Article

Specific Differential Entropy Rate Estimation for Continuous-Valued Time Series

David Darmon

Department of Military and Emergency Medicine, Uniformed Services University, Bethesda, MD 20814, USA; david.m.darmon@gmail.com

Abstract: We introduce a method for quantifying the inherent unpredictability of a continuous-valued time series via an extension of the differential Shannon entropy rate. Our extension, the specific entropy rate, quantifies the amount of predictive uncertainty associated with a *specific* state, rather than averaged over all states. We provide a data-driven approach for estimating the specific entropy rate of an observed time series. Finally, we consider three case studies of estimating the specific entropy rate from synthetic and physiological data relevant to the analysis of heart rate variability.

1. Introduction

The analysis of time series resulting from complex systems must often be performed "blind": in many cases, mechanistic or phenomenological models are not available because of the inherent difficulty in formulating accurate models for complex systems. In this case, a typical analysis might treat the data as the model, in the spirit of nonparametric statistics, and attempt to generalize from the available data to the system more generally. For example, a common question to ask about a times series is how "complex" it is, where we place complex in quotes to emphasize the lack of a satisfactory definition of complexity at present [1]. An answer is then sought that agrees with a particular intuition about what makes a system complex. For example, trajectories from periodic or entirely random systems appear simple, while trajectories from chaotic systems appear quite complicated. On a more practical level, it is common to compare two time series from similar systems, in which case one wants to meaningfully ask: is the phenomenon resulting from System A more or less complex than the phenomenon resulting from System B?

There are many possible definitions of the complexity of a time series. See [1,2] for comprehensive reviews. Some notable attempts at formal definitions include Kolmogorov complexity [3], stochastic complexity [4], forecast complexity [5], and Grassberger–Crutchfield–Young statistical complexity [6]. Perhaps the most well-developed theory of complexity, which incorporates and expands on many of these quantities in the special case of discrete-valued time series, is computational mechanics [7]. For example, see [8] for an elucidation of the amount of information, in a formal sense, stored in a single observation from a discrete-valued stochastic process.

Practical definitions of complexity for continuous-valued time series are much less well-developed. The most common definitions rely on some notion of the difficulty in predicting a time series. There are currently at least two schools of thought for the (un)predictability-based notions of complexity when applied to systems with continuous states: Kolmogorov–Sinai entropy [9,10] and the Shannon entropy rate [11]. Approaches based on the former treat the data as a trajectory from a deterministic dynamical system and seek to estimate the Kolmogorov–Sinai entropy based on the trajectory [12]. This school of thought goes back to some of the earliest work applying nonlinear dynamics to observational data [13].

Approaches based on the latter treat the data as a realization from a stochastic process and focus on the entropy rate from a statistical perspective [14]. While these approaches *seem* very similar and are typically treated as such in much of the applied literature, they in fact give diverging answers to similar questions. In particular, the Kolmogorov–Sinai entropy of a stochastic dynamical system is infinite, while the differential Shannon entropy rate of a deterministic dynamical system is infinite [15]. These facts have been noted in some of the earliest work on estimating entropy rates from continuous-valued time series [16], but are largely ignored in the applied literature. Moreover, methods proposed to estimate Kolmogorov–Sinai entropy may in fact be estimating the Shannon entropy rate, and *vice versa*. The situation may be further confused by the fact that the Kolmogorov–Sinai entropy *does* correspond to a Shannon entropy rate, in this case the supremum over the *discrete* Shannon entropy rates induced by finite partitions of the state space of a dynamical system [17].

In addition to the methodological divide between the two dominant approaches to entropy rate estimation, neither has been used to provide a *specific* entropy rate for the system as a function of its state. That is, estimates are typically reported as time averages, which under certain conditions, converge to state space averages. However, it may be desired to know the entropy rate associated with a system *now*, at the present state, rather than on average. It is difficult to define such a state-specific entropy rate in the Kolmogorov–Sinai framework. For stochastic dynamics, such a state-specific entropy rate can be defined over *ensembles* of the system starting at the specified state. Thus, one of the aims of this paper is to provide an estimator for such a *specific* entropy rate.

The contributions of this paper are three-fold. First, we reemphasize the dependence of the short-term predictability of a nonlinear dynamical system on its current state and propose an information theoretic quantity, the specific entropy rate, that captures this dependence. Second, we propose a statistically-principled approach to estimating the specific entropy rate from a continuous-valued time series that takes advantage of recent advances in conditional density estimation. Finally, we demonstrate the new approach with both synthetic and real data to highlight its strengths and weaknesses, with a special emphasis on inter-event interval data as found in heart rate variability analysis. Throughout, we also make connections to modern practices in entropy rate estimation, both of the Kolmogorov–Sinai and differential schools, and seek to highlight how our estimator fits into those frameworks.

2. Methodology

In the following sections, we define the specific entropy rate of a stochastic dynamical system and develop an approach for its estimation from data. In Section 2.1, we fix our notation and define a stochastic dynamical system. In Section 2.2, we review the entropy rate of a stochastic dynamical system and define the specific entropy rate. In Sections 2.3 and 2.4, we propose a method for estimating the specific entropy rate from finite data. Finally, in Section 2.5, we make connections between the specific entropy rate and other commonly-used entropy rate estimators.

2.1. Stochastic Dynamical System

Consider an observed scalar real-valued time series $x_1, x_2, \ldots, x_T$. We explicitly model the time series as a realization from an autonomous stochastic dynamical system [18,19]. That is, unlike for autonomous deterministic dynamical systems, which assume that a deterministic update rule acts on the precisely-known state of the system, we assume that the states are stochastic and, moreover, that transitions from state to state occur according to a transition density. Thus, we view $x_1, x_2, \ldots, x_T$, as a realization from the system $\{X_t\}_{t \in \mathbb{Z}}$, where we use the standard convention of using upper/lower case to denote a random variable/its realization. For $n > m$, let $X_m^n = (X_m, X_{m+1}, \ldots, X_{n-1}, X_n)$ denote the $n - m + 1$ block of states for the dynamical system from time m to time n. Similarly, let $X_{-\infty}^m = (\ldots, X_{m-1}, X_m)$ denote the semi-infinite past until time m and let $X_n^\infty = (X_n, X_{n+1}, \ldots)$ denote

the semi-infinite future starting at time n. Then, a general model [19] for how the state evolves assumes that the future state X_t can be expressed as a random transformation of its past $X_{-\infty}^{t-1}$,

$$X_t = F\left(X_{-\infty}^{t-1}; \epsilon_t\right) \tag{1}$$

where ϵ_t represents *dynamical* noise, that is noise that influences the *dynamics* of the system, to be contrasted with *observational* noise, which impacts the observations of the system, but not its dynamics. Equivalently, Equation (1) can be expressed explicitly in terms of the transition density $f\left(x \mid x_{-\infty}^{t-1}\right)$ as:

$$X_t \sim f\left(x \mid X_{-\infty}^{t-1}\right). \tag{2}$$

More typically, the dynamical noise is taken to be additive, in which case:

$$X_t = G(X_{-\infty}^{t-1}) + \epsilon_t \tag{3}$$

where typically, $\{\epsilon_t\}$ is taken to be independent and identically distributed and ϵ_t is taken to be independent of previous values of $X_s, s < t$. Finally, we note that we consider solely *scalar* time series in this paper. While much of the theory can be translated to the case of multivariate time series by replacing the scalar observable X_t with a d-dimensional vector observable $\mathbf{X}_t$, the impact of this change on the computational and statistical burdens of an approach such as the one we develop here is less easily overcome.

2.2. Differential Entropy Rate and Its Estimation

Let $\{X_t\}_{t \in \mathbb{Z}}$ be a discrete-time, continuous-state stochastic dynamical system as defined in the previous section. Recall that for a continuous-valued random variable X with density $f(x)$, the differential entropy [20] of X is given by:

$$h[X] = -E[\log f(X)] \tag{4}$$

$$= -\int_{\mathbb{R}} f(x) \log f(x) \, dx. \tag{5}$$

We will always take the logarithm with base e, and thus, all differential entropies are in nats. For the remainder of this paper, because our focus is on continuous-state systems, when we use the term entropy, we refer to differential entropy. For random variables (X, Y) with joint density $f(x, y)$, the joint entropy of X and Y is defined similarly as:

$$h[X, Y] = -E[\log f(X, Y)] \tag{6}$$

$$= -\int_{\mathbb{R}^2} f(x, y) \log f(x, y) \, dx \, dy. \tag{7}$$

Applying Equation (7) to a stochastic dynamical system $\{X_t\}_{t \in \mathbb{Z}}$ with a block-p joint distribution f_t at time t, the block-p entropies at time t are given by:

$$h\left[X_t^{t+p-1}\right] = h[X_t, \ldots, X_{t+p-1}] = -E[\log f_t(X_t, \ldots, X_{t+p-1})]. \tag{8}$$

There are two definitions of the differential entropy rate that are equivalent for a strong-sense stationary stochastic process [21,22]. The first, which we denote as $\bar{h}_1(X)$, defines the entropy rate in terms of the rate of growth of block-p entropies,

$$\bar{h}_1(X) = \lim_{t \to \infty} \frac{h[X_1, \ldots, X_t]}{t}. \tag{9}$$

The second, which we denote as $\bar{h}_2(X)$, defines the entropy rate in terms of the entropy of a one-step-ahead future conditional on a sufficiently long past,

$$\bar{h}_2(X) = \lim_{t \to \infty} h\left[X_{t+1} \mid X_1^t\right]. \tag{10}$$

While these are equivalent for strictly stationary stochastic processes, they need not be for an arbitrary process. Because we are interested in quantifying the *predictability* of a stochastic process over time, we take Equation (10) as our definition of the entropy rate, $\bar{h}(X) \equiv \bar{h}_2(X)$.

Clearly, care must be taken when interpreting the densities that appear in the definitions of the entropies and entropy rates we have defined thus far, and this interpretation depends on the assumptions that the practitioner is willing or able to make about the system under consideration. In practice, the assumption is typically made that $\{X_t\}_{t \in \mathbb{Z}}$ is strong-sense stationary [23] or at least can be made so via transformations, such as differencing or detrending. These assumptions are typically violated in practice. We make a less restrictive assumption on the system under consideration, namely that it is *conditionally stationary* [24]. A process is conditionally stationary if the conditional distribution function of X_{t+1} given $(X_t, \ldots, X_{t-p+1}) = \mathbf{x}$ does not depend on t for some fixed p: that is, the statistical future of the process conditional on a past of sufficient length does not depend on *when* that past was observed. Strong-sense stationary processes and Markov processes are special cases of this type.

