Title
Testing time symmetry in time series using data compression dictionaries

Journal
Physical Review E, 69(5)

ISSN
1063-651X

Author
Kennel, Matthew B

Publication Date
2004-05-01

Peer reviewed

Testing time symmetry in time series using data compression dictionaries

Matthew B. Kennel*
Institute For Nonlinear Science, University of California, San Diego, La Jolla, California 92093-0402, USA
(Received 21 July 2003; published 14 May 2004; publisher error corrected 20 May 2004)

Time symmetry, often called statistical time reversibility, in a dynamical process means that any segment of time-series output has the same probability of occurrence in the process as its time reversal. A technique, based on symbolic dynamics, is proposed to distinguish such symmetrical processes from asymmetrical ones, given a time-series observation of the otherwise unknown process. Because linear stochastic Gaussian processes, and static nonlinear transformations of them, are statistically reversible, but nonlinear dynamics such as dissipative chaos are usually statistically irreversible, a test will separate large classes of hypotheses for the data. A general-purpose and robust statistical test procedure requires adapting to arbitrary dynamics which may have significant time correlation of undetermined form. Given a symbolization of the observed time series, the technology behind adaptive dictionary data compression algorithms offers a suitable estimate of reversibility, as well as a statistical likelihood test. The data compression methods create approximately independent segments permitting a simple and direct null test without resampling or surrogate data. We demonstrate the results on various time-series-reversible and irreversible systems.

DOI: 10.1103/PhysRevE.69.056208 PACS number(s): 05.45.Tp

I. INTRODUCTION

A well-known issue in the analysis of observed data is to distinguish colored noise produced from a Gaussian linear process from data produced from nonlinear sources. The tools of traditional, linear, signal processing and time-series statistics, power spectra, transfer functions, autoregressive modeling, etc., often fail in such cases when their assumptions are violated; but when these assumptions are fulfilled they are often provably optimal.

The technique [1–3] most commonly employed for this task is to generate Monte Carlo simulations of "surrogate data," a linear Gaussian noisy data set, with similar characteristics (e.g., power spectrum, autocorrelation, or autoregressive coefficients) as the original data and compare the original and surrogates on some statistic of the user's choice which is sensitive to various nonlinear features. This method is quite general but there are a number of subtle and tricky technical issues [4–7] which are not always appreciated, and it may be computationally intensive.

Testing for time asymmetry (e.g., Ref. [8,9] and their references) is a useful alternative to surrogate methods for distinguishing linear noise and static nonlinear transformations thereof from nonlinear dynamics. This idea relies on the fact that a stationary linear Gaussian stochastic process is statistically *time symmetrical*, also often called *time reversible*: the literal time reverse of the observed series would have the same probability to be emitted from the source as the observed one [10]. Any fixed static nonlinear transformation of such a process—including *nonmonotonic* transformations which have proven to be problematic in the surrogate-data method [7]—stays time reversible. Importantly for this work, one such transformation is the *symbolization* or discretization of a continuous state space to a coarse alphabet of a small number of symbols. Dissipative chaos, by contrast, will produce a statistically time-irreversible signal as the creation of information via instability in the time-forward direction is distinct from the destruction of past state information via dissipation. The meaning of statistical "irreversibility" used herein is not exactly the same as the "irreversibility" of physical processes in the traditional thermodynamic sense. Herein, we assume that the measured process is already in its statistically stationary condition, and use "reversibility" in its statistical sense: and the word "reversible" is a synonym for "time symmetrical."

This work does not give an explicit description of the "null hypothesis" (e.g., a linear Gaussian process) as would be done with a parametric estimate for the entire process, i.e., it is not feasible to directly evaluate the two likelihoods for seeing the observed set in its original orientation and its time-reversed orientation. With the usual requirements of stationarity and the absence of very long time dependence, one may empirically estimate the likelