sample Entropy (SampEn) was developed by Richman and Moorman [27] as a counterpoint of ApEn. We do not present the details of SampEn of a signal here, since it is well defined in [27]. We will say only that the codes are widely accessible on the author's webpage, and they serve as a good estimator of K_2. However, any code which adequately estimates K_2 (for some q) could be substituted in its place, such as the subroutines of the TISEAN package [17] or the maximum likelihood estimator of [21]. A typical application is to use the algorithms to compute a regularity measure on a time series and then to use that computed value to classify the time series as being of one type or another. In many biological applications, they have been used to distinguish "healthy" from "unhealthy" biological signals [24, 34, 13].

As an extension of these applications, we consider the problem of continuous health monitoring, where the time series is not a fixed and complete set, but is "streaming." If we can associate a change in signal complexity with a change in health of the system, then we might hope that an entropy-like measure might detect a developing problem (and possibly provide some warning before system failure). An inherent difficulty is that the basis for the entropy estimators makes an assumption of sufficient stationarity of the process such that the estimates of probabilities p_i can be accurately and efficiently collected from the finite data sample. Moreover, our goal of detecting a change in the system implies an essential non-stationary at

some time-scale. We remark that generally, empirical KS analysis assumes that the process that generates the time series is sufficiently stationary that the underlying attractor in the phase space of the dynamical system becomes sufficiently sampled so that probability estimates become representative of the ergodic statistic. However, estimators such as ApEn and SampEn may be viewed as simply a statistic of the finite sample, without requiring stationarity. But even under this less formal interpretation, for these statistics to reasonably interpret the complexity of the signal requires that there is sufficient recurrence so that computed values can be interpreted as estimates of transition probabilities. Our main question in this work will be the development of a method to infer observations from a time series, of solutions from the system, whether there has been a significant parameter drift within the system.

Many real world systems evolve in time may be idealized by a general deterministic model of the form,

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t), \lambda), \ \mathbf{x}(t) \in \Re^m, \lambda \in \Re^q. \tag{3}$$

In this form, we may explicitly define a stationary system to be a dynamical system of the form (3), where the parameter(s) λ is (are) constant and where there is no explicit dependence on time in the description of f. One may argue that no real world system maintains its "constant" parameters λ, to be constant indefinitely. For example, consider the simple forced spring oscillator, $\ddot{y} + \gamma \dot{y} + \lambda = A \sin wt$, which may be written in the form of (3) with constant parameters for short times scales, but on much longer time-scales, the spring constant λ may actually be time varying and perhaps decaying. To emphasize fast variables $\mathbf{x}$ and slow variables λ, it is common to write the standard singularly perturbed system [5],

$$\begin{aligned} \dot{\lambda} &= \mathbf{f}_1(\lambda, \mathbf{x}), \\ \epsilon \dot{\mathbf{x}} &= \mathbf{f}_2(\lambda, \mathbf{x}). \end{aligned} \tag{4}$$

whose long term dynamics of the slow variable λ corresponds to the $\epsilon = 0$ slow manifold, with corresponding differential-algebraic system, $\dot{\lambda} = \mathbf{f}_1(\lambda, \mathbf{x}), \mathbf{f}_2(\lambda, \mathbf{x}) = 0$. However, when calling λ to be *approximately constant*, we are referring to the so-called associated system [2]; in the fast time-scale, $s = t/\epsilon$, hence (4) becomes,

$$\begin{aligned} \lambda' &= \epsilon \mathbf{f}_1(\lambda, \mathbf{x}), \\ \mathbf{x}' &= \mathbf{f}_2(\lambda, \mathbf{x}), \end{aligned} \tag{5}$$

and in the limit $\epsilon \to 0$, (5) yields the associate system,

$$\mathbf{x}' = \mathbf{f}_2(\lambda, \mathbf{x}), \ \lambda = constant. \tag{6}$$

In this sense, it can be said that the parameter λ changes slowly in time.

In the rest of this paper, we will write $\mathbf{x}(t)$ to denote a vector valued function parameterized by time-t,

$$\mathbf{x} : \Omega \in \Re \to: \Re^m, \tag{7}$$

where Ω is the scalar time domain. Similarly, $x(t)$ will be written to denote a scalar measurement of the vector value,

$$h : \Re^m \to \Re, h[\mathbf{x}(t)] \mapsto x(t), \tag{8}$$

typically one of the vector components. Our measurements may come from a dynamical system, or a stochastic process, with measurement error. It is also notable

that real world data sets from a continuous process are necessarily discretely sampled, and we will write,

$$x_i \equiv x(t_i), i = 0, 1, ..., n, \tag{9}$$

where typically, $t_{i+1} - t_i = constant$, is a uniform time grid controlled largely by the engineering of the experimental measuring devices. For this reason, it is natural to consider not just continuous time processes such as flows from differential equations, or solutions from stochastic differential equations. For the rest of the presentation of this paper, we restrict our model problems to be discrete time maps, $\mathbf{F} : \Re^m \to \Re^m, \mathbf{x}_i \mapsto \mathbf{x}_{i+1} = \mathbf{F}(\mathbf{x}_i)$, which may be related to a flow by the Poincaré section method or by the stroboscopic method.

2. **Development of control entropy (CE) .** Our goal is to develop an entropy like regularity statistic that can be applied against (possibly) nonstationary time series data in a way that allows insight into the (possibly) time varying parameters of the system. While in the case of a chaotic dynamical system with an invariant measure, it is suitable to discuss the full Renyi spectrum of entropies [17], it becomes less well founded for our goal problem. For this reason, we adopt the philosophy of ApEn [23, 25], and similarly SampEn [27], developing a statistical measure of a data stream without a specific claim that the statistics should be converging to some "true" entropy value. Consequently, we may tune the algorithm to improve its utility in a particular setting. As such, our techniques are meant to allow the user to exploit (where possible) any expert knowledge of the underlying physical system that generates the data. The general characteristics of our statistic, which we call *control entropy*, are as follows:

1. The control entropy should be calculated in a moving window along the time series, generating an entropy value for each of many (possibly overlapping) windows. As such, the original data time series may be viewed as generating an associated control entropy series. In some cases, the physical system that generates the data might lead to reasoned choices of window, but one might also consider under multiple window sizes provides a multiscale view of the data.

2. Computing entropy on only a window of the data necessitates a rapidly converging statistic. As such, we use entropy based on correlation sums [17], closely related to the technique of SampEn, but where we allow for removing the effects of short time correlation by using a Theiler window [33].

3. To increase the stationarity of the analyzed signal, when we consider time series signal $x(t)$, control entropy of $x(t)$ is based on a correlation entropy of $\frac{\mathrm{d}x}{\mathrm{d}t}$. For data signals sampled in discrete time, we simply compute on the finite difference (or increment) of the signal. That differences often render the signal stationary, can be viewed as an empirical observation of many real world time series [11, 17] without trying to establish a rigorous basis. We remark that many of the stochastic processes that arise in modeling (Levi process, Brownian motion, Poisson process, etc.) have *stationary increments* as a fundamental assumption of the model, making our choice to analyze the differences a very natural approach to the non-stationarity.

4. Standard methods for computing correlation entropy estimates on a continuous space require the choice of a "neighborhood size," which is related to the partitioning required to compute a numerical approximation of (2). As an alternative, one may transform the continuous signal into a sequence of

symbols, and analyze the entropy of the symbol stream. We will often apply this symbolization approach, using the methodology of [20], with a primary goal of reducing the number of tunable parameters to increase the robustness of the approach [18].

To define control entropy via correlation sums, we proceed as follows: Let $\{z_i\}_{i=1}^N$ be a scalar time series where data is taken on a uniform time grid. Without loss of generality, we assume unit spacing in time. For embedding dimension of m, let $\mathbf{v}_i = (z_i, z_{i-1}, \ldots, z_{i-m+1})$ be a delay embedding, where we assume a unit embedding delay. The correlation sum [17] is given by

$$C_2(\{z_i\}; m, r, T) = \frac{1}{N_{pairs}} \sum_{i=m}^{N} \sum_{j<i-T} \Theta(r - \|\mathbf{v}_i - \mathbf{v}_j\|_\infty), \tag{10}$$

where Θ is the heaviside function, r is a parameter defining a neighborhood, and $N_{pairs} = (N-m+1)(N-m-w+1)/2$ is the total number of pairs of delay vectors. Integer parameter $T \geq 1$ is the *Theiler window*, which mitigates the effects of time correlation in the data. We define

$$\hat{h}_2(\{z_i\}; m, r, T) = \ln\left(\frac{C_2(\{z_i\}, m, r, T)}{C_2(\{z_i\}, m+1, r, T)}\right). \tag{11}$$

Using this formulation as a regularity measure (without trying to consider scaling regimes and limiting values) is essentially equivalent to SampEn, which motivates are choice of symbology for the windowed version. Consider time series $\{z_i\}_{i=1}^n$: Define

$$SE(j + J, w, \{z_i\}_{i=1}^n; m, r, T) := \hat{h}_2(\{z_i\}_{i=1+j}^{w+j}; m, r, T),$$
$$0 \leq j \leq n - w, \tag{12}$$

where J represents a time offset. SE assigns an entropy value to a time window of the dataset. J allows us to associate that entropy with a specific instant in time. Choosing $J = w$ associates the entropy of the window with the ending time of that window, while $J = w/2$ would associate that entropy with the time at the middle of the window. We now define control entropy CE by applying SE to the first difference,

$$CE(j + J, w, \{x_i\}_{i=0}^n; m, r, T) := SE(j + J, w, \{x_i - x_{i-1}\}_{i=1}^n; m, r, T),$$
$$0 \leq j \leq n - w - 1. \tag{13}$$

The CE statistic may be reasonable applied to both deterministic and stochastic process. In 2.1, we show a short example based on Pincus' MIX_p signal [23] and a stochastic random walk. Additionally, in Appendix A, we apply the method to a simple system of deterministic chaos. That example focuses on the robustness of the procedures in the presence of significant observational noise and added non-stationarity of the measured variable.