The value of $\bar{h}(X)$ depends on $h\left[X_{t+1} \mid X_1^t\right]$ and, thus, on the conditional structure of the stochastic process. Consider the conditional entropy of X_t given the block X_{t-p}^{t-1} of length p. Under the assumption of the conditional stationarity of order p, this conditional entropy can be rewritten as:

$$h\left[X_t \mid X_{t-p}^{t-1}\right] = -E[\log f_t(X_t \mid X_{t-p}^{t-1})] \tag{11}$$

$$= -\int_{\mathbb{R}^{p+1}} f_t(x_1^{p+1}) \log f_t(x_{p+1} \mid x_1^p)\, dx_{p+1}\, dx_1^p \tag{12}$$

$$= -\int_{\mathbb{R}^{p+1}} f_t(x_1^p) f_t(x_{p+1} \mid x_1^p) \log f_t(x_{p+1} \mid x_1^p)\, dx_{p+1}\, dx_1^p \tag{13}$$

$$= -\int_{\mathbb{R}^{p+1}} f_t(x_1^p) f(x_{p+1} \mid x_1^p) \log f(x_{p+1} \mid x_1^p)\, dx_{p+1}\, dx_1^p \tag{14}$$

$$= -\int_{\mathbb{R}^p} f_t(x_1^p) E\left[\log f(X_t \mid X_{t-p}^{t-1}) \mid X_{t-p}^{t-1} = x_1^p\right] dx_1^p \tag{15}$$

$$= -E\left[E\left[\log f(X_t \mid X_{t-p}^{t-1}) \mid X_{t-p}^{t-1}\right]\right]. \tag{16}$$

where going from Equations (13) to (14), we have applied conditional stationarity. Thus, we see that the order p conditional entropy depends on two properties of the stochastic process: the state-specific entropy conditional on a particular past x_1^p and the overall density of the pasts X_1^p. This decomposition motivates defining the state-specific entropy rate of order p at time t as:

$$h_t^{(p)} \equiv h[X_t \mid X_{t-p}^{t-1} = x_{t-p}^{t-1}] \tag{17}$$

$$= -E\left[\log f(X_t \mid X_{t-p}^{t-1}) \mid X_{t-p}^{t-1} = x_{t-p}^{t-1}\right] \tag{18}$$

$$= -\int_{\mathbb{R}} f(x_{p+1} \mid x_1^p) \log f(x_{p+1} \mid x_1^p)\, dx_{p+1}. \tag{19}$$

We will call $h_t^{(p)}$ the *specific entropy rate* of order p or simply the *specific entropy rate* where the order p is clear. We will specify a procedure for choosing p in Section 2.4. The specific entropy rate quantifies the unpredictability of the process conditional on the specific past x_{t-p}^{t-1} observed immediately before time t. We see that Equation (19) emphasizes the well-known fact that the *difficulty in prediction* can depend on the current state for both deterministic and stochastic nonlinear dynamics [25,26]. This is *not* the case for linear time series models, where the specific entropy rate is independent of the present

state of the system. We note that our specific entropy rate is similar in spirit to the specific information of a stimulus [27] from computational neuroscience, local information measures from [28,29], and the Lyapunov-like index [25] from statistical nonlinear time series analysis. The specific information of a stimulus notes that the mutual information between two random variables R and S can be decomposed as $I[R \wedge S] = H[R] - H[R \mid S]$ where H denotes the *discrete* Shannon entropy. Thus, the specific information of a particular stimulus s for a response R is taken to be $I[R \wedge s] = H[R] - H[R \mid S = s]$, using a similar decomposition as Equation (16). The local information measures go one further step back, defining the local information measures in terms of the argument of the expectation associated with the information measure. For example, the local entropy rate of order p at x_1^{p+1} under this formalism is defined as $-\log f(x_{p+1} \mid x_1^p)$, rather than as $-E[\log f(X_{p+1} \mid X_1^p) \mid X_1^p = x_1^p]$ in our definition. The Lyapunov-like index is defined in terms of divergences with respect to the past of conditional expectations of the future given the past and, thus, measures uncertainty about the future given the past using solely the first moment of the predictive density.

In practice, the predictive density $f(x_{p+1} \mid x_1^p)$ is unknown and must be inferred from observations of the system under consideration. Thus, we consider the plug-in estimator for the specific entropy rate, namely:

$$\hat{h}_t^{(p)} \equiv -E\left[\log \hat{f}(X_t \mid X_{t-p}^{t-1}) \mid X_{t-p}^{t-1} = x_{t-p}^{t-1}\right] \tag{20}$$

where we substitute an estimator $\hat{f}(x_{p+1} \mid x_1^p)$ for the unknown predictive density $f(x_{p+1} \mid x_1^p)$. Finally, if an estimator for the overall entropy rate (10) of the system is desired, we define the estimator:

$$\hat{\bar{h}}^{(p)} = \frac{1}{T-p} \sum_{t=p+1}^{T} -E\left[\log \hat{f}(X_t \mid X_{t-p}^{t-1}) \mid X_{t-p}^{t-1} = X_{t-p}^{t-1}\right] \tag{21}$$

$$= \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{h}_t^{(p)}, \tag{22}$$

a time-average of the specific entropy rates, using the empirical distribution over the pasts as an estimator for $f_t(x_1^p)$ in Equation (15).

Before considering the problem of estimating the predictive density $\hat{f}(x_{p+1} \mid x_1^p)$, we note that we are really interested in the specific entropy of the predictive density and not the predictive density outright. Thus, the predictive density $f(x_{p+1} \mid x_1^p)$ is a nuisance parameter and a difficult one to estimate, especially in higher dimensions. Based on this insight, many information theoretic estimators has been proposed that directly estimate the quantity of interest without first estimating the underlying density. For example, many estimators have been proposed based on the statistics of k-nearest neighbors amongst the sample points [30–35]. In fact, many of these estimators correspond to plug-in estimators using *variable* bandwidth kernel density estimators [36], with the bandwidth varying with the evaluation point: the bandwidth is taken to be the distance to the k-th nearest neighbor. A key aspect of our estimator, which we turn to in Section 2.4, is the use of model selection to directly learn which lags are relevant to prediction. A similar approach could be taken with the k-th nearest neighbor-based estimators, letting k vary with each lag. We return to a discussion of this approach and its relation to our method in Section 4.

2.3. Conditional Density Estimation

The problem of estimating a conditional density goes back to the pioneering work of Rosenblatt [37]. We estimate the predictive density using the conditional kernel density estimator proposed in [38,39]. See [40] for additional theoretical results for density estimators for general

stochastic processes. Consider a continuous-valued time series $\{X_t\}_{t=1}^{T}$ for which we desire to estimate the predictive density $f(x_{p+1} \mid x_1^p)$. Recalling that the predictive density is given by:

$$f(x_{p+1} \mid x_1^p) = \frac{f(x_1^p, x_{p+1})}{f(x_1^p)}, \tag{23}$$

we can estimate the predictive density by estimating the joint density $f(x_1^p, x_{p+1})$ and the marginal density $f(x_1^p)$ and taking their ratio. We estimate the marginal and joint densities using the kernel density estimators:

$$\hat{f}(x_1^p) = \frac{1}{T-p} \sum_{t=p+1}^{T} K_{\mathbf{k}}(x_1^p, X_{t-p}^{t-1}) \tag{24}$$

and:

$$\hat{f}(x_1^p, x_{p+1}) = \frac{1}{T-p} \sum_{t=p+1}^{T} K_{\mathbf{k}}(x_1^p, X_{t-p}^{t-1}) L_{k_{p+1}}(x^{p+1}, X_t), \tag{25}$$

respectively, where $K_{\mathbf{k}}$ is the product kernel:

$$K_{\mathbf{k}}(x_1^p, X_{t-p}^{t-1}) = \prod_{j=1}^{p} \frac{1}{k_j} K\left(\frac{x_j - X_{t-p+j-1}}{k_j}\right), \tag{26}$$

$L_{k_{p+1}}$ is the univariate kernel:

$$L_{k_{p+1}}(x_{p+1}, X_t) = \frac{1}{k_{p+1}} K\left(\frac{x_{p+1} - X_t}{k_{p+1}}\right), \tag{27}$$

$k_1, \ldots, k_{p+1}$ are the bandwidths and $K(\cdot)$ is a kernel function, *i.e.*, a positive, symmetric probability density with finite second moment. The estimator for the conditional density $\hat{f}(x_{p+1} \mid x_1^p)$ is then:

$$\hat{f}(x_{p+1} \mid x_1^p) = \frac{\hat{f}(x_1^p, x_{p+1})}{\hat{f}(x_1^p)}. \tag{28}$$

Note that the joint and marginal density estimators are *coupled* since they use the same bandwidths $k_1, \ldots, k_p$ for both the marginal and joint density estimators. This coupling is necessary to ensure that, for example, the conditional density integrates to one with respect to x_{p+1}. On a more practical level for time series, this coupling allows us to screen out the distant past. Consider, for example, the extreme case where the past is irrelevant to the future in terms of prediction. By this coupling, we can ignore the past by setting the bandwidths $k_1, \ldots, k_p$ to large values. This has the effect of giving $\hat{f}(x_{p+1} \mid x_1^p) \approx \hat{f}(x_{p+1})$ and recovering the appropriate independence relationship. More generally, if $q < p$ lags are sufficient to screen off the distant past, then by setting the bandwidths $k_1, \ldots, k_{p-q}$ sufficiently large, we can recover $\hat{f}(x_{p+1} \mid x_1^p) \approx \hat{f}(x_{p+1} \mid x_{p-q+1}^p)$. We discuss how to take advantage of this property of conditional kernel density estimators in more detail in the next section.

2.4. Bandwidth and Order Selection

The estimator of the conditional density function (28) and, thus, the estimator of the specific entropy rate (20) depend on the choice of the order p and bandwidths $k_1, \ldots, k_{p+1}$. We therefore require a principled and repeatable procedure for selecting them. For example, in the context of transfer entropy estimation, [41] noted how, depending on the choice of these parameters, the direction of causality can be reversed. Because our approach explicitly builds a *statistical* model for the dynamical system, we

choose the order and bandwidths via *l*-block cross-validation [42] of the negative log-likelihood of the conditional density (note that [42] calls their method *h*-block cross-validation, which we rename in this manuscript to avoid confusion with differential entropy). *l*-block cross-validation is an extension of leave-one-out cross-validation where instead of leaving out a *single* observation at each evaluation, we remove the observation and *l* observations on either side of that observation. That is, we seek the values of p and $\mathbf{k} = (k_1, \ldots, k_{p+1})$ that minimize:

$$\mathrm{CV}_l(p, \mathbf{k}) = -\frac{1}{T-p} \sum_{t=p+1}^{T} \log \hat{f}_{-t:l}(X_t \mid X_{t-p}^{t-1}), \tag{29}$$

where $\hat{f}_{-t:l}$ is the estimate of conditional density after removing the $2l+1$ observations about t. This accounts for a bias in zero-block cross-validated likelihood resulting from the dependence inherent in temporally nearby realizations of a time series. We immediately see that Equation (29) takes the form of an entropy rate, so this cross-validation procedure can also be thought of as minimizing the entropy rate of the model. Thus, cross-validation provides a principled means for choosing the order of the entropy rate in analogy to common practices in the discrete-valued case. For example, when computing the entropy rate for discrete-valued data, it is frequently recommended to choose the order of the entropy rate by searching for an asymptotic value for the order-p entropy rate as a function of p [43]. Thus, our approach extends this heuristic to the continuous-valued case, with an additional penalty on p induced by the nature of cross-validation. Moreover, both theoretical and empirical work have shown that choosing the bandwidth via cross-validation can automatically "smooth out" irrelevant predictors by setting their bandwidths very large [38,44]. This is clearly desirable in the time series case, since we expect to induce conditional independence between the distant past and the future after accounting for a sufficient portion of the recent past. By using cross-validation, we get this dimension reduction for free.

Because of the computationally-intensive nature of *l*-block cross-validation, we begin by fixing p and choosing the bandwidths $(k_1, \ldots, k_{p+1})$ using zero-block cross-validation, which reduces to leave-one-out-cross-validation. Then, using these bandwidths, we choose p via *l*-block cross-validation. In all of the reported results, we use $l = 50$, thus leaving out 101 points about any evaluation in Equation (29). In principle, the block size could be chosen using the autocorrelation time or lagged mutual information [12] or a data-driven approach [45]. We leave the exploration of these approaches for future work.

2.5. Relationship to Other Entropy Rate Estimators

In the nonlinear dynamics community, especially in applications to biological systems, two popular measures of the uncertainty associated with the dynamics of a system are approximate entropy [46] and sample entropy [47]. Despite their names, both of these quantities correspond to estimators of entropy *rates* rather than entropies. Approximate entropy, as originally proposed by Pincus, was motivated by a finite-time, finite-resolution approximation to the Kolmogorov–Sinai entropy of a deterministic dynamical system. The sample entropy was proposed as a modification to the approximate entropy that addressed several of its deficiencies. In [14], Lake elucidates the key connection between the approximate and sample entropies and information theoretic entropy rates. In particular, Lake shows that the approximate entropy corresponds to a kernel density-based estimator of the Shannon differential entropy rate using uniform kernels and fixed bandwidths $k_1 = k_2 = \ldots = k_{p+1}$, while the sample entropy corresponds to a kernel density-based estimator of the Rényi entropy rate with order $\alpha = 2$, the so-called collision entropy, with a particular choice of definition for the conditional Rényi entropy (unlike conditional Shannon entropy, no standard definition of conditional Rényi entropy exists for arbitrary α [48]). In later work, recommendations were made for choosing the model order p [49], for setting the common bandwidth [50], and for incorporating an adaptive bandwidth [51]. In the Appendix to this paper, we reproduce the derivation

made in [14] connecting the approximate entropy statistic to kernel density-based estimators of the differential entropy rate.