2.1. **A motivating, historical example.** In [23], Pincus defines a family of signals called MIX_p as a prototype for use in study of ApEn. In that setting, fixed parameter $0 = p = 1$ determined the amount of randomness. For any p, the signal is mean 0 with unit variance. Pincus showed that an entropy measure could be used to assess the randomness of the signal. We will use MIX_p as a building block to create a signal that will illustrate the CE tool.

Let us assume that instead of being fixed, the parameter p is slowly varying in time. (For our example, $p(t)$ is a prescribed function, graphed in Fig-1. At each instant in time $(t = j)$, the MIX process uses the current value of $p(t)$ to generate a step size m_j which feeds a random walk, given by

$$x_j = x_{j-1} + m_j,$$

where we take $x(0) = 0$ as a starting point. The result sequence is a sample from a generalized random walk, with strong correlations in time. Although the increments are not stationary, they are always mean 0 with unit variance. (See Appendix B for a detailed description of this process.) The goal of the CE analysis attempts to recover the shape of $p(t)$ through entropy measurement of $\{x_j\}$. Figure 1 shows a plot of signal $\{x\}$, parameter $p(t)$ and an windowed SE and CE analysis. We remark that on windows that are small enough to follow the changes in p, there is insufficient recurrence in $x(t)$ to provide reliable entropy analysis, and SE provides little insight to the parameter, while CE provides a reasonable recovery of the variations in $p(t)$.
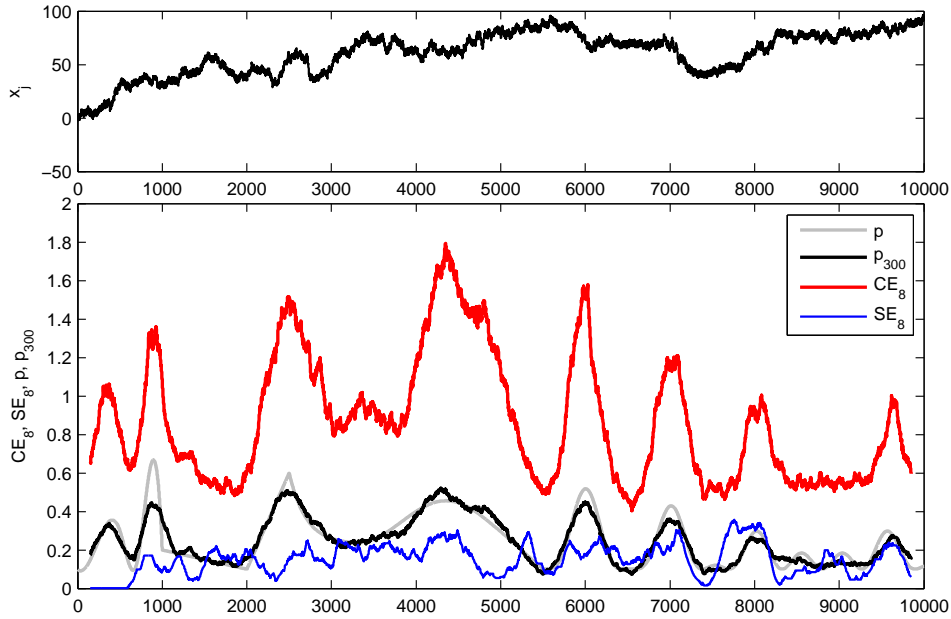

FIGURE 1. **A random walk from** Mix_p. (Top) The signal $\{x_j\}$ to be analyzed. The "constant width" of the noise (small scale oscillations) is due to unit variance of the Mix_p signal, which provides the increment for the random walk. (Bottom) Drive signal $p(t)$ (gray) was specified. The black signal p_300 shows resultant (random) realization, where the plotted value shows the fraction of random steps within a length $w = 300$ moving window. CE_8 (red) (using a 8 symbol discretization of data, as described in Sec-2.2, and window size $w = 300$) provides a very reasonable tracking of $p(t)$. However SE_8 (blue) is unable to follow the changes in the parameter.

2.2. **Symbolization and SAX partitions.** Symbolic methods have a long history in application to dynamical systems. In many statistical and datamining environment, converting from continuous signals to a symbol space has been shown to be a robust way of detecting changes in signal characteristics [20, 18]. Let us assume that we have a reasonable method for converting continuous values x_i to symbols s_i, where the cardinality of the symbol set is b. If we assume $\|s_i - s_j\| = 0$ when $s_i = s_j$, and 1 otherwise, then we can compute control entropy on the symbolized space by

$$CE_b(j + J, w, \{x_i\}_{i=0}^n ; m, T) = CE(j + J, w, \{s_i\}_{i=0}^n ; m, 0.5, T). \qquad (14)$$

We acknowledge that the symbolization procedure can give misleading results; we have written extensively about the role of generating partitions in dynamical systems and the computation of entropy with so-called "misplaced" partitions [4, 3]. The relevant message here, from our work, is that the role of the partition is important to the specific results. At best, one expects the numerical values to be biased, but it may also lead to nonintuitive non-monotonicity (inconsistency) of the results. With respect to symbolic dynamics from dynamical systems, we take a broad view here, toward a principle that the partition should be designed to allow the data set to demonstrate its full "complexity." While, for an infinite data set, refining the partition to allow more symbols may lead to better results, refinement eventually leads to poor results for *finite* data sets because of an under-sampling of recurrences through the partition. On the other hand, as analyzed extensively in [4, 3], when investigating a highly complex signal with a theoretical entropy $> \ln b$, then an b-symbol partition cannot reveal the full complexity.

We take the design goal of a useful partition to be to facilitate the finite signal to "maximally express itself" in the sense of frequent recurrences through the designed partition while maintaining a simplicity of implementation. As a general approach, one would like to have approximately equal usage of each of the symbols in the chosen symbol set. In Fig-2, we show an empirical histogram of the power data, $\{x(t_i)\}$, for the Giro d'Italia data set from Fig-11, which suggests a normal distribution, in this case. Further, in Fig-11 (Right), we show a windowed collection of histograms resulting from sliding block windows across the same data set. In this case, we see a generally unimodal distribution is maintained throughout. This motivates our standard partitioning method, called SAX, Symbolic Aggregate approXimation, [20]. The method fits an appropriate normal distribution, $\mathcal{N}(\mu, \sigma)$, to the data set $\{x(t_i)\}$, and then converting to the standard normal distribution $\mathcal{N}(0, 1)$, by the usual $Z = \frac{x(t_i) - \overline{x}}{\sigma}$, we may associate a integer $j \in 1, 2, ..., b$ by a uniform partition of the z-score of each $\{x(t_i)\}$. In particular applications, we may find that the data set requires other partitioning schemes, but the SAX approach provides an "off-the-shelf" approach that appears to perform satisfactorily against a wide variety of signals.

Although these symbolic procedures reduce the number of parameters, we must still choose an appropriate alphabet size. We note that the theoretical maximum entropy on b symbols is $\ln b$. In practice, if we find $CE_b(\{x(t_i)\} \approx \ln b$, then we say that the signal is saturating the entropy measurement, indicating that the partition is not fine enough to accurately capture the relevant dynamics. The competing limitations to balance in choosing b are that,

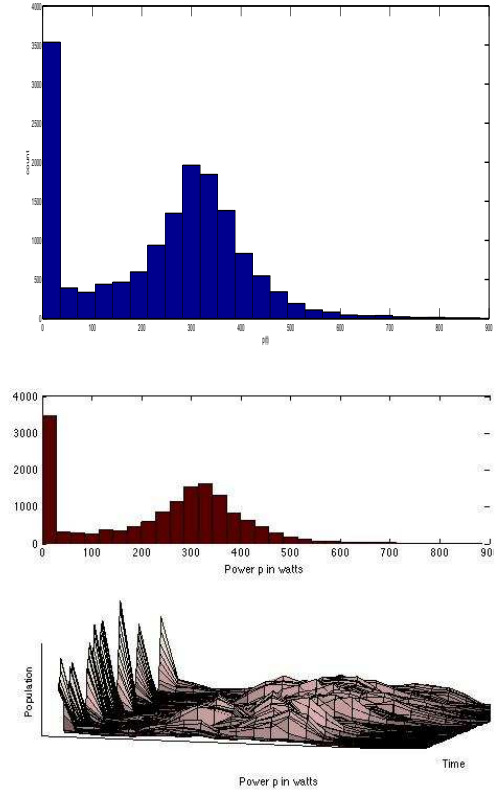- if b is too small, there may be saturation, but,

FIGURE 2. Data histogram(s) of the power data from a professional bicycle rider on the open road, as shown in Fig-11. (Top) A pdf suggestive of a normal distribution, supporting the use of the SAX algorithm. The notable deviation from normality is the $p(t) = 0$ bin, which is due to the moments when the bicycle racer "coasts" during a ride. (Right) Windowed Histograms, moving in time, maintains a generally unimodal distribution, for this data set, suggesting some validity in this case in maintaining the partition developed using the full time series.
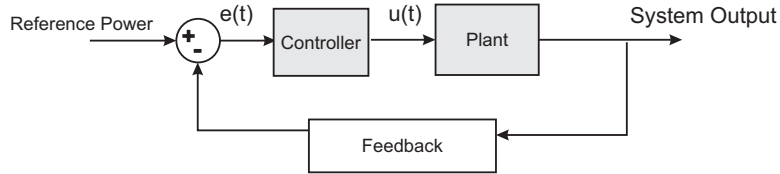

FIGURE 3. Basic feedback controller. $e(t)$ is the error between the reference power and the actual power. $u(t)$ is the control signal.

- if b is too large, we may observe lack of recurrence, making estimates of transition probabilities (and the entropy estimate) unreliable.

See the example data set in Fig-11 (Left) of a power data set which suggested to us the need of just such a non-saturating CE_b as in Fig-11 (Right) in which $b = 8$ symbols are chosen for SAX-8.
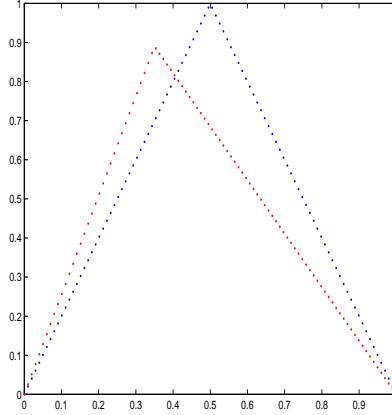

FIGURE 4. Two Tent Maps with different underlying entropy. The time series experiment in Fig-14 is run for half of the time with the higher entropy blue map, and then is switched abruptly to the lower entropy red map.

2.3. **Biomedical and physiological examples.** The kind of natural data sets with nonstationarity we are now able to address, consider the samples demonstrated in Figs. 5-10. These are each biomedical and physiological examples for purpose of this paper, each with the characteristic nonstationarity renders traditional entropy methods more difficult. These examples range from cyclists power output during exercise in the lab during a Vo2Max test, Fig-5, a maximal aerobic power test Fig-6, during an intervals session in the lab, Fig- 7, to heart rate data taken during ergonomic testing, Fig- 8, to accelerometer on muscles during isometric exercise data, Fig- 9, and accelerometer data taken during treadmill testing for VO2max Fig- 10. One may compare this to nonlaboratory data taken on the open road during a professional Grande Tour bicycle race, Fig- 11.

2.4. **Time-scales and variability.** We are primarily interested in how the data set may evolve in time. In particular, we hope to assess the behavior of some underlying, slowly varying parameter. We use a windowed fraction of the data, we compute the statistic CE on windows of size w, and then observe the windowed CE value as it evolves in time. The choice of window size must balances competing factors of:

- Smoothing and convergence of the estimated statistic CE with *larger* w. Larger numerical samples reduce statistical fluctuations in the computed statistic for each sliding window. If w is too small, the fluctuations may be so large as to mask the underlying parameter change of interest.
- Detecting changes on (relatively) short time-scales. If w is chosen too large, the window may stretch across several "features" of a parameter change. The
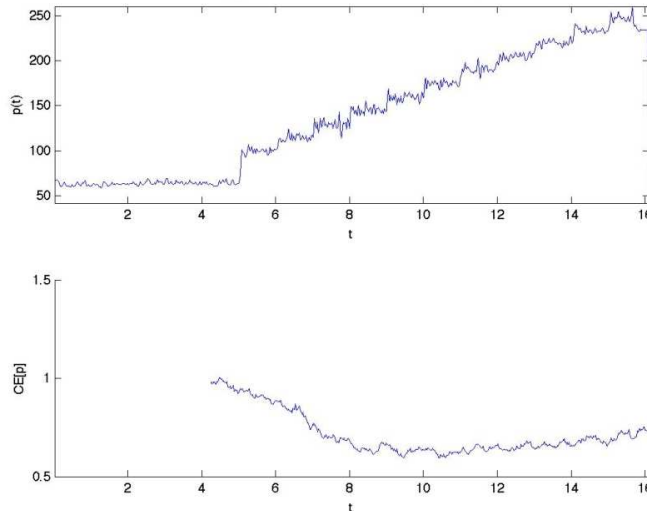
FIGURE 5.   CE of power output during incremental cycling VO2max test on laboratory ergometer. (Upper) Plot of $p(t)$, measured power in Watts, versus t in minutes. The subject completes a warm up at 60 Watts. At 5 minutes, the workrate is increased to 100 Watts, and increases thereafter 15 Watts per minute until exhaustion while metabolic gasses are collected (data not shown). (Bottom) The CE of power as a function of time t displays what we have observed to be a typical scenario during stressful exertions, that being a decline in CE.

CE algorithm assigns an "average" value to the window. Real fluctuations of the parameter that occur on time-scales that are shorter than the window length may be difficult to discern.

The balance is similar to that of standard moving average filters, with longer windows providing smoother curves while limiting the capability to follow higher frequency fluctuations in the signal of interest.

The examples, shown in Fig-13, as well as Fig-12, illustrate the smoothness gained by longer windows, while shorter time-scales pushes the limits of limited data for convergence of the entropy statistic. We remark that the average period of the underlying system drives the sampling rate that is needed for the discrete time series to capture the full dynamics, and such sampling requirements are well understood. However, since the signal of interest is the changing behavior of some system parameter, that theory cannot be directly applied. To find the appropriate data scale, we run the the windowed CE analysis, across many time-scales, as shown in Figures 12 and 13.

3. **The name control entropy.** In this section, we motivate the choice of the term *Control* Entropy to define our statistic. Although the formulation of section 2 may be viewed simply as an estimator of correlation entropy of the derivative of a time series, we focus our description on a primary application of interest,
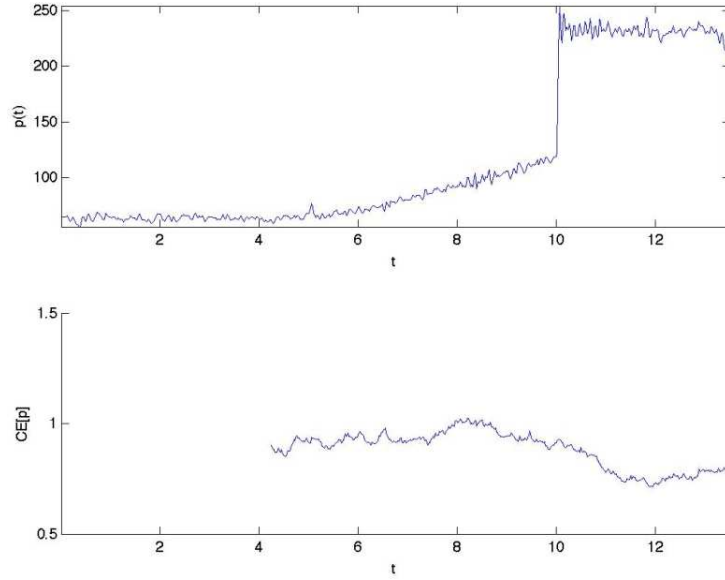
FIGURE 6. A cycling effort on laboratory ergometer at maximal aerobic power (MAP) established by VO2max test shown in Fig-5. (Upper). Again, $p(t)$, measured power in Watts, plotted versus t in minutes. The subject performs a warm up at 25%, ramp to 50%, then maximal effort at MAP until exhaustion. (Bottom). CE of power is elevated during the warm up, and declines as workrate is ramped to 50%, then reaches a nadir during the effort at MAP despite the relatively short duration of the maximal effort.

where the measured time series is taken from some physical process where control is required. We remark that a vast number of physical systems of interest can be interpreted within the framework of control systems theory such that our discussions are relevant to the broader set of examples.