3. Results

We consider entropy rate estimation in three examples of increasing realism. The first example, described in Section 3.1, applies the specific entropy rate estimator to a second-order Markov model. This example was designed to emphasize, in a simple way, the potential dependence of the specific entropy rate h_t on the state of the system. In Section 3.2, we consider the entropy rate of inter-event intervals resulting from an integrate-and-fire model driven by synthetic chaotic signals. This type of model is typically implicit in many of the analyses of biological signals ranging from heart rate variability to neural firing. This example demonstrates how our entropy rate estimator performs when the assumption of a *stochastic* dynamical system is violated. Finally, in Section 3.3, we demonstrate the specific entropy rate estimator using interbeat interval sequences resulting from a tilt table experiment.

Throughout these examples, we use the R package `np` [39] to estimate the conditional densities using second-order Gaussian kernels [52]. As recommended in the Methodology section, for a particular model order p, we choose the bandwidths using leave-one-out cross-validation on the log-likelihood and choose the model order p using l-block cross-validation with $l = 50$. We then estimate the specific entropy rate using Equation (20).

3.1. A Second-Order Markov Process

Our first example is chosen to highlight the state-dependent nature of the specific entropy rate Equation (19). We consider a stochastic dynamical system with three effective states. One of the states corresponds to a *crossing event*, when the system switches from positive to negative outputs or *vice versa*. This state has a high specific entropy rate. The other two states correspond to when the system settles into either a run of positive outputs or a run of negative outputs. In these states, the specific entropy rate is smaller. Explicitly, consider the second-order Markov process with the transition density:

$$f(x_t \mid x_{t-2}, x_{t-1}) = \begin{cases} p_+ \phi(x_t; 5, 1) + (1 - p_+) \phi(x_t; -5, 1) & : x_{t-2}, x_{t-1} > 0 \\ p_- \phi(x_t; -5, 1) + (1 - p_-) \phi(x_t; 5, 1) & : x_{t-2}, x_{t-1} < 0 \\ \phi(x_t; 0, 3^2) & : \text{otherwise} \end{cases} \tag{30}$$

where $p_+ = p_- = 0.9$ and $\phi(x; \mu, \sigma^2)$ is the probability density function for a normal random variable with mean μ and variance σ^2,

$$\phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}. \tag{31}$$

The transition densities for each effective state are shown in the left panel of Figure 1. The first effective state (red solid) corresponds to when the two previous observations were positive; the second effective state (blue dashed) corresponds to when the two previous observations were negative; and the third effective state (green dash-dotted) corresponds to when the two previous observations had opposite signs. The right panel of Figure 1 shows a scatter plot representation of the marginal density (X_t, X_{t+1}) with the quadrants colored by the effective states.
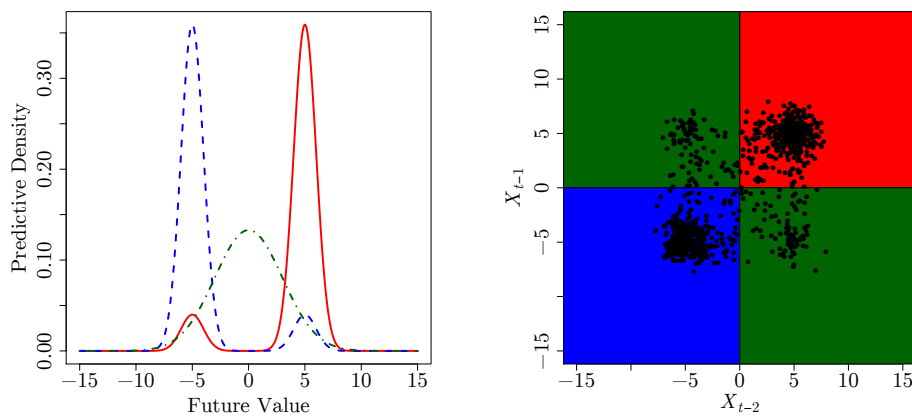
Figure 1. (**Left**) The predictive densities associated with each of the effective states for the Markov process (30); (**Right**) a scatter plot representation of the marginal density of (X_{t-2}, X_{t-1}) with the effective states colored according to the convention in the left panel.

The top panel of Figure 2 shows an example realization with $T = 1000$, which we use to estimate the specific entropy rate. We can compute the specific entropy rate h_t for each effective state exactly. By symmetry, the first two effective states have the same specific entropy rate, which we compute by evaluating Equation (19) numerically: 1.744 nats per symbol. The third effective state's predictive density corresponds to a normal density with a variance of nine and, thus, has specific entropy rate $\frac{1}{2} \log(2\pi e \cdot 3^2) \approx 2.518$ nats per symbol. The bottom panel shows the specific entropy rate (dashed blue), along with the estimated specific entropy rate with $p = 2$ (solid red). From the specific entropy rate, we can clearly see when the system switches from one of the low specific entropy rate states to the high specific entropy rate state and *vice versa*. Moreover, we see that the estimated entropy also displays these transitions, though not as cleanly.
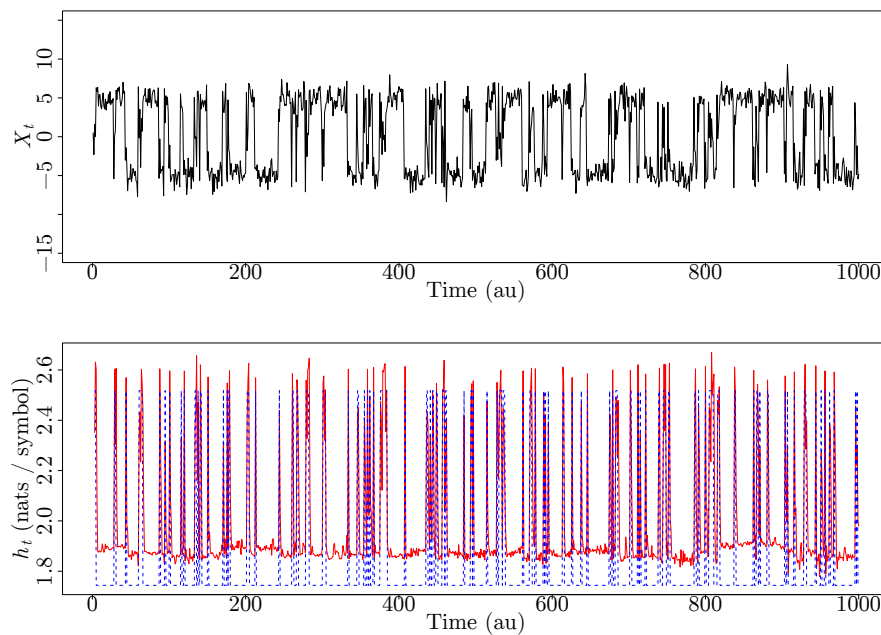

Figure 2. An example realization from Equation (30) (top), along with the specific entropy rate (bottom). The dashed blue line indicates the true specific entropy rate, while the solid red line indicates the entropy rate estimated using Equation (20).

To see the performance of the estimator as a function of the history, for each time point t, we compute both the estimator of the specific entropy rate $\hat{h}_t$, as well as the empirical bias between the estimated and true value,

$$\text{Bias}\left(\hat{h}_t\right) = \hat{h}_t - h_t. \tag{32}$$

Figure 3 displays the estimated specific entropy rate (left) and bias (right) as a function of the history (x_{t-2}, x_{t-1}). As we saw in Figure 2, the estimator successfully distinguishes between the high entropy rate effective state (colored purple) and the low entropy rate effective states (colored yellow). Because the estimated specific entropy rate is always positive for this system, a positive bias indicates that the estimated entropy rate is *larger* (greater predictive uncertainty) than it should be, and a negative bias indicates that the entropy rate is *smaller* (lower predictive uncertainty) than it should be. We see that a large positive bias occurs for those pasts that belong to either the first (red) or second (blue) effective states, but lie near the border with the third (green) effective state. This occurs because of the discontinuous transition in the predictive density between each state. It is especially pronounced for those (rare) pasts near the origin, again because of the discontinuity.
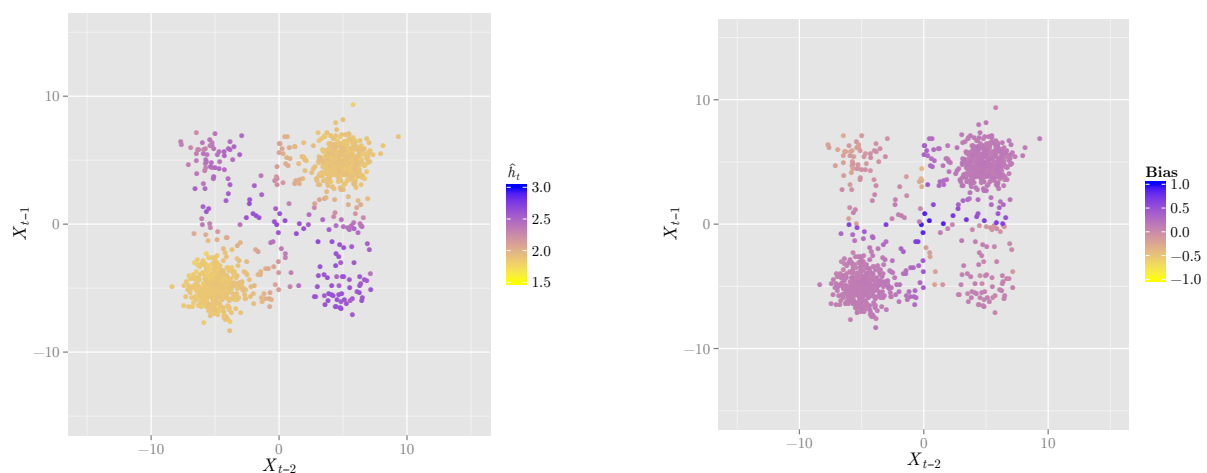

Figure 3. The estimated specific entropy rate $\hat{h}_t$ (**left**) and its bias $\hat{h}_t - h_t$ (**right**) as a function of the history (X_{t-2}, X_{t-1}) for the Markov model. Note that the estimator correctly identifies the high and low specific entropy rate histories, and its largest bias occurs near the transitions between quadrants.

Finally, we demonstrate two snap shots of the system in Figure 4 to recall the intuition behind the specific entropy rate and how it relates to the predictive density of the stochastic dynamical system. Each panel shows the state of the system (top) with the present state x_t marked by a blue circle and the past (x_{t-2}, x_{t-1}) marked by red circles, the estimated predictive density $\hat{f}(\cdot \mid x_{t-2}, x_{t-1})$ (middle) and the estimated specific entropy rate (bottom). The left panel corresponds to when the two past observations were positive, and thus, the system is in one of the low entropy rate effective states. The right panel corresponds to when the two past observations were opposite in sign, and thus, the system is in the high entropy rate effective state. We see that in both cases, the estimated predictive densities and estimated entropy rates agree with the effective states.
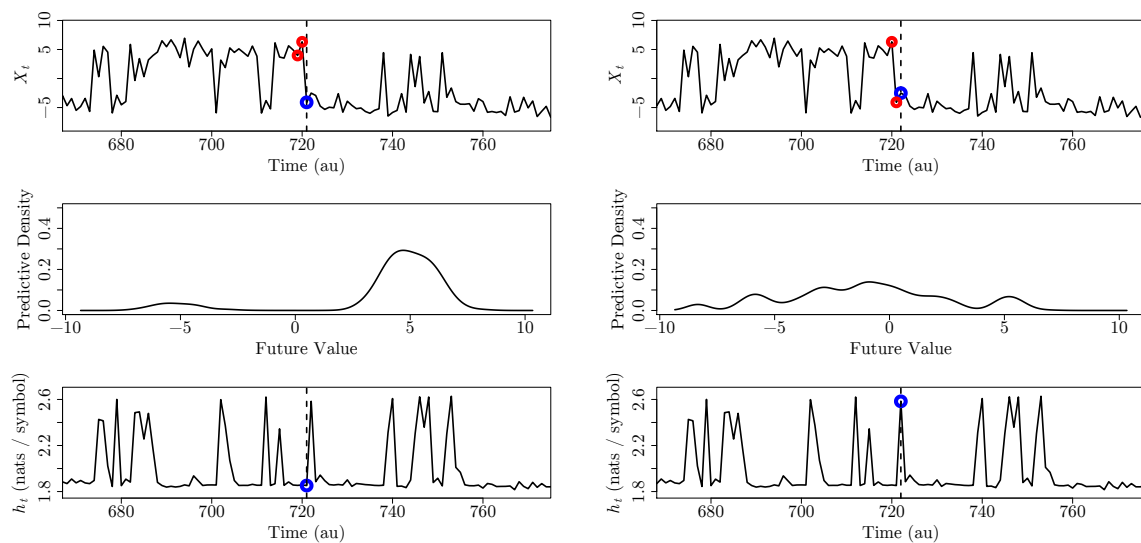
Figure 4. A demonstration of two adjacent time points of (top) a realization from the second order Markov model, (middle) the estimated predictive density $\hat{f}(x_t \mid x_{t-2}, x_{t-1})$ and (bottom) the specific entropy rate for the second-order Markov process in low (left) and high (right) specific entropy rate states. In the top panels, the dashed vertical bar indicates the time t; the red points correspond to the specific pasts (x_{t-2}, x_{t-1}); and the blue points correspond to the future values x_t.