For specificity, let us suppose that we desire to analyze power signal $P(t)$, the measured output by a bicycle rider. Furthermore, suppose the rider is trying to achieve some specified power level, denoted P_{ref}. Obviously, over the course of time that value may change, but let us treat it as a constant over some time interval of interest, with the rider doing something to try to reach and maintain that level. We would expect that some type of feedback control description of the system would provide a reasonable characterization of the important dynamics. (See Fig 3)

Let $x(t)$ be defined by

$$x(t) = P(t) - P_{ref}, \tag{15}$$

where we now view $x(t)$ as carrying the state variable information (where some embedding may be required to properly characterize the state equations). A standard control theory approach is to assume that we are near the desired control point, so that a linearized approximation is reasonable, with the system governed by the dynamical system

$$\dot{x} = Ax + Bu, \tag{16}$$

where $\dot{x}$ denotes derivative with respect to time. $u = u(t)$ is a control signal, generated based on the sensed error. In our simple description, $x(t)$ is the difference
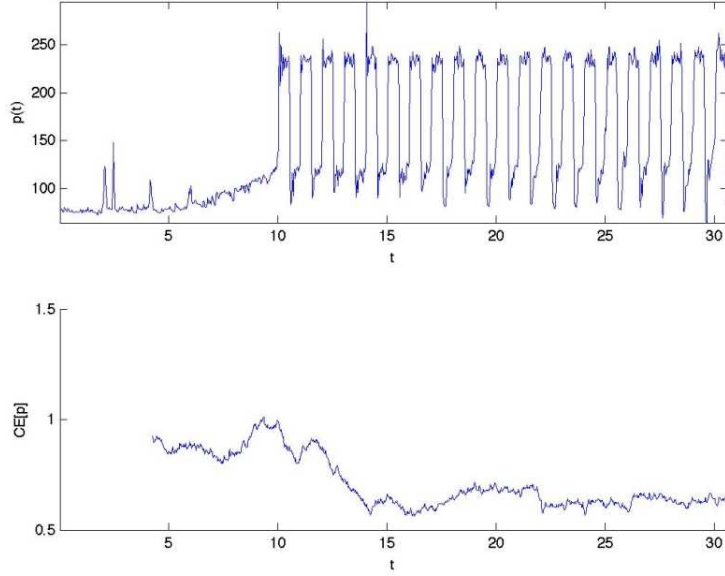
FIGURE 7.   An intermittent cycling effort on laboratory ergometer at MAP. (Upper). $p(t)$, measured power in Watts, plotted versus t in minutes. The subject performs a warm up at 25%, ramp to 50%, then effort is alternated between 100 and 50% of MAP until exhaustion. (Bottom). Again, despite the variable nature of the effort, CE of power declines at the end of the warm up and stays suppressed during the remainder of the protocol.

from reference power, and can be treated as the error, and the *controller* generates signal $u(t)$ as some function g of the error. If we continue the assumed simplification of a linear form, then this relationship can be described by

$$u(t) = g(x(t)) = Kx(t), \tag{17}$$

where K is a matrix that defines the controller. Then (16) can be rewritten as

$$\dot{x} = (A + BK)\, x. \tag{18}$$

The control system designer *chooses* K so that the resultant system of (18) is stable. CE is computed as an entropy of $\dot{P}$. If we assume that the rider is trying to maintain a constant power, then $dP_{ref}/dt = 0$ most of the time, deviating only during those short intervals when the rider decides to ride at a different power level. Differentiating (15), we have

$$\dot{P} = \dot{x} = (A + BK)\, x, \tag{19}$$

which we interpret as indicating that the CE computation as allowing direct insight into the control system behavior, providing an entropy measure of the control signal.

The linear analysis above is meant to justify the view that CE is descriptive of the controller is not rich enough to show why we consider the *evolution* of CE, as described in the time-window statistic of (13). To provide a heuristic argument, we consider the following general description of the controlled dynamical system: If x describes a vector of state variables for some system, then the controlled behavior
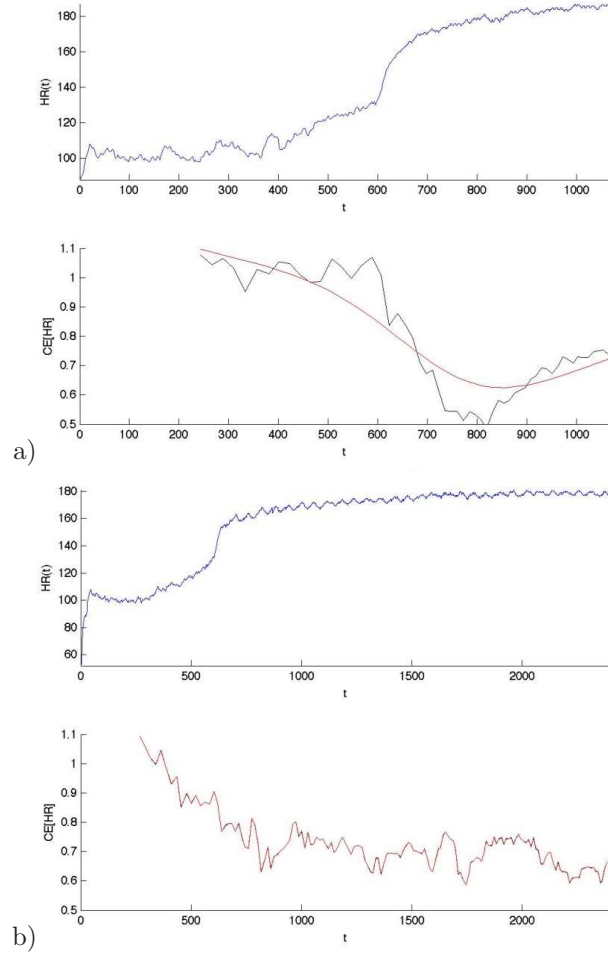
FIGURE 8. CE of heart rate during cycling on laboratory ergometer. In both (a) and (b), (Upper) HR(t), measured heart rate (bpm), versus t in seconds. (Lower). CE of heart rate. a) Constant effort cycling on laboratory ergometer at MAP: The same exercise protocol was used as presented in Fig-6. (Bottom). CE of heart rate declines, as did CE of power, once the maximal effort begins at 600 seconds. b) CE of heart rate during intermittent cycling effort at MAP. The same exercise protocol was used as presented in Fig-7. (Lower). As with CE of power, CE of heart rate declines despite the variable nature of the effort, and cardiovascular response.

is

$$\dot{x} = f(x, u_s(t), u_c(t), t), \qquad (20)$$

where f gives the functional dependence of the evolution of the system on state x at time t, with control u decomposed into a subconscious component u_s, and conscious component u_c. While the control signals will certainly depend upon some
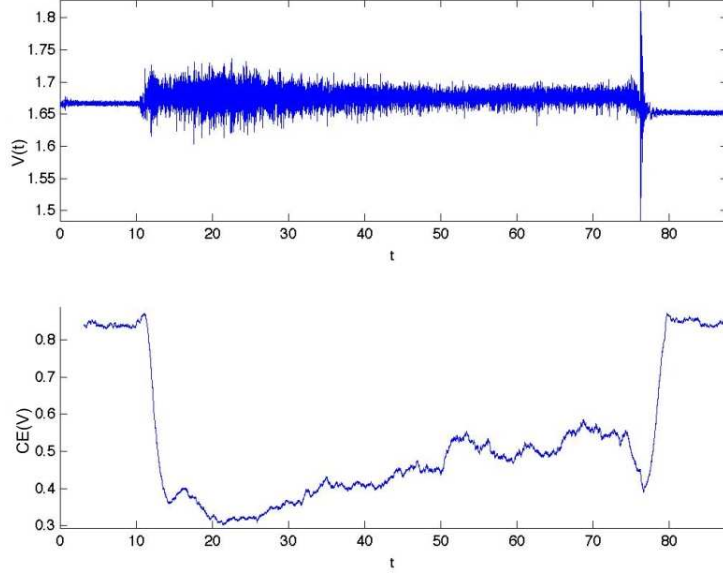
FIGURE 9. CE of mechanomyography (MMG) during isometric muscle action at 100% maximal voluntary contraction. (Upper). $V(t)$, measured voltage, versus t in seconds. Voltage is signal from a uniaxial accelerometer placed on the surface of the Vastus lateralis muscle of the thigh. Signal is collected during a 10 sec baseline rest, 70 sec maximal isometric leg extension, and 10 sec recovery. The accelerometer measures muscle vibration as a result of muscle action. (Bottom). CE of accelerometer voltage declines dramatically at the start of the contraction, and remains depressed until fatigue and return to baseline rest.

desired reference behavior, we make no specific effort to distinguish feedback and feedforward controllers, allowing that these control signals might also depend upon the state x. Then CE can be viewed as measuring the complexity of the signal

$$s(t) = f(x, u_s, u_c, t), \tag{21}$$

Changes in the complexity of this signal would indicate changes in the characteristics of the attractor of this system, which would logically be associated with either changes in the control system or changes in the underlying physiology of the system. Analysis of human gait experiments indicate that subjects walking at their natural pace (called unconstrained walking) show a complex pattern in stride interval, with long time correlations in the time series data, with these correlations disappearing when the person was required to walk to a metronome [15]. CE measurement of the signals indicates higher entropy for the metronome walking as compared to the unconstrained case. One interpretation of the stride walking data is that without a metronome, the body is required to generate a clocking signal to drive the overall cadence. With the goal of a steady walking pace, this requires periodic control signal u_c. The resultant signal is correlated in time, which implies that the future states is correlated with current state, so that less information (per-unit time) is required to describe the evolution of the system (a lowe entropy). When stepping with a

metronome, the periodic drive is generated by an external source, with entrainment to that signal requiring only a small amount of conscious (periodic) control. In other dynamic exercise situations, the "constraint" might be the individual's desire to maintain a certain work output (e.g. running speed). During extended efforts, as the body fatigues, the natural (unconstrained) action by the body would tend to slow down the system. Anecdotal evidence indicates that the athlete must concentrate harder to maintain a constant work rate. If we associate this concentration with an increased conscious control u_c, (asking the body to maintain a steady level), then the entropy of signal $s(t)$ would fall as the athlete becomes more fatigued. Additionally, we might expect that if the athlete loses concentration, u_c would be small, and the signal entropy would rebound to previous levels, though now the work rate output by the system might be significantly reduced. Although we use the u_c and u_s designations for this heuristic justification, from a practical standpoint, CE does not depend on conscious control. Simply stated, the time-varying methodology is meant to detect complexity shifts that would be indicative of changes in an underlying system parameter.

4. **An effect of observation noise on CE .** Since our methods are focused on application to measured data from physical systems, the effect of noisy observations must be understood. The methodology of [27] provides a means of assessing the expected fluctuations in the measurement due to the randomness of the model, but they do not account for noise. Specifically, if a process has a continuous noise input, its "true" entropy is infinite [17], with the measured entropy strongly dependent upon the observation scale. Although our choice of symbolization mitigates some of these effects, if one is attempting to assess parameter changes of the system through entropy measurement, one must be able to quantify the effects of observational noise. CE computations are fundamentally counting statistics. Counting a particular symbolization amounts to application of a characteristic function, which is inherently a nonlinear function, where standard statistical methods are difficult to apply. The goal of this section is to present a method to develop an algorithmic approach to assess an empirical distribution of error in the noisy CE computation.

Suppose we model the noise of the time-dependent random variables $X(t)$ leading to individual measurement realizations $x(t)$, where we acknowledge both a continuous measurement noise (error) as well as discretization errors which may occur in certain digital devices that quantize the reported signals: Let,

$$x(t) = \mathcal{X}(t) + \xi(t), \tag{22}$$

where $\mathcal{X}(t)$ is the "true" physical value being measured, with error due to added random noise variable $\xi(t)$. We take $\xi(t)$ to be a time-dependent random variable, with a time-dependent probability distribution function $\nu(t)$. Notice that we have assumed added noise to the signal, which should be sufficient for modeling measurement noise, (whereas other models might be required to study the stochastic evolution of the system.)

Suppose we wish to compute a time series of CE from measured data $\{x(t_i)\}$. Since different realizations of the added noise $\xi(t)$ at each instant t may lead to different symbolizations, different computed control entropy values could result. Therefore, *if we have a good model for the noise profile*, $\nu(t)$, we may use a random number generator to produce appropriately distributed instances of possible, $-\xi(t)$. Thus we will define a Monte-Carlo style simulation of the noise, from which we will simulate appropriate samples. In practice as computer code, we declare pseudo

random numbers $\hat{\xi}(t)$ (a coded subroutine to be built), to be distributed as $\nu(t)$ at each instant t. With $\hat{\xi}(t)$, we may generate the Monte-Carlo instance,

$$\hat{x}(t) = x(t) - \hat{\xi}(t), \tag{23}$$

For each sampled instance, $\hat{x}(t_i)$, we may compute instances of CE,

$$\hat{CE} = CE_b(j + J, w, \{\hat{x}(t_i)\}_{i=0}^n \, ; m, T), \tag{24}$$

where all parameters are chosen as in the analysis of the original time series. Repeated sampling, Monte-Carlo style, of this generated time series may be performed to sample the associate distribution of $\hat{CE}(t)$, at each t, with error bounds based on these distributions.

For the simulations shown here, we specialize to assume a typical form of $\nu(t)$. We take as an example the power in watts from the professional bicyclist in Fig-11 as measured by a device called an SRM Pro powermeter, which uses several piezo-electrical strain gauges built into the bicycle crank armature to infer applied torque. Measured values are reported in whole integer valued numbers of watts, in the range 0-2500 watts, with a margin of error imprinted from the factory of +/-2.5%, which we assume to indicate normally distributed and unbiased noise, which we take (as worst case) to be identically distributed. So we take

$$\hat{\xi}(t) \sim N(0, \sigma) + U(-0.5, 0.5), \text{ again where, } \sigma = 0.025x(t), \tag{25}$$

where the uniform noise term is meant to describe the discretization error by assuming a uniform rounding procedure.

Proceeding with the given measured signal, $\{x(t_i)\}$, and computing $\hat{CE}$ by (24) and with discretization and noise errors modeled (25), we generate many samples, N, of possible instances of $\hat{CE}$ time series. In Fig-11 (Right), we see the result of such repeated Monte-Carlo simulations of possible CE, resulting from possible original signals, through $N = 250$ simulated pseudo randomized signals. In the Figure, we have displayed 95% (blue) error bars from the simulation, which is the CE-value for each t such that 0.025 of the simulated pseudo randomized signals lie above (below) the sampled population. Also shown in red is the computed $CE[j; w, \{x(t_i)\}]$ from the measured signal $\{x(t_i)\}$ without any pseudo randomization. The expectation is that the CE-signal will lie within the blue error bars roughly 95% of the time. We observe in Fig-12 that as the window length increases, the error bands of CE become tighter. This is due in part to the particular form of the noise, for which repeated sampling gives better averages, and the essentially counting nature of the CE computation.

On the other hand, as the window length becomes too long, the computation averages through the slower-time scaled parameter variations, which are the very events which are being investigated. That, the time-scale of slow varying parameters is changed with window lengths, is also discussed in Sec-2.4. See also Fig-13 for further study of this time-scales issue.

5. **Discussion.** We have introduced a new entropy statistic designed to assess the behavior of real systems under the condition of nonstationarity. In particular the focus is on those systems where the nonstationarity is due to some slowly varying parameter of the system. Our focus is on application to systems where the change in parameter would be indicative of a change in the "health" of the system. *Control Entropy* is based on a windowed application of correlation entropy, but uses techniques from symbol dynamics (as well as the increased stationarity of signal

increments) to improve the robustness of the statistic. A key component of the approach is that to analyze for a slowly changing parameter, one must employ some type of windowed analysis of the signal. CE provides a sufficiently rapid convergence of the statistic to allow exploitation of this irregularity measure on real signals. In Sec-2.3, we have shown a vast array of examples on physiological data. In Sec-3 we give an heuristic argument as to how changes in CE might be related to the underlying control mechanism. These initial studies seem to indicate that CE may serve as a non-invasive, objective means of determining physiological stress under non-steady state conditions such as competition or acute clinical pathologies. Crucial to being able to properly interpret these signals, especially in a clinical setting, requires that the effects of measurement noise be understood. Section 4 explores a statistical methodology for determining when the changes in CE can be interpreted as "real" changes in the underlying system. With appropriate statistical tools in hand, it appears that CE could serve as a valuable tool for dynamically monitoring health status in a wide range of nonstationary systems.

Appendix A. Example of CE applied to deterministic chaos system with noisy observation. Consider a discrete-time chaotic map,

$$\mathbf{F} : \Re^1 \to \Re^1, \mathbf{x}_i \mapsto \mathbf{x}_{i+1} = \mathbf{F}(\mathbf{x}_i), \tag{26}$$

for which we choose the specific example, a family of skew tent maps, $F_{a,b} : [0,1] \to [0,1]$, with

$$F_{a,b}(x) = \begin{array}{l} b + ((1-b)/a)x \text{ if } 0 \le x < a, \\ (1-x)/(1-a) \text{ if } a \le x \le 1 \end{array} . \tag{27}$$

When $a = 1/2$, $b = 0$, it is known [28, 4] that this map has entropy exactly $h_2(F) = \ln(2)$.

In Fig-4, we show two different tent maps, one with $a = 1/2$, $b = 0$, and the other with slightly lower entropy, which will be the subject of demonstration in Fig-14.