3.2. Inter-Event Intervals from an Integrate-And-Fire Model Driven by Chaotic Signals

For our second example, we consider inter-event intervals resulting from an integrate-and-fire model driven by a chaotic signal. This model implicitly motivates many of the embedding-based analyses used with neural and heart rate variability data. For example, it is common to consider the times between heart beats (called interbeat intervals or RR intervals because the interbeat intervals are taken between consecutive R waves on the electrocardiogram) as if they are equispaced samples from a continuous time process, and then apply methods from nonlinear dynamics. There is not, *a priori*, any reason to assume that such an analysis of inter-event interval data through this "wrong" lens (e.g., treating the inter-event times from a point process as the output from a map) should give rise to meaningful results. However, a surprising result by Sauer [53] demonstrates at least one scenario where this type of analysis *does* give rise to meaningful results. In particular, Sauer demonstrated that when the state of a chaotic dynamical system is mapped into an inter-event interval sequence via an integrate-and-fire model, a one-to-one mapping exists between the full, unobserved state of the system and an embedding of the inter-event interval sequence as long as the embedding is of a dimension at least twice the box counting dimension of the underlying chaotic system. Thus, it is possible to recover the true state of the entire system by considering sufficiently long inter-event interval sequences.

This fact poses a problem for the analysis of inter-event interval data using quantities such as approximate entropy or sample entropy, since as we have noted, those can be seen as estimators of differential entropy rates and differential entropy rates of deterministic dynamical system are infinite. Thus, the quantity being used is at least potentially misspecified for the phenomenon being studied. Nevertheless, it seems unlikely that the popularity of approximate entropy or sample entropy will abate in the near future [54], and thus, it is interesting to consider how a more principled entropy rate estimator performs in the misspecified case. Moreover, in practice, the deterministic dynamical system model is almost certainly misspecified for complex systems. As noted in [16], there is hope that observational and dynamical noise might smooth out the infinities, thus resulting in useful estimates of entropy rates.

Consider a non-negative signal $S(t) = g(\mathbf{x}(t))$ mapping the m-dimensional state $\mathbf{x}(t) \in \mathbb{R}^m$ of a chaotic dynamical system to a scalar value. The integrate-and-fire model generates a series of discrete events based on when the integrated signal crosses a fixed threshold Θ. Setting $T_0 = 0$, for a fixed threshold value Θ, the threshold crossing events $\{T_i\}$ are defined recursively as:

$$\int_{T_i}^{T_{i+1}} S(t)\, dt = \Theta \tag{33}$$

and the inter-event intervals are given by the time between event $i - 1$ and i, $\text{IEI}_i = T_i - T_{i-1}$.

We consider signals generated by two classic chaotic systems, the Lorenz system evolving according to:

$$\begin{aligned}
\dot{x} &= \sigma(y - x) \\
\dot{y} &= x(\rho - z) - y \\
\dot{z} &= xy - \beta z
\end{aligned} \tag{34}$$

with the canonical values $\sigma = 10, \beta = 8/3$ and $\rho = 28$, and the Rössler system evolving according to:

$$\begin{aligned}
\dot{x} &= -y - z \\
\dot{y} &= x + ay \\
\dot{z} &= b + z(x - c)
\end{aligned} \tag{35}$$

with the canonical values of $a = 0.1, b = 0.1$ and $c = 14$. For both the Lorenz and Rössler systems, following [53], we take the signal to be:

$$S(t) = (x(t) + 2)^2 \tag{36}$$

and fix $\Theta = 60$ and $\Theta = 125$, respectively.

Figure 5 demonstrates example realizations of the inter-event intervals $\text{IEI}_i = T_i - T_{i-1}$ by event index i (left), as well as a lag-lag plot of consecutive inter-event intervals (right) for the Lorenz (top) and Rössler (bottom) systems. We see that the two systems give rise to very different time courses of inter-event intervals, as we would expect from differing dynamics of the two systems. In particular, since both the x- and y-coordinates of the Rössler system evolve in a nearly-linear fashion, we see that the inter-event intervals are relatively regular. By comparison, the inter-event intervals for the Lorenz system are much more erratic. Thus, we might intuitively expect the inter-event intervals from the Lorenz system to give higher specific entropy rates than the inter-event intervals from the Rössler system.

Next, we turn to estimating the specific entropy rate for each of these systems. For each system, we generated inter-event interval sequences of length $T = 1000$. We then chose the model order p and bandwidths $(k_1, \ldots, k_{p+1})$ as described in Section 2.4. The 50-block cross-validated log-likelihood (29) as a function of p is shown in Figure 6. Based on the embedology [55] result from [53], an embedding of at least twice the box counting dimension of the underlying attractor is required. Both the Lorenz and Rössler attractors have box counting dimensions between two and three; thus, we expect that a value of p around six should be sufficient for the predictive density. We see that the 50-block cross-validated log-likelihood chooses $p = 9$ and $p = 8$ for the Lorenz and Rössler systems.
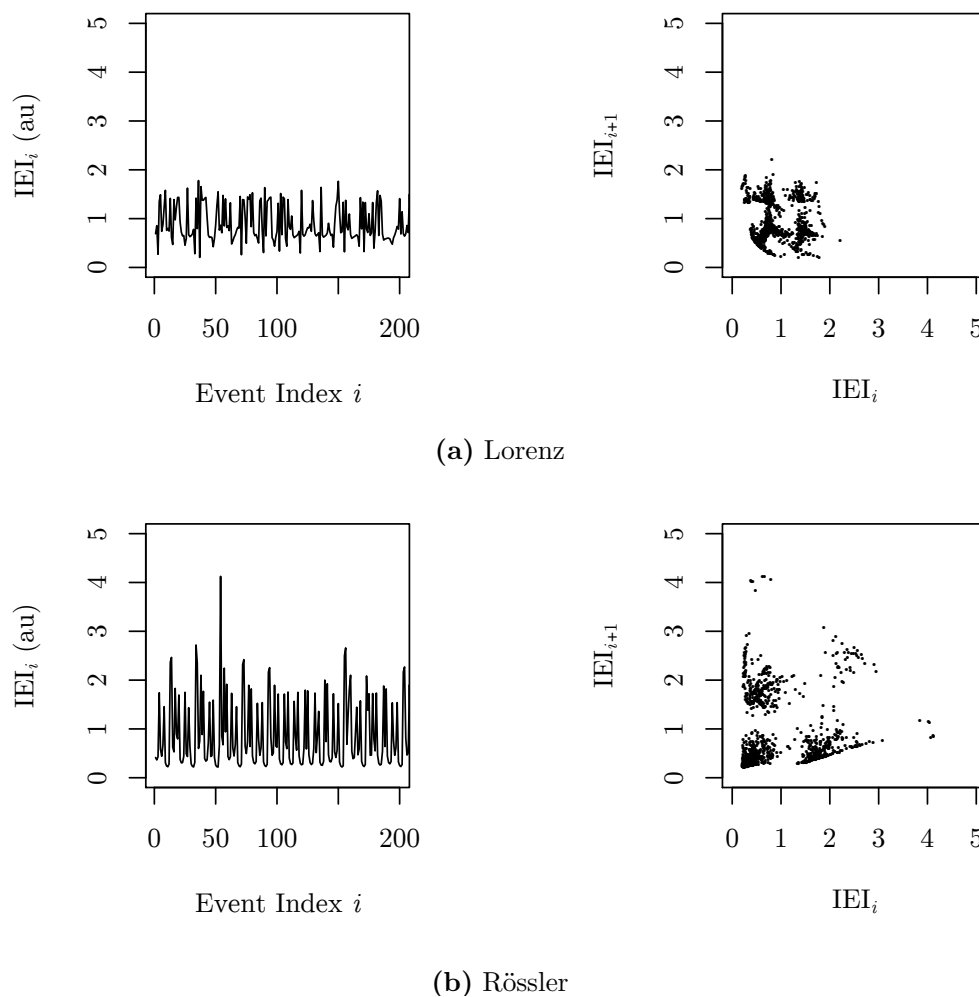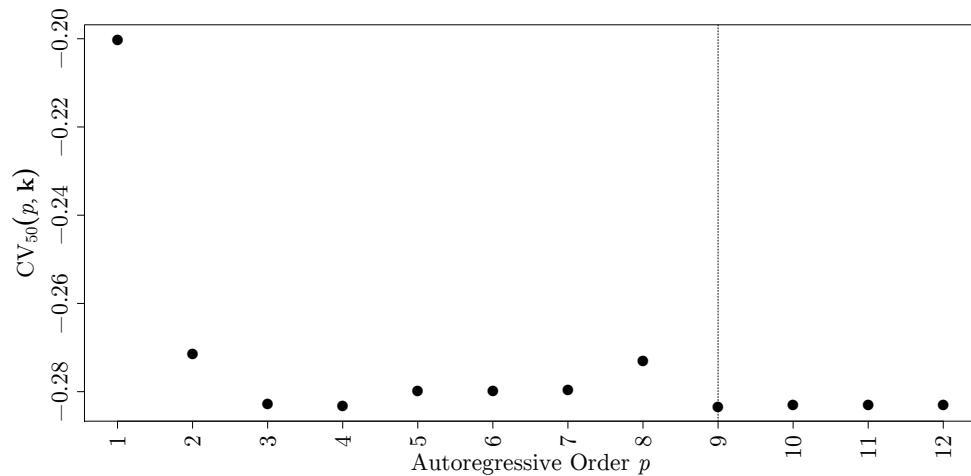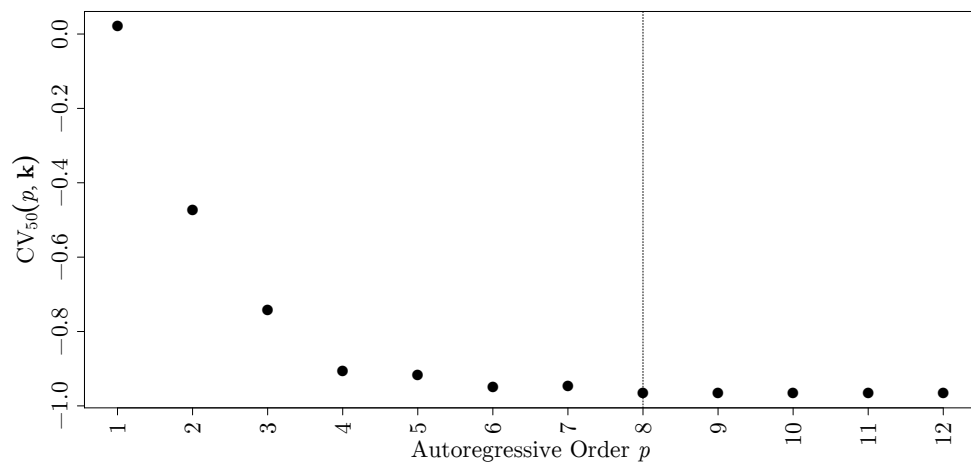
(a) Lorenz

(b) Rössler

Figure 5. Example inter-event intervals from an integrate-and-fire model driven by the $x(t)$ states of the Lorenz (top) and Rössler (bottom) systems. The inter-event interval lengths *versus* the event index (left) and the lag plots of the inter-event interval sequences (right) for both systems.