In Fig-14 (Upper Left), we show a chaotic time-series $x(t)$ of $25,000$ iterates, where the parameters of the underlying dynamical system are abruptly changed at time $12,500$, from the maximal entropy tent map, to the submaximal tent map shown in Fig-4. Because the data-set is stationary, a windowed application of the popular SE successfully detects the parameter change, as Fig-14 (Middle Left) demonstrates an abrupt decrease in measured entropy. For reference, the vertical blue line is placed at the moment of the theoretical change in the map's parameters.

Next, we modify the stationary chaotic signal to produce what our "benchmark" of a nonstationary signal, shown in Fig-14(Upper Right). The dark curve, which we name $y(t)$, running through the middle of the erratic signal shown is an artificial and deliberately wandering baseline signal designed explicitly to produce a definitive nonstationarity. The erratic signal, $\tilde{x}(t)$, shown is then a sum of erratic components,

$$\tilde{x}(t) = x(t) + y(t) + \xi(t), \tag{28}$$

where $x(t)$ is the deterministic chaotic signal, $y(t)$ is the wandering baseline signal, and $\xi(t)$ is a random component as follows: Let $N(0, \sigma)$ be a normal random variable for which we choose a large variance, $U(0,1)$ the uniform random variable, and $H_{1/20}[s]$ is a Heaviside function, with trigger $s = 1/20$. Define $\xi(t)$ by,

$$\xi(t) = N(0, \sigma)H_{1/20}[U(0,1)]. \tag{29}$$

which yields a noise signal is "off" 95% of the time, but gives a large normal noise component on about 5% of the measurements. The designed signal $\tilde{x}(t)$ is highly

erratic with both chaotic and large wandering random components. Due to the high degree of nonstationarity, application of SE algorithm to the $\tilde{x}(t)$ signal fails to indicate the moment of the changed parameters, as no apparent change in SE at the time of parameter change (see Fig-14 (Middle Right)). However, as seen in Fig-14, CE also detects the event change in both the nonstationary and the stationary signal.

Appendix B. **A random walk from** Mix_p. **.** Pincus [23] defines a family of signals called MIX_p as a prototype for use in study of ApEn. In that setting, fixed parameter $0 = p = 1$ determined the amount of randomness. For any p, the signal is mean 0 with unit variance. Pincus showed that an entropy measure could be used to assess the randomness of the signal. We consider a slight modification where we allow p to be slowly varying in time and use this signal as the step size drive for a generalized random walk. To define our signal, let

$$w_j = a^{-l/2}\sin(\frac{2\pi j}{12}),\tag{30}$$

a deterministic sinusoid, where

$$a := \sum_{j=1}^{12}\frac{1}{12}sin(\frac{2\pi j}{12})),$$

is chosen so that the sinusoidal signal has unit variance. Let Y_j be a family of iid variables taken from the uniform distribution on the interval $[-\sqrt{3}, \sqrt{3}]$, observing that each Y_j also has unit variance. Define

$$Z_j = \left\{ \begin{array}{lll} 1 & \text{with probability} & p \\ 0 & \text{with probability} & 1-p, \end{array} \right.$$

The standard MIX_p signal (from [23]) is

$$MIX_p := (1 - Z_j)w_j + Z_jY_j.\tag{31}$$

At each time increment, the MIX_p is either a random value Y_j, with probability p, or it is a deterministic, periodic signal, with probability $1 - p$. The deterministic signal has entropy $h = 0$. With increasing p, the randomness of the signal increases, and entropy should increase. This signal is typically studied for fixed parameter value p. However, suppose the parameter p is varying with time to create a non-stationary signal:

$$w_j = a^{-l/2}\sin(\frac{2\pi j}{12}),\tag{32}$$

$$Z_j = \left\{ \begin{array}{lll} 1 & \text{with probability} & p_j \\ 0 & \text{with probability} & 1-p_j \end{array} \right.,\tag{33}$$

$$M_j = (1 - Z_j)w_j + Z_jY_j,\tag{34}$$

$$p_j = g(j).\tag{35}$$

The M_j are independent, with mean 0 and unit variance (though not identically distributed). We now take this signal to be the step size for a random walk process, defining

$$X_j = X_{j-1} + M_j,$$

where we presume a particular realization of the form

$$x_j = s_{j-1} + m_j.$$

For simplification, note that we choose to present p_j as a simple function of time. We reason that in a physical problem, there is (typically) not sufficient data to recover a phase space representation of the *slow dynamics* which govern the evolution of p. Rather, we only expect that we might recover the basic shape of the signal $\{p_j\}$. Consequently, for these study problems, we simply ignore the dynamics of p and prescribe an interesting test signal $\{p_j\}_{j=1}^N$ that will help us to explore the performance of our computational algorithms. We remark that the resultant generalized random walk is time-inhomogeneous Markov chain with strong time correlations and non-stationary increments.

REFERENCES

[1] S. Alesci, P. E. Martinez, S. Kelkar, I. Ilias, D. S. Ronsaville, S. J. Listwak, A. R. Ayala, J. Licinio, H. K. Gold, M. A. Kling, et al, "Major Depression Is Associated with Significant Diurnal Elevations," in "Plasma Interleukin-6 Levels, a Shift of Its Circadian Rhythm, and Loss of Physiological Complexity in Its Secretion: Clinical Implications," 2005.

[2] N. Berglund and B. Gentz, *Geometric singular perturbation theory for stochastic differential equations*, Journal of Differential Equations, **191** (2003), 1–54.

[3] E. M. Bollt, T. Stanford, Y. C. Lai and K. Życzkowski, *Validity of threshold-crossing analysis of symbolic dynamics from chaotic time series*, Physical Review Letters, **85** (2000), 3524–3527.

[4] E. M. Bollt, T. Stanford, Y. C. Lai and K. Życzkowski, *What symbolic dynamics do we get with a misplaced partition? On the validity of threshold crossings analysis of chaotic time-series*, Physica D: Nonlinear Phenomena, **154** (2001), 259–286.

[5] J. Carr, "Applications of Centre Manifold Theory," Springer, 1981.

[6] J. T. Cavanaugh, K. M. Guskiewicz and N. Stergiou, *A nonlinear dynamic approach for evaluating postural control: New directions for the management of sport-related cerebral concussion*, Sports Medicine, **35** (2005), 935–950.

[7] J. H. Challis, *Aging, regularity and variability in maximum isometric moments*, Journal of Biomechanics, **39** (2006), 1543–1546.

[8] M. Costa, A. L. Goldberger and C. K. Peng, *Multiscale entropy analysis of complex physiologic time series*, Physical Review Letters, **89** (2002), 68102.

[9] M. Costa, A. L. Goldberger and C. K. Peng, *Multiscale entropy analysis of biological signals*, Physical Review E, **71** (2005), 21906.

[10] T. M. Cover and J. A. Thomas, "Elements of Information Theory," Wiley-Interscience New York, 2006.

[11] C. Diks, "Nonlinear Time Series Analysis," World Scientific River Edge, NJ, 1999.

[12] J.-P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*, Rev. Mod. Phys., **57** (1985), 617–656.

[13] M. Ferrario, MG Signorini, G. Magenes and S. Cerutti, *Comparison of entropy-based regularity estimators: application to the fetal heart rate signal for the identification of fetal distress*, Biomedical Engineering, IEEE Transactions on, **53** (2006), 119–125.

[14] P. Grassberger and I. Procaccia, *Estimation of the Kolmogorov entropy from a chaotic signal*, Physical Review A, **28** (1983), 2591–2593.

[15] JM Hausdorff, PL Purdon, CK Peng, Z. Ladin, JY Wei and AL Goldberger, *Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations*, Journal of Applied Physiology, **80** (1996), 1448–1457.

[16] K. K. L. Ho, G. B. Moody, C. K. Peng, J. E. Mietus, M. G. Larson, D. Levy and A. L. Goldberger, *Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics*, Circulation, **96** (1997), 842–848.

[17] H. Kantz and T. Schreiber, "Nonlinear Time Series Analysis," Cambridge University Press, Cambridge, 2004.

[18] M. B. Kennel and A. I. Mees, *Context-tree modeling of observed symbolic dynamics*, Physical Review E, **66** (2002), 56209.

[19] J. Kurths, A. Voss, P. Saparin, A. Witt, H. J. Kleiner and N. Wessel, *Quantitative analysis of heart rate variability*, Chaos, **5** (1995), 88–94.

[20] J. Lin, E. Keogh, S. Lonardi and B. Chiu, *A symbolic representation of time series, with implications for streaming algorithms*, Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, (2003), 2–11.

[21] E. Olofsen, J. Degoede and R. Heijungs, *A maximum likelihood approach to correlation dimension and entropy estimation*, Bulletin of Mathematical Biology, **54** (1992), 45–58.

[22] E. Ott, T. Sauer and J. A. Yorke, *Coping with chaos. Analysis of chaotic data and the exploitation of chaotic systems*, Wiley Series in Nonlinear Science, New York: John Wiley,— c1994, edited by Ott, Edward; Sauer, Tim; Yorke, James A., 1994.

[23] S. M. Pincus, *Approximate entropy as a measure of system complexity*, Proceedings of the National Academy of Sciences of the United States of America, **88** (1991), 2297–2301.

[24] S. M. PINCUS, *Assessing serial irregularity and its implications for health*, Annals of the New York Academy of Sciences, **954** (2001), 245.

[25] SM Pincus and AL Goldberger, *Physiological time-series analysis: what does regularity quantify?* American Journal of Physiology- Heart and Circulatory Physiology, **266** (1994), 1643–1656.

[26] A. Renyi, *On measures of entropy and information*, Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, **1** (1961), 547–561.

[27] J. S. Richman and J. R. Moorman, *Physiological time-series analysis using approximate entropy and sample entropy*, American Journal of Physiology- Heart and Circulatory Physiology, **278** (2000), 2039–2049.

[28] C. Robinson, "Dynamical Systems; Stability, Symbolic Dynamics, and Chaos," CRC Press, second edition, 1999.

[29] T. Sauer, J. A. Yorke and M. Casdagli, *Embedology*, Journal of Statistical Physics, **65** (1991), 579–616.

[30] C. E. Shannon and W. Weaver, "The Mathematical Theory of Information," Urbana: University of Illinois Press, 97, 1949.

[31] A. G. Sinai, *On the concept of entropy of a dynamical system*, Dokl. Akad. Nauk. SSSR,, **124** (1959), 768–771.

[32] F. Takens et al, *Detecting strange attractors in turbulence*, Lecture Notes in Mathematics, **898** (1981), 366–381.

[33] J. Theiler, *Efficient algorithm for estimating the correlation dimension from a set of discrete points*, Physical Review A, **36** (1987), 4456–4462.

[34] DE Vaillancourt, AB Slifkin and KM Newell, *Regularity of force tremor in Parkinson's disease*, Clin Neurophysiol, **112** (2001), 1594–603.

[35] K. Życzkowski, *Rényi extrapolation of Shannon entropy*, Open Systems & Information Dynamics, **10** (2003), 297–310.

E-mail address: bolltem@clarkson.edu

E-mail address: jskufca@clarkson.edu

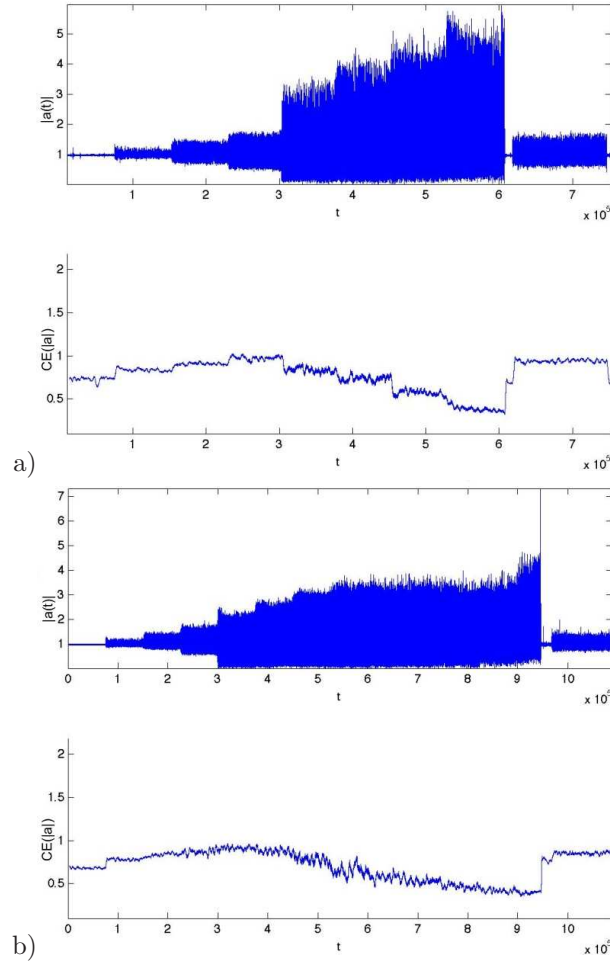*E-mail address*: stephen.mcgregor@emich.edu