As mentioned in Sections 2.3 and 2.4, using cross-validation to choose the bandwidths of the conditional kernel density estimator introduces a form of feature selection into the conditional density estimation process: lags that are not relevant, as measured by the cross-validation score, are smoothed out by setting their associated bandwidths to infinity (in practice, to a large value). We demonstrate this phenomenon now for the bandwidths estimated for the inter-event intervals derived from the Lorenz and Rössler systems. For a fixed maximal lag p, Table 1 shows the bandwidths estimated for the Lorenz (top) and Rössler (bottom) systems. The first row indicates the bandwidths chosen by cross-validation for the future k_0 and past k_{-1} when we include only a single lag; the second row indicates the bandwidths chosen for the future k_0 and past (k_{-1}, k_{-2}) when we include two lags, *etc*. A horizontal dash (—) indicates that cross-validation has set the bandwidth associated with that lag to a value of five or greater, which is large with respect to the scale of the dynamics, thus in effect ignoring the lag in the estimation of the predictive density. Note that these bandwidths are for Gaussian kernels and, thus, are not immediately at the scale of the data. A transformation from the Gaussian scale to the uniform scale could be performed using the concept of canonical kernels [56]. Comparing Table 1 to Figure 6, we see that for the inter-event intervals generated by the Lorenz system, Intervals 4 through 7 can be ignored. This agrees with the sharp drop in Figure 6 at $p = 3$. Then, the Intervals 8 and 9 are included, but no others, thus giving the minimum at $p = 9$. A similar result holds for the bandwidths for the Rössler-governed inter-event intervals, where the bandwidths stabilize at $p = 8$, which also

corresponds to the minima in the 50-block cross-validated log-likelihood. Beyond this automatic selection of relevant lags, we see that the *magnitudes* of the bandwidths are very different amongst the $\mathbf{k} = (k_0, k_{-1}, \ldots, k_{-p})$: as one might expect, the bandwidths for the near past are smaller than the bandwidths for the distant past, *i.e.*, we should pay more attention to the recent past for prediction. Compare this inherent dynamic range in the bandwidths across lags to the fixed bandwidths across lags used in other statistics, such as approximate entropy, sample entropy, and multiscale entropy. If viewed as estimators of different differential entropy rates, these estimators would be severely biased by the fixed bandwidths.

(a) Lorenz

(b) Rössler

Figure 6. The 50-block cross-validated log-likelihoods (29) for the Lorenz (top) and Rössler (bottom) inter-event interval sequences as a function of the autoregressive order p. The vertical lines mark the minimum 50-block cross-validated log-likelihoods, which occur at $p = 9$ and $p = 8$, respectively.

Table 1. The optimal bandwidths $\mathbf{k} = (k_0, k_{-1}, \ldots, k_{-p})$ chosen using Equation (29) with p fixed from 1 to 12 for the inter-event intervals derived from the Lorenz (top) and Rössler (bottom) systems. A horizontal dash (—) indicates that cross-validation set the bandwidth associated with that lag to a value of 5 or greater, in effect ignoring the lag in the estimation of the predictive density. The bold rows correspond to bandwidths selected for the minimal values of p, as shown in Figure 6.

p	k_0	k_{-1}	k_{-2}	k_{-3}	k_{-4}	k_{-5}	k_{-6}	k_{-7}	k_{-8}	k_{-9}	k_{-10}	k_{-11}	k_{-12}
1	0.048	0.035											
2	0.059	0.039	0.055										
3	0.059	0.039	0.051	0.559									
4	0.059	0.039	0.051	0.558	—								
5	0.059	0.039	0.051	0.563	—	—							
6	0.059	0.039	0.051	0.564	—	—	—						
7	0.059	0.039	0.051	0.576	—	—	—	—					
8	0.070	0.050	0.057	0.450	0.541	0.625	—	—	0.674				
9	**0.059**	**0.039**	**0.052**	**0.570**	**—**	**—**	**—**	**—**	**1.263**	**0.826**			
10	0.059	0.039	0.052	0.573	—	—	—	—	1.194	0.816	—		
11	0.059	0.039	0.052	0.571	—	—	—	—	1.188	0.819	—	—	
12	0.059	0.039	0.052	0.574	—	—	—	—	1.184	0.816	—	—	—

(a) Lorenz

p	k_0	k_{-1}	k_{-2}	k_{-3}	k_{-4}	k_{-5}	k_{-6}	k_{-7}	k_{-8}	k_{-9}	k_{-10}	k_{-11}	k_{-12}
1	0.047	0.087											
2	0.062	0.054	0.052										
3	0.064	0.049	0.044	0.058									
4	0.065	0.048	0.046	0.072	0.078								
5	0.065	0.049	0.047	0.073	0.087	0.575							
6	0.065	0.053	0.051	0.082	0.089	0.751	0.185						
7	0.064	0.052	0.051	0.088	0.086	0.787	0.359	0.732					
8	**0.065**	**0.053**	**0.055**	**0.086**	**0.100**	**—**	**0.360**	**0.820**	**0.553**				
9	0.064	0.054	0.055	0.086	0.100	—	0.366	0.805	0.613	—			
10	0.065	0.053	0.054	0.085	0.100	—	0.359	0.810	0.573	—	—		
11	0.064	0.054	0.055	0.087	0.099	—	0.369	0.812	0.592	—	—	—	
12	0.065	0.054	0.054	0.086	0.101	—	0.366	0.808	0.580	—	—	—	—

(b) Rössler.
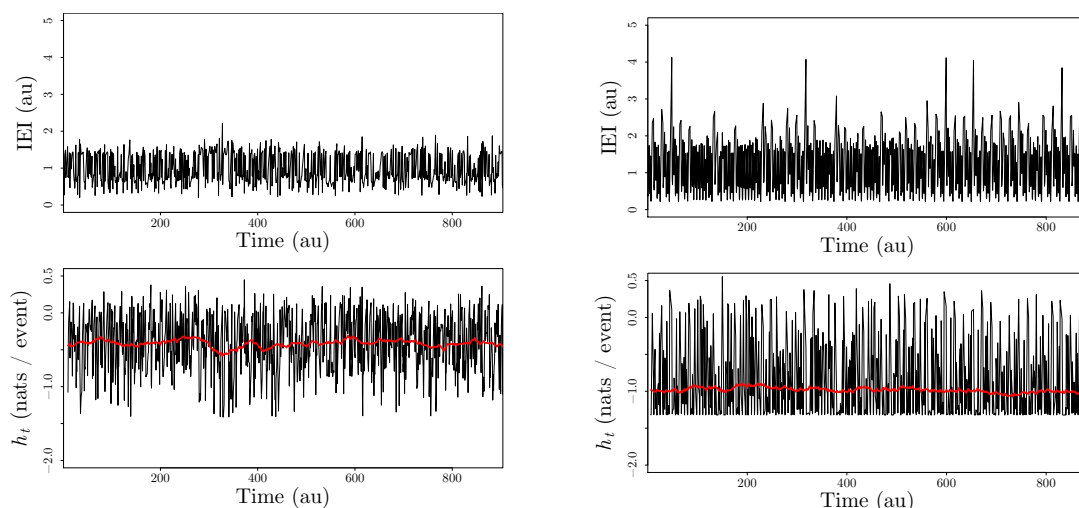
(a) **Lorenz** (b) **Rössler**

Figure 7. The inter-event interval sequence (top) and specific entropy rate (bottom) for the Lorenz (left) and Rössler (right) systems. Note that both the inter-event intervals and specific entropy rates are plotted as a function of the event times rather than the event index. The solid red line indicates a time-windowed average of the specific entropy rate with a uniform kernel with a window length of 60 au.

Now, consider the specific entropy rate of two inter-event interval sequences as a function of time, shown in Figure 7. Note that both the inter-event intervals and specific entropy rates are shown as a function of the *time* rather than the *event* index. That is, for each inter-event interval sequence, we show (T_i, IEI_i) and (T_i, h_i). The estimate of the time-averaged specific entropy rate (20) for the Lorenz and Rössler inter-event interval sequences are -0.41 nats/event and -1.0 nat/event, respectively. In addition, we also show a moving windowed average of the specific entropy rate using a uniform kernel of a width of 60 au in red in the bottom panel of Figure 7. This can be thought of as a local (in time) version of Equation (20) and allows us to determine if there are periods of time when the inter-event intervals are more, or less, predictable. For example, we see a drop in the specific entropy rate for the Lorenz inter-event intervals around 300 au, which corresponds to a run of relatively long and regular inter-event intervals.

We see from both Equation (20) and its time-local version that the interbeat interval sequence derived from the Rössler system is more predictable, which matches our intuition as outlined above based on the near-linear dynamics of the *x*-coordinate of the Rössler system. The thresholds Θ were chosen such that each system has approximately equal mean inter-event interval lengths: 0.90 au and 0.88 au for the Lorenz and Rössler systems, respectively. However, the pointwise standard deviations of the two inter-event interval sequences *are* different: 0.39 au and 0.73 au for the Lorenz and Rössler systems, respectively. Recall that, unlike discrete entropy, differential entropy is *not* scale invariant. This motivates determining a scale-invariant analog of the specific entropy rate that teases apart inherent unpredictability from the natural scale of the system. We will consider this point in the Discussion section.
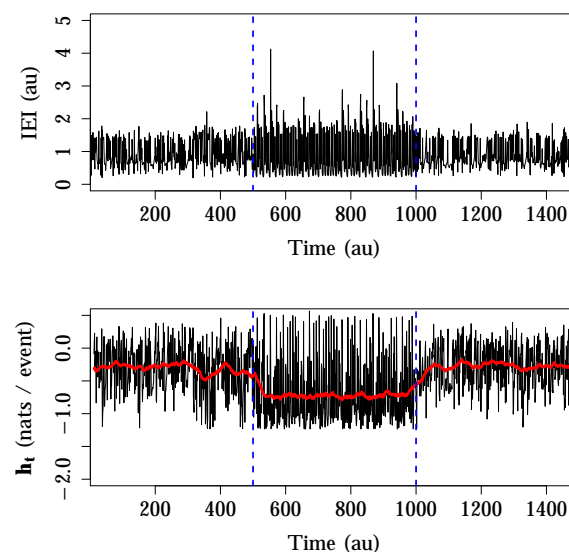

Figure 8. The inter-event interval sequence (top) and specific entropy rate (bottom) for the concatenation of Lorenz, Rössler, and Lorenz inter-event intervals. The dashed blue lines indicate the transitions from one system to the other. Compare to Figure 7, where the specific entropy rates were estimated individually for each system.

As a final example, we consider the estimation of the specific entropy rate where the inter-event interval sequence transitions from being generated by the Lorenz system to being generated by the Rössler system and back again. In this case, the inter-event interval sequence is clearly non-stationary. However, conditional stationarity is only violated locally in time around the transitions. To generate this time series, we concatenate 500 inter-event intervals each from the Lorenz, Rössler, and then Lorenz systems, and thus, $T = 1500$. This sequence is shown in the top panel of Figure 8. We estimate the autoregressive order p over the entire time series using Equation (29). The 50-block cross-validated

log-likelihood as a function of p is shown in Figure 9. The minima occurs at $p = 11$. Note that this is a higher order than chosen for either the Lorenz ($p = 9$) or Rössler ($p = 8$) systems when estimated in isolation. We see that additional information about the past is required when we need to distinguish between the two systems. Finally, Table 2 demonstrates the bandwidths chosen by cross-validation as a function of the maximal lag p. Again, we see that cross-validation provides both model selection and adaptive smoothing.
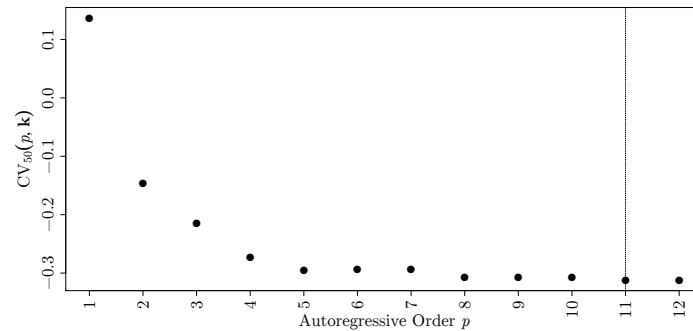

Figure 9. The 50-block cross-validated log-likelihood (29) for the concatenation of the Lorenz, Rössler, and Lorenz inter-event interval sequences as a function of the autoregressive order p. The vertical line marks the minimum log-likelihood, which occurs at $p = 11$.

Table 2. The optimal bandwidths $\mathbf{k} = (k_0, k_{-1}, \ldots, k_{-p})$ chosen using Equation (29) with p fixed from 1 to 12 for the inter-event intervals derived from the concatenation of the Lorenz, then Rössler, then Lorenz systems. A horizontal dash (—) indicates that cross-validation set the bandwidth associated with that lag to a value of 5 or greater, in effect ignoring the lag in the estimation of the predictive density. The bold row correspond to bandwidths selected for the minimal value of p as shown in Figure 9.

p	k_0	k_{-1}	k_{-2}	k_{-3}	k_{-4}	k_{-5}	k_{-6}	k_{-7}	k_{-8}	k_{-9}	k_{-10}	k_{-11}	k_{-12}
1	0.048	0.063											
2	0.064	0.046	0.059										
3	0.074	0.046	0.047	0.370									
4	0.071	0.049	0.051	0.417	0.459								
5	0.070	0.047	0.058	0.431	0.512	0.650							
6	0.070	0.047	0.058	0.431	0.513	0.649	—						
7	0.070	0.047	0.058	0.432	0.513	0.646	—	—					
8	0.070	0.050	0.057	0.450	0.541	0.625	—	—	0.674				
9	0.070	0.051	0.059	0.455	0.531	0.661	—	—	0.710	—			
10	0.070	0.050	0.057	0.454	0.542	0.620	—	—	0.666	—	—		
11	**0.071**	**0.051**	**0.058**	**0.470**	**0.548**	**0.632**	**—**	**—**	**0.622**	**—**	**—**	**0.985**	
12	0.071	0.051	0.057	0.471	0.548	0.634	—	—	0.628	—	—	0.997	—

The bottom panel of Figure 8 shows the specific entropy rate as a function of time for the concatenated system. As before, the black line is the specific entropy rate, and the red line is a moving windowed average of the specific entropy rate. Again, we see that the specific entropy rate drops as the system transitions from the Lorenz inter-event intervals to the Rössler inter-event intervals and then increases after the transition back to the Lorenz inter-event intervals. There is, however, a slight penalty to estimating the specific entropy rate for the concatenated inter-event interval sequences all at once. During the Lorenz-governed inter-event interval sequence, the time-averaged specific entropy rates are -0.30 nats/event and -0.28 nats/event, compared to -0.41 nats/event when estimated in isolation. Similarly, the time-averaged specific entropy rate for the Rössler-governed inter-event interval sequence is -0.72 nats/event compared to -1.0 nats/event when estimated in isolation. In both cases, we see that the specific entropy rates have increased. This is largely due to the fact

that the optimal bandwidths $k_1, \ldots, k_{p+1}$ when estimating the predictive density for either system in isolation are *not* optimal for estimating the concatenation of the two systems. This will lead to larger bandwidths overall and, thus, higher specific entropy rates. For this system, the difference in the dynamics is very large and the transition point relatively obvious, and thus, a better approach might be to estimate the predictive densities separately for each segment. However, in those cases where such transitions are non-obvious or where manual transition detection is not desirable, we see that estimating the predictive density all at once still leads to discrimination between high and low specific entropy rates.

Figure 10 demonstrates the inter-event interval sequence (top), predictive density (middle), and specific entropy rate (bottom) for the inter-event interval sequence for two time instants during the Lorenz (left) and Rössler (right) portions. The time instant during the portion governed by the Lorenz system has a higher specific entropy rate, as we would expect given the multi-modal nature of the estimated predictive density in the middle panel. In contrast, the time instant during the portion governed by the Rössler system has a lower specific entropy rate, as we would expect from the uni-modal and narrow estimated predictive density. However, we see that in both cases, the specific entropy rate can vary widely depending on the state of the system. For example, during periods around the long inter-event intervals, the inter-event intervals generated by the Rössler system can have higher specific entropy rates than those governed by the Lorenz system (the peaks in the specific entropy rate).
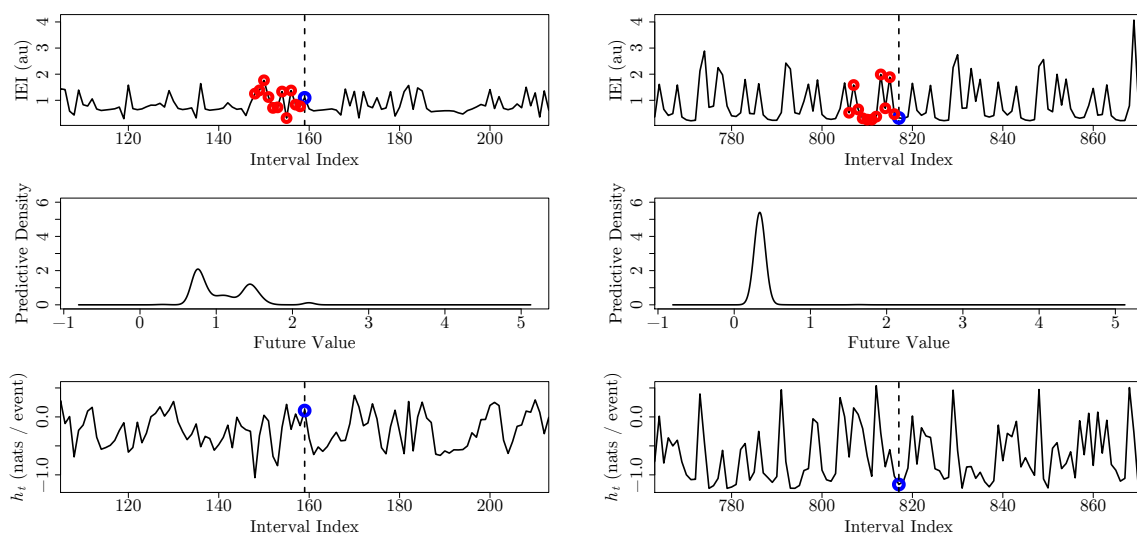

Figure 10. A demonstration of (top) the inter-event interval sequence, (middle) the estimated predictive density $\hat{f}\left(\text{IEI}_i \mid \text{IEI}_{i-11}^{i-1}\right)$ and (bottom) the specific entropy rate for the concatenated Lorenz, Rössler, Lorenz system during the Lorenz (left) and Rössler (right) portions of the sequence. In the top panels, the dashed vertical bar indicates the event index i; the red circles correspond to the specific past IEI_{i-11}^{i-1}; and the blue circles correspond to the future value IEI_i.

3.3. Specific Entropy Rate from a Tilt Table Experiment

As a final example, we consider the specific entropy rates of interbeat interval sequences from subjects participating in a tilt table experiment. It is well known by anyone with a heart that the *rate* of their pulse, the average number of beats within a specified window of time, can vary widely based on environmental, physiological, and psychological factors. However, it was not until the 20th century that researchers came to realize that *beat-to-beat* variations in heart rate convey information about the health of individuals. The study of beat-to-beat variations in heart rate is typically referred to under the umbrella term of heart rate variability. See [57–59] for a historical perspective on heart rate variability.

The nonlinear dynamics community has contributed a large number of methods for the analysis of interbeat intervals. See [60] for an extensive historical and methodological review.

In what follows, we use the term interbeat interval (IBI) to refer to the times between the R components of adjacent QRS complexes in the electrocardiograms. Common statistics computed from heart rate variability data include the mean interbeat interval and the standard deviation of the interbeat intervals. In addition, it is common to interpolate the interbeat interval sequence to obtain an equi-spaced sequence for spectral analysis [61], from which the power of high frequency and low frequency components, and their ratio, are commonly reported. It is also very common to compute approximate and/or sample entropies of interbeat interval sequences. Any, and sometimes all, of these statistics are referred to as heart rate variability (HRV), and thus, we will refrain from using that term. Many of these quantities can be computed by off-the-shelf software tailored for heart rate variability analysis, such as Kubios [62], though we recommend caution when using such software, since many of the parameters involved in both pre-processing of the data and its analysis are set in an *ad hoc* fashion.

As before, our approach to analyzing an interbeat interval sequence is to view it as the realization of some conditionally stationary stochastic dynamical system. This perspective naturally handles the fact that heart beats occur as a point process in time, as we saw in the previous section. Thus, we can compute the specific entropy rate associated with the time until the next heart beat, conditional on the most recent interbeat intervals. That is, if we denote the time between the $(i-1)$-th and i-th heart beat by IBI_i, we consider the specific entropy rate as $h\left[IBI_i \mid IBI_{i-p}^{i-1}\right]$.

We will investigate the specific entropy rate from the interbeat interval sequences of five subjects participating in a tilt table experiment. The population consisted of two males and three females between the ages of 27 and 44. In the experiment, the subject initially positioned himself/herself in a prone position on the table and was secured to the table. The subject was then kept in the supine position for five minutes, then tilted upright for five minutes and finally was returned to a supine position for five minutes. An electrocardiogram (ECG) was continuously recorded throughout the experiment. The interbeat intervals were extracted using the first amplitude-and-first derivative (AF1) algorithm from [63].

Specific entropy rates were computed for each subject using model orders p and bandwidths $(k_1, \ldots, k_{p+1})$ chosen as described in Section 2.4. The interbeat interval sequences (top) and specific entropy rates (bottom) for each subject are shown in Figure 11. For each subject, we see the expected decrease in interbeat interval length (increase in heart rate) as they move from a supine to upright position. However, for subjects (a)–(d), this change in mean interbeat interval length is also associated with a change in the overall dynamics of the interbeat interval sequence, which results in a drop in the specific entropy rate during the upright time period. With the return to supine position, the interbeat interval lengths again increase (the heart rate decreases), and the specific entropy rates of subjects (a)–(d) return to the same level as the start of the experiment.

Clearly, with only five subjects and a single session from each subject, we cannot say much about either the typical or atypical evolution of specific entropy rates in a tilt table experiment. However, it is interesting to note that Subject e, the only outlier in terms of the evolution of their specific entropy rate over time, is also the only subject with a traumatic brain injury in their past. Head trauma has been associated with changes in both spectral and information theoretic properties of interbeat interval sequences at rest [64,65]. Our results corroborate these findings and suggest that additional studies that include a physiological stressor, such as the tilt table, may be even more disclosing.
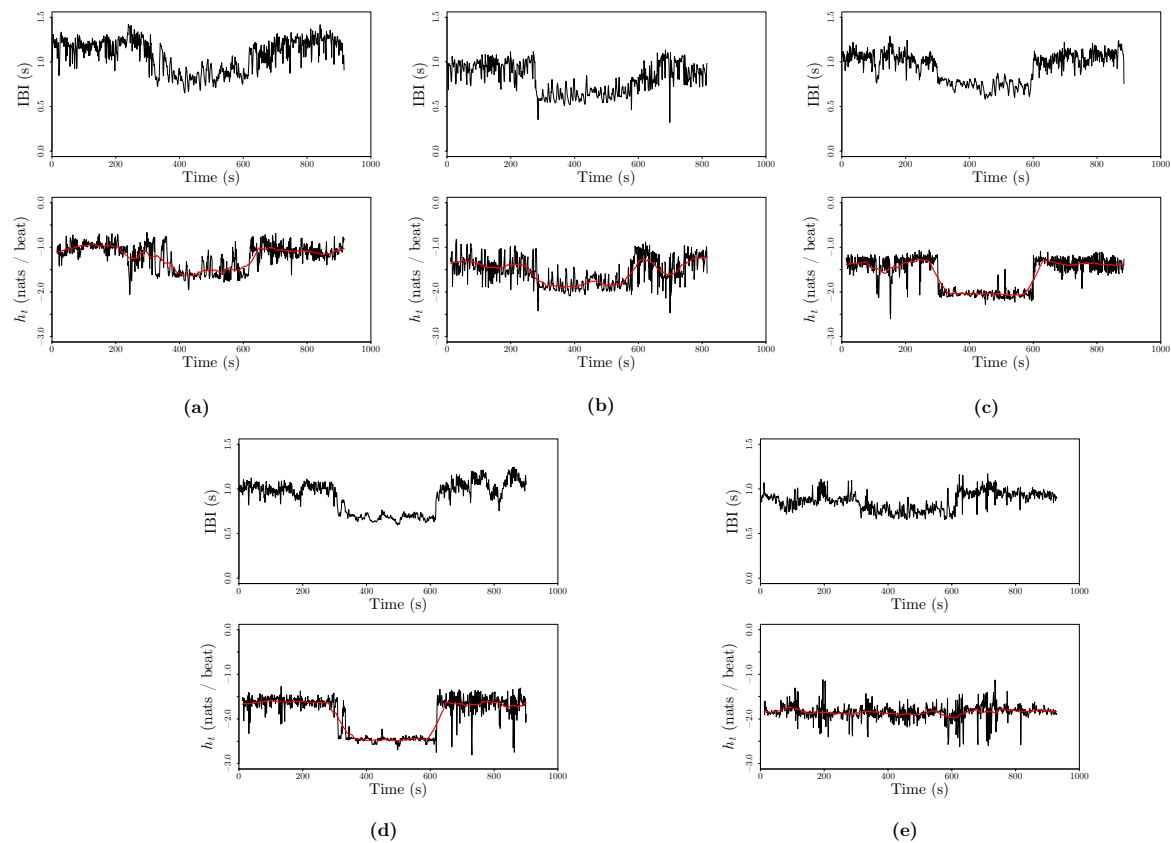
Figure 11. The interbeat interval sequences (top) and specific entropy rates (bottom) for each of the five subjects (a)–(e) in the tilt table experiment. The solid red line indicates a time-windowed average of the specific entropy rate with a uniform kernel with a window length of 60 s.