FIGURE 10. CE of accelerometry during incremental treadmill VO2max test. In both a) and b), (Upper). $|a(t)|$, measured resultant acceleration magnitude in g's of gravity, obtained by triaxial accelerometer, versus t, sample number. Samples were recorded at 667 Hz for each axis while subjects wore the accelerometer placed at the approximate center of mass on the lower back. The protocol consisted of standing for 2 min, then walking on the treadmill at 2 kph, and speed was increased thereafter by 2 kph every 2 min until exhaustion. (Bottom). CE of acceleration increases during the walking phase until the run transition (red arrow) where CE declines until exhaustion. The biphasic response indicates that biomechanical constraints of walking are reduced as speed increases toward natural pace because energy provision is not limiting. In contrast, during running, energy provision becomes a limiting factor and increases constraints as speed increases. This interpretation is supported by comparing plots a) Untrained runner and b) Trained elite runner. Both exhibit the same biphasic response, yet, CE of accelerometry remains higher in b) than a) after the run transition (red arrow), presumably because of greater running capacity.
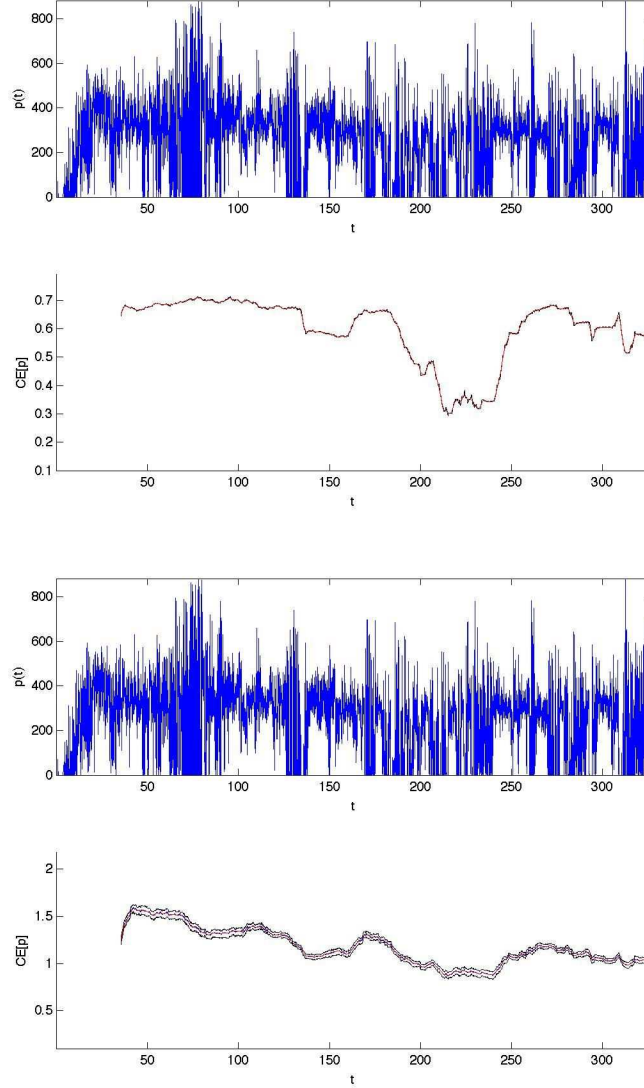
FIGURE 11. Power data from a professional bicycle rider on the open road, during a major European Grande-Tour, the 2001 Giro d'Italia. (Upper Row) Power $p(t)$ data set in watts is plotted versus time t in minutes. Apparently, and notably, there is a great deal more variability in the power of a cyclist in the professional peleton compared to a amateur in the laboratory setting from Fig-5-5. (Bottom-Left) CE computed by sign-partition method, (2). With essentially 2-symbols allow, and the highly variable and complex signal of the open road racing, a great deal of time yields saturation of the computed CE≈ ln 2, suggesting the need for a finer partition. (Bottom-Right) CE_8, using the SAX method and 8-symbols reveals a common scenario here , which is decreasing entropy with rider stress as the control signal becomes more constrained. Statistical error bands are shown, as explained in Sec-4.
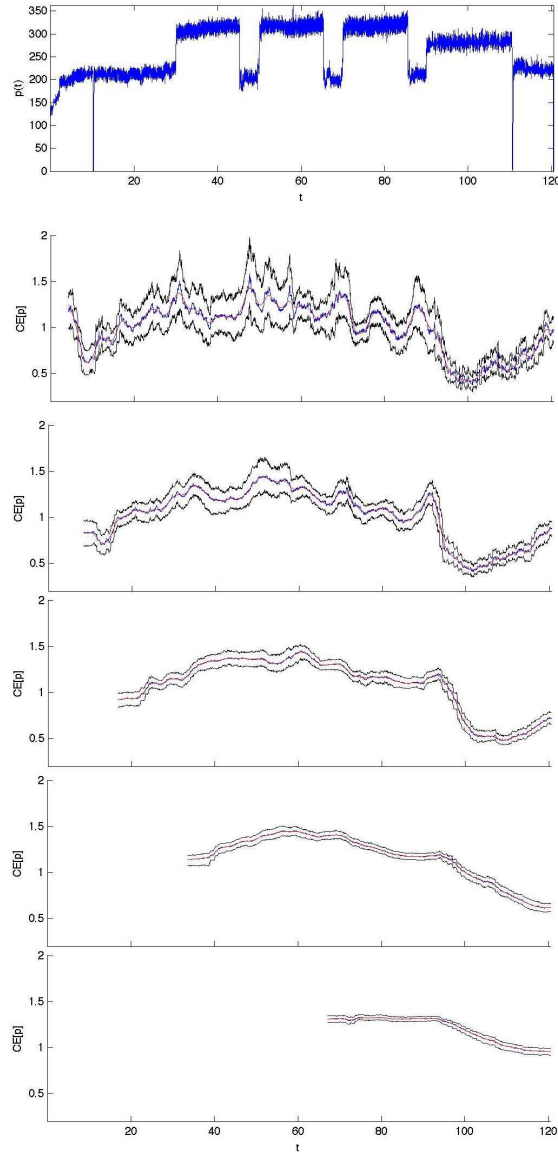
FIGURE 12. Window length study of CE of power during submaximal cycle ergometer trial. The subject performed warm up and cool down at 200 Watts, with three 15 min efforts at 315 Watts and 20 min at 280 Watts (50, 78 and 68% of VO2max, respectively). (Top) Power $p(t)$ in watts as a function of time t in minutes. (Successively from 2nd to Bottom) CE for increasing window lengths $w = 201, 401, 801, 1601, 3201$ data points, with $w = 201$ corresponding to a window duration of ≈ 4.2 minutes. Error bands (black) are as described in Sec-4. Observe that smaller windows are able to detect the high intensity intervals, but at the cost of greater statistical fluctuation. The longer windows provide a time-averaged interpretation that indicates the overall trend toward exhaustion without regard to changes in intensity of the workout.
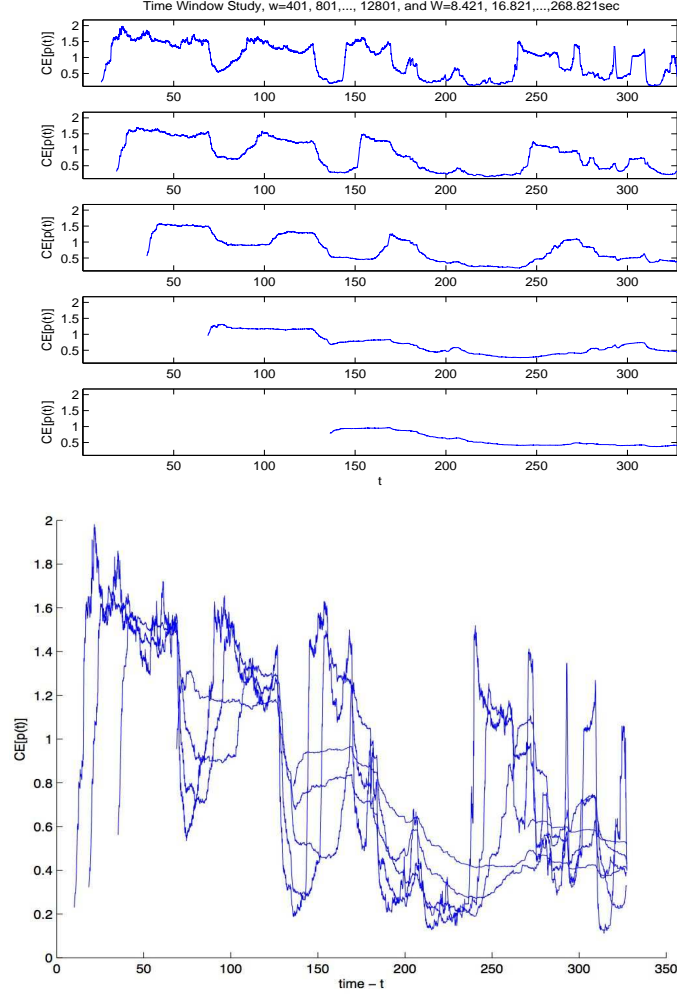
FIGURE 13. CE time-scale study of power data from a professional bicycle rider on the open road. (Left) $CE[j; w, \{p(t_i)\}]$, according to (13), for increasing time windows, from top, $w = 401, 801, ..., 12801$, for a CE filtering of data in windows of natural units, $W = 8.41sec, 16.821sec, ..., 268.821sec$. (Right) The same data plotted simultaneously in the same window. We see for small w variability in $CE[j; w, \{p(t_i)\}]$ due both to statistical variability of the data and corresponding convergence of the CE statistic, as well as true parameter drift. On the other hand, as w is increased, we see smoothing of the CE variability, to the point that we can see variations due to slow drift in parameters more clearly. However, at some point, when the window is longer than natural trends in parameter variations, there may be a loss in variability of CE computation due to averaging between several trends. By the second and third CE plots from the bottom-left, we see a general decline in CE as the ride progresses, overlaying some general oscillations which perhaps correspond to ebbs and waves of the race pace due to mountains and strategy.
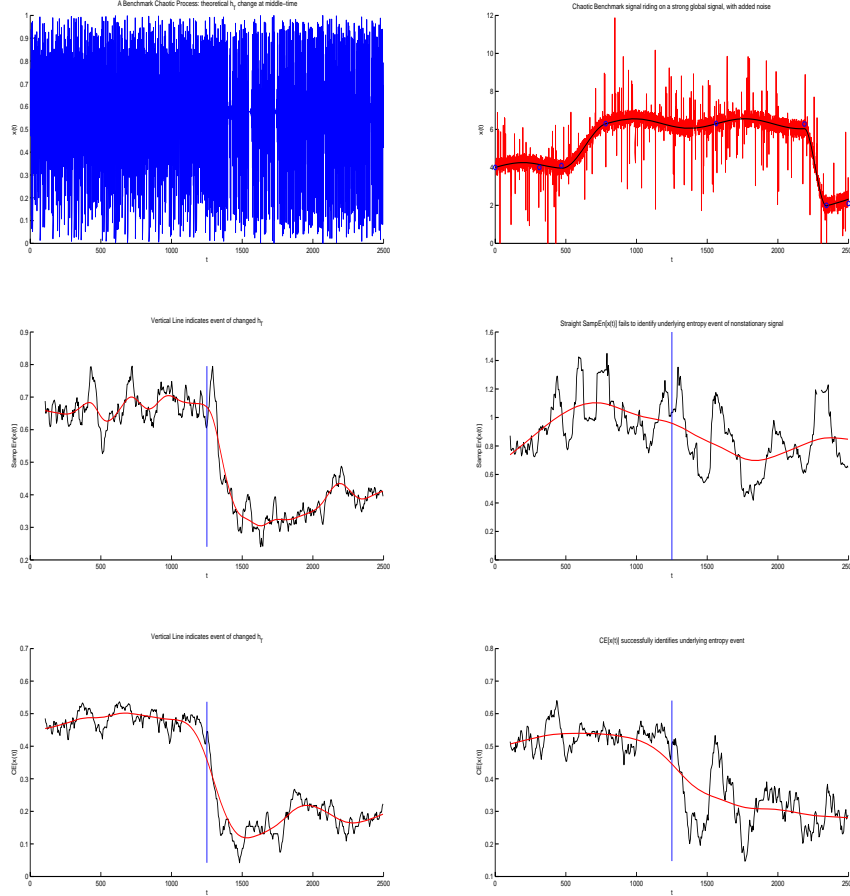
FIGURE 14. **CE** locates changes in entropy of nonstationary signals. (Top Left) A chaotic signal, determined by a chaotic tent map is designed with a change in from the larger entropy blue chaotic tent map to the smaller entropy red map as shown in Fig-4. (Mid Left) a moving window estimate of entropy of the time-series signal using SampEn method successfully detects the change in the driving process. (Lower Left) CE also successfully detects the change in parameter corresponding to changed topological entropy. (Top Right) The stationary signal $x(t)$ (shown Top Left) is added to riding signal, shown in blue through the middle, and the resultant signal is contaminated with measurement noise, which gives the nonstationary stochastic signal shown in red. (Mid Right) SampEn of this noisy signal; note that SampEn cannot "find" the parameter change in the nonstationary signal. (Lower Right) CE detects the change, despite the highly nonstationary signal.