4. Discussion and Future Directions

An important consideration for any estimator relates to how it behaves under error or in the presence of noise. Care must be taken with respect to how one defines error, however. For example, does error refer to observational noise, model uncertainty/misspecification, or unobserved factors [43]? We have not considered the impact of observational noise, for example, because the measurements we have considered, namely inter-event and interbeat intervals, can be treated as relatively noise free. However, if observational noise *is* a major concern, then the estimation of the specific entropy rate must be carefully applied in this context, since direct estimation from the observed signal will combine dynamical and observational uncertainties. Possible solutions include the errors-in-variables model for density estimation [40] or more general nonlinear filtering approaches [66].

We have considered only *fixed* bandwidths for the conditional kernel density estimator in estimating the specific entropy rate: regardless of the past and future states of the system, we use the same bandwidths in estimating the predictive density. In Section 3.2, we saw a scenario where this estimation strategy may be problematic: the typical scale of the inter-event intervals differed between the Lorenz-governed and Rössler-governed periods, and this led to suboptimal bandwidths. Alternative variable bandwidth density estimation schemes allow the bandwidths to vary with either the data used in estimation of the density or the point of evaluation [36]. For example, the estimator for the differential entropy of a random vector developed in [30] based on *k*-nearest neighbor statistics is equivalent to a plug-in estimator of the differential entropy using a kernel density estimator with a bandwidth that varies with the point of evaluation, in this case the distance to the *k*-th nearest neighbor of the evaluation point, along with an additional bias correction term. Many other estimators, such as

the popular Kraskov–Stögbauer–Grassberger mutual information estimator [31], fall into this category. Future work will explore the tradeoff between the resolution gained by variable bandwidth estimators of a specific entropy rate and the statistical and computational burden imposed. One recent approach along these lines used a variable bandwidth kernel density estimator to estimate the transfer entropy for various simulated systems [67].

Another potential issue with scaling, as we again saw in Section 3.2, is that differential entropy and, thus, the differential entropy rate are not invariant to scaling. For example, changing the units used to measure the system under consideration will result in an affine shift to the differential entropy. Depending on the application at hand, this may or may not be problematic. If comparing the entropy rate of multiple time series, all with the same units, then the lack of invariance to scale washes out. However, if one is analyzing a single time series that has large variations in its characteristic scale over time, then the dependence on scaling may be problematic. One potential alternative is to normalize the differential entropy rate using the typical scale of the system at any given instant. A good candidate for this is the *negentropy* [68] of a random variable, which normalizes the differential entropy by the differential entropy associated with a Gaussian density with the same variance. The negentropy, unlike the differential entropy, is invariant to affine transformations of a random variable. Thus, we might define a *specific negentropy rate* by normalizing the specific entropy rate by an instantaneous measure of the variance. This is analogous to the *redundancy* [43] of discrete-state stochastic processes, which normalizes the entropy rate of a stochastic process by the entropy rate of a uniformly-distributed process with the same alphabet.

Any method that utilizes either approximate or sample entropy could be modified to use our specific entropy rate estimator. For example, the multiscale entropy [69], which is defined as the sample entropy of a time series at varying levels of aggregation, could easily be modified by direct substitution with the specific entropy rate. This would allow for not only an analysis of the unpredictability across *scales*, but also across *time*. Similarly, the point process model of interbeat interval sequences introduced in [70–72] is a particular parametric form for the stochastic dynamical system Equation (2). In a sequel [73], the authors propose using the filtered state from this model to estimate what they call the inhomogeneous point-process entropy. They estimate this quantity using either the approximate or sample entropy, and thus, based on the analysis from [14], we see that their estimator is for the unscaled Shannon or Rényi entropy rate of the filtered state. Thus, the specific entropy rate could be used on the filtered state.

Our approach to *specific* entropy rate estimation via conditional kernel density estimation can also be extended to any of the various other information theoretic measures gaining popularity, including transfer entropy [41,74], causation entropy [75], and co-/multi-information [76]. Many of these quantities would benefit from a data-driven approach to bandwidth selection, in addition to the automatic dimension reduction such approaches induce. However, we also note that with each additional probabilistic conditioning required by these measures, we increase both the statistical and computational burden for constructing the appropriate estimator. For example, the convergence rate of kernel density-based estimators for many information theoretic quantities scale exponentially in the reciprocal of the dimension of the random vector [77], while their time complexities scale exponentially in the dimension of the random vector [34].

5. Conclusions

Via a decomposition of the entropy rate of a discrete-time, continuous-valued stochastic dynamical system, we have proposed a measure of state-specific uncertainty: the specific entropy rate. We have shown how to estimate the specific entropy rate from finite data using kernel density estimators and provided a data-driven method for choosing the free parameters in the kernel density estimation. Given the immense popularity of heuristic approaches to entropy rate estimation, such as approximate entropy and sample entropy, it is our hope that a more principled approach to entropy rate estimation will be found useful by the larger research community.

All of the software used in this paper was developed in R via extensions to the `np` library for kernel density estimation. We have made this implementation available on GitHub [78]. In an effort to match the naming convention applied to Approximate Entropy (ApEn) and Sample Entropy (SampEn), we call our R implementation `spenra` for **Sp**ecific **En**tropy **Ra**te.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix: Relationship between the Kernel Density Estimator for the Differential Entropy Rate and Approximate Entropy

In this Appendix, we make the connection first noted in [14] between the kernel density estimator for the differential entropy rate and approximate entropy, emphasizing the implicit assumptions on the kernel, bandwidths, *etc.*, that result from the default parameters used by most approximate entropy-based analyses. However, we also note that [46] did not motivate approximate entropy as a kernel density-based estimator of the entropy rate, but rather as a family of statistics for comparing two time series. This explains, for example, the inclusion of both self-matching and sample size-independent bandwidths, which would lead to estimation bias from the perspective of kernel density estimation.

We begin by recalling the standard formulation of approximate entropy from [46]. Consider a time series $\{X_t\}_{t=1}^{T}$. For an embedding dimension p, we form the embedding vectors $\left\{\mathbf{X}_t^{(p)}\right\}_{t=1}^{T-p+1}$ where $\mathbf{X}_t^{(p)} = (X_t, X_{t+1}, \ldots, X_{t+p-1})$. For each vector $\mathbf{X}_t^{(p)}$, we compute the number of other vectors (including the vector indexed by t) that are within a tolerance r of $\mathbf{X}_t^{(p)}$ under the infinity norm,

$$C_t^{(p)}(r) = \frac{\#\left\{\mathbf{X}_{t'}^{(p)} : \left|\left|\mathbf{X}_t^{(p)} - \mathbf{X}_{t'}^{(p)}\right|\right|_\infty \le r\right\}}{T - p + 1}, \tag{A1}$$

where we recall that the infinity norm $||\cdot||_\infty$ of a vector $\mathbf{u} = (u_1, \ldots, u_p)$ is given by:

$$||\mathbf{u}||_\infty = \max_i |u_i|. \tag{A2}$$

Finally, we compute the average logarithm of Equation (A1) across all of the vectors, giving:

$$\Phi^{(p)}(r) = \frac{1}{T - p + 1} \sum_{t=1}^{T-p+1} \log C_t^{(p)}. \tag{A3}$$

For fixed p, r, and T, the approximate entropy is defined as:

$$\text{ApEn}(p, r, T) = \Phi^{(p)}(r) - \Phi^{(p+1)}(r). \tag{A4}$$

We next show that Equation (A4) is almost equivalent to a plug-in entropy rate estimator based on kernel density estimation. We begin by rewriting the $C_t^{(p)}(r)$ terms using the uniform/boxcar kernel $K_{\text{uniform}}(u) = \mathbf{1}_{[-1,1]}(u)$ as:

$$C_t^{(p)}(r) = \frac{\#\left\{\mathbf{X}_{t'}^{(p)} : \left\|\mathbf{X}_t^{(p)} - \mathbf{X}_{t'}^{(p)}\right\|_\infty \leq r\right\}}{T - p + 1} \tag{A5}$$

$$= \frac{1}{T - p + 1} \sum_{t=1}^{T-p+1} K_{\text{uniform}}\left(\frac{\left\|\mathbf{X}_t^{(p)} - \mathbf{X}_{t'}^{(p)}\right\|_\infty}{r}\right) \tag{A6}$$

$$= \frac{1}{T - p + 1} \sum_{t=1}^{T-p+1} \prod_{i=0}^{p-1} K_{\text{uniform}}\left(\frac{|X_{t+i} - X_{t'+i}|}{r}\right). \tag{A7}$$

We see that Equation (A7) is equivalent to the kernel density estimator for the density of $\left\{\mathbf{X}_t^{(p)}\right\}_{t=1}^{T-p+1}$ using a product of uniform kernels up to a normalization factor of $(2r)^{-p}$. The true kernel density estimator therefore would be given by:

$$\hat{f}\left(\mathbf{x}^{(p)}\right) = \frac{1}{T - p + 1} \sum_{t=1}^{T-p+1} \frac{1}{(2r)^p} K_{\text{uniform}}\left(\frac{\|\mathbf{x}^{(p)} - \mathbf{X}_t^{(p)}\|_\infty}{r}\right) \tag{A8}$$

$$= \frac{1}{T - p + 1} \sum_{t=1}^{T-p+1} \prod_{i=0}^{p-1} \frac{1}{2r} K_{\text{uniform}}\left(\frac{|x_i - X_{t+i}|}{r}\right). \tag{A9}$$

Therefore, we see that Equation (A1) is the unnormalized form of Equation (A9) evaluated at $\mathbf{X}_t^{(p)}$. If we include the normalization, the summation Equation (A3) becomes:

$$\Phi_{\text{normalized}}^{(p)}(r) = \frac{1}{T - p + 1} \sum_{t=1}^{T-p+1} \log \hat{f}\left(\mathbf{X}_t^{(p)}\right). \tag{A10}$$

If $\hat{f}$ were replaced with the true density f, then for large T, $\Phi_{\text{normalized}}^{(p)}(r)$ approximates the negative joint differential entropy:

$$h\left[\mathbf{X}^{(p)}\right] = -E\left[\log f\left(\mathbf{X}^{(p)}\right)\right] \tag{A11}$$

$$= -\int_{\mathbb{R}^p} f\left(\mathbf{x}^{(p)}\right) \log f\left(\mathbf{x}^{(p)}\right) d\mathbf{x}^{(p)} \tag{A12}$$

by the law of large numbers. However, because we evaluate the estimator $\hat{f}$ with the same data used to estimate it, Equation (A10) is a biased estimator of the negative differential entropy $-h\left[\mathbf{X}^{(p)}\right]$. A simple modification of Equation (A10), due to [77,79], provides an estimator for the joint differential entropy with a fast rate of convergence in the i.i.d. case. In particular, let $\hat{f}_{-t}$ be the kernel density estimator for the joint density formed by leaving out the t-th vector $\mathbf{X}_t$. That is, we estimate the joint density using Equation (A9) with all of the vectors, except $\mathbf{X}_t$. This gives the leave-one-out (LOO) estimator for the joint differential entropy,

$$-\Phi_{\text{normalized, LOO}}^{(p)}(r) = -\frac{1}{T - p + 1} \sum_{t=1}^{T-p+1} \log \hat{f}_{-t}\left(\mathbf{X}_t^{(p)}\right). \tag{A13}$$

Thus, we see that with the proper normalization, a modification of the approximate entropy gives an estimator for the finite-p differential entropy rate,

$$h[X_{p+1} \mid X_p, \dots, X_1] = h[X_1, \dots, X_{p+1}] - h[X_1, \dots, X_p]. \tag{A14}$$

References

1. Shalizi, C.R. Methods and techniques of complex systems science: An overview. In *Complex Systems Science in Biomedicine*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 33–114.
2. Peliti, L.; Vulpiani, A. *Measures of Complexity*; Springer-Verlag: Berlin/Heidelberg, Germany, 1988.
3. Li, M.; Vitányi, P. *An Introduction to Kolmogorov Complexity and Its Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1993.
4. Rissanen, J. *Stochastic Complexity in Statistical Inquiry*; World Scientific: Singapore, Singapore, 1989.
5. Grassberger, P. Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.* **1986**, *25*, 907–938.
6. Crutchfield, J.P.; Young, K. Inferring statistical complexity. *Phys. Rev. Lett.* **1989**, *63*, doi:10.1103/Phys RevLett.63.105.
7. Shalizi, C.R.; Crutchfield, J.P. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.* **2001**, *104*, 817–819.
8. James, R.G.; Ellison, C.J.; Crutchfield, J.P. Anatomy of a bit: Information in a time series observation. *Chaos Interdiscip. J. Nonlinear Sci.* **2011**, *21*, 037109.
9. Kolmogorov, A.N. A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. *Dokl. Akad. Nauk SSSR* **1958**, *119*, 861–864.
10. Sinai, J. On the concept of entropy for a dynamic system. *Dokl. Akad. Nauk SSSR* **1959**, *124*, 768–771.
11. Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
12. Kantz, H.; Schreiber, T. *Nonlinear Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2004; Volume 7.
13. Crutchfield, J.P.; McNamara, B.S. Equations of motion from a data series. *Complex Syst.* **1987**, *1*, 417–452.
14. Lake, D.E. Renyi entropy measures of heart rate Gaussianity. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 21–27.
15. Ostruszka, A.; Pakoński, P.; Słomczyński, W.; Życzkowski, K. Dynamical entropy for systems with stochastic perturbation. *Phys. Rev. E* **2000**, *62*, 2018–2029.
16. Fraser, A.M. Information and entropy in strange attractors. *IEEE Trans. Inf. Theory* **1989**, *35*, 245–262.
17. Badii, R.; Politi, A. *Complexity: Hierarchical Structures and Scaling in Physics*; Cambridge University Press: Cambridge, UK, 1999; Volume 6.
18. Fan, J.; Yao, Q. *Nonlinear Time Series: Nonparametric and Parametric Methods*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2003.
19. Chan, K.S.; Tong, H. *Chaos: A Statistical Perspective*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
20. Michalowicz, J.V.; Nichols, J.M.; Bucholtz, F. *Handbook of Differential Entropy*; CRC Press: Boca Raton, FL, USA, 2013.
21. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
22. Ihara, S. *Information Theory for Continuous Systems*; World Scientific: Singapore, Singapore, 1993; Volume 2.
23. Grimmett, G.; Stirzaker, D. *Probability and Random Processes*; Oxford University Press: Oxford, UK, 2001.
24. Caires, S.; Ferreira, J. On the non-parametric prediction of conditionally stationary sequences. *Stat. Inference Stoch. Process.* **2005**, *8*, 151–184.
25. Yao, Q.; Tong, H. Quantifying the influence of initial values on non-linear prediction. *J. R. Stat. Soc. Ser. B Methodol.* **1994**, *56*, 701–725.
26. Yao, Q.; Tong, H. On prediction and chaos in stochastic systems. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1994**, *348*, 357–369.
27. DeWeese, M.R.; Meister, M. How to measure the information gained from one symbol. *Netw. Comput. Neural Syst.* **1999**, *10*, 325–340.

28. Lizier, J.T.; Prokopenko, M.; Zomaya, A.Y. Local information transfer as a spatiotemporal filter for complex systems. *Phys. Rev. E* **2008**, *77*, 026110.

29. Lizier, J.T. Measuring the dynamics of information processing on a local scale in time and space. In *Directed Information Measures in Neuroscience*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 161–193.

30. Kozachenko, L.; Leonenko, N.N. Sample estimate of the entropy of a random vector. *Probl. Peredachi Inf.* **1987**, *23*, 9–16.

31. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138.

32. Sricharan, K.; Wei, D.; Hero, A.O. Ensemble Estimators for Multivariate Entropy Estimation. *IEEE Trans. Inf. Theory* **2013**, *59*, 4374–4388.

33. Gao, S.; Ver Steeg, G.; Galstyan, A. Estimating Mutual Information by Local Gaussian Approximation. 2015, arXiv:1508.00536.

34. Singh, S.; Póczos, B. Analysis of k-Nearest Neighbor Distances with Application to Entropy Estimation. 2016, arXiv:1603.08578.

35. Lombardi, D.; Pant, S. Nonparametric k-nearest-neighbor entropy estimator. *Phys. Rev. E* **2016**, *93*, 013310.

36. Terrell, G.R.; Scott, D.W. Variable Kernel Density Estimation. *Ann. Stat.* **1992**, *20*, 1236–1265.

37. Rosenblatt, M. Conditional probability density and regression estimators. In *Multivariate Analysis II*; Academic Press: New York, NY, USA, 1969; Volume 25, p. 31.

38. Hall, P.; Racine, J.; Li, Q. Cross-validation and the estimation of conditional probability densities. *J. Am. Stat. Assoc.* **2004**, *99*, 1015–1026.

39. Hayfield, T.; Racine, J.S. Nonparametric Econometrics: The np Package. *J. Stat. Softw.* **2008**, *27*, 1–32.

40. Bosq, D. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 110.

41. Kaiser, A.; Schreiber, T. Information transfer in continuous processes. *Phys. D Nonlinear Phenom.* **2002**, *166*, 43–62.

42. Burman, P.; Chow, E.; Nolan, D. A cross-validatory method for dependent data. *Biometrika* **1994**, *81*, 351–358.

43. Crutchfield, J.P.; Feldman, D.P. Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos Interdiscip. J. Nonlinear Sci.* **2003**, *13*, 25–54.

44. Efromovich, S. Dimension reduction and adaptation in conditional density estimation. *J. Am. Stat. Assoc.* **2010**, *105*, 761–774.

45. Lahiri, S.N. *Resampling Methods for Dependent Data*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.

46. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301.

47. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049.

48. Teixeira, A.; Matos, A.; Antunes, L. Conditional rényi entropies. *IEEE Trans. Inf. Theory* **2012**, *58*, 4273–4277.

49. Lake, D.E.; Richman, J.S.; Griffin, M.P.; Moorman, J.R. Sample entropy analysis of neonatal heart rate variability. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2002**, *283*, R789–R797.

50. Lake, D.E. Nonparametric entropy estimation using kernel densities. *Methods Enzymol.* **2009**, *467*, 531–546.

51. Lake, D.E. Improved entropy rate estimation in physiological data. In Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 1463–1466.

52. Wand, M.P.; Schucany, W.R. Gaussian-based kernels. *Can. J. Stat.* **1990**, *18*, 197–204.

53. Sauer, T. Reconstruction of integrate-and-fire dynamics. *Nonlinear Dyn. Time Ser.* **1997**, *11*, 63.

54. Yentes, J.M.; Hunt, N.; Schmid, K.K.; Kaipust, J.P.; McGrath, D.; Stergiou, N. The appropriate use of approximate entropy and sample entropy with short data sets. *Ann. Biomed. Eng.* **2013**, *41*, 349–365.

55. Sauer, T.; Yorke, J.A.; Casdagli, M. Embedology. *J. Stat. Phys.* **1991**, *65*, 579–616.

56. Marron, J.S.; Nolan, D. Canonical kernels for density estimation. *Stat. Probab. Lett.* **1988**, *7*, 195–199.

57. Acharya, U.R.; Joseph, K.P.; Kannathal, N.; Lim, C.M.; Suri, J.S. Heart rate variability: A review. *Med. Biol. Eng. Comput.* **2006**, *44*, 1031–1051.

58. Berntson, G.G.; Bigger, J.T.; Eckberg, D.L. Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology* **1997**, *34*, 623–648.

59. Billman, G.E. Heart rate variability—A historical perspective. *Front. Physiol.* **2011**, *2*, doi:10.3389/fphys. 2011.00086.

60. Voss, A.; Schulz, S.; Schroeder, R.; Baumert, M.; Caminal, P. Methods derived from nonlinear dynamics for analysing heart rate variability. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2009**, *367*, 277–296.

61. Deboer, R.W.; Karemaker, J.M. Comparing spectra of a series of point events particularly for heart rate variability data. *IEEE Trans. Biomed. Eng.* **1984**, *4*, 384–387.

62. Tarvainen, M.P.; Niskanen, J.P.; Lipponen, J.A.; Ranta-Aho, P.O.; Karjalainen, P.A. Kubios HRV—Heart rate variability analysis software. *Comput. Methods Progr. Biomed.* **2014**, *113*, 210–220.

63. Friesen, G.M.; Jannett, T.C.; Jadallah, M.A.; Yates, S.L.; Quint, S.R.; Nagle, H.T. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Trans. Biomed. Eng.* **1990**, *37*, 85–98.

64. Su, C.F.; Kuo, T.B.; Kuo, J.S.; Lai, H.Y.; Chen, H.I. Sympathetic and parasympathetic activities evaluated by heart-rate variability in head injury of various severities. *Clin. Neurophysiol.* **2005**, *116*, 1273–1279.

65. Papaioannou, V.; Giannakou, M.; Maglaveras, N.; Sofianos, E.; Giala, M. Investigation of heart rate and blood pressure variability, baroreflex sensitivity, and approximate entropy in acute brain injury patients. *J. Crit. Care* **2008**, *23*, 380–386.

66. Tanizaki, H. *Nonlinear Filters: Estimation and Applications*; Springer Science & Business Media: Berlin/ Heidelberg, Germany, 1996.

67. Zuo, K.; Bellanger, J.J.; Yang, C.; Shu, H.; Le Jeannes, R.B. Exploring neural directed interactions with transfer entropy based on an adaptive kernel density estimator. In Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 4342–4345.

68. Comon, P. Independent component analysis, a new concept? *Signal Process.* **1994**, *36*, 287–314.

69. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102.

70. Barbieri, R. A point-process model of human heartbeat intervals: New definitions of heart rate and heart rate variability. *AJP Heart Circ. Physiol.* **2004**, *288*, H424–H435.

71. Chen, Z.; Brown, E.N.; Barbieri, R. Characterizing nonlinear heartbeat dynamics within a point process framework. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1335–1347.

72. Valenza, G.; Citi, L.; Scilingo, E.P.; Barbieri, R. Point-process nonlinear models with laguerre and volterra expansions: Instantaneous assessment of heartbeat dynamics. *IEEE Trans. Signal Process.* **2013**, *61*, 2914–2926.

73. Valenza, G.; Citi, L.; Scilingo, E.P.; Barbieri, R. Inhomogeneous point-process entropy: An instantaneous measure of complexity in discrete systems. *Phys. Rev. E* **2014**, *89*, 052803.

74. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464.

75. Sun, J.; Bollt, E.M. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. *Phys. D Nonlinear Phenom.* **2014**, *267*, 49–57.

76. Bell, A.J. The co-information lattice. In Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA, Nara, Japan, 1–4 April 2003; Volume 2003.

77. Kandasamy, K.; Krishnamurthy, A.; Poczos, B.; Wasserman, L.; Robins, J.M. Influence Functions for Machine Learning: Nonparametric Estimators for Entropies, Divergences and Mutual Informations. 2014, arXiv:1411.4342.

78. Darmon, D. `spenra` GitHub Repository. Available online: http://github.com/ddarmon/spenra (accessed on 17 May 2016).

79. Kandasamy, K.; Krishnamurthy, A.; Poczos, B. Nonparametric von mises estimators for entropies, divergences and mutual informations. In *Advances in Neural Information Processing Systems*; Morgan Kaufmann Publishers: Burlington, MA, USA, 2015; pp. 397–405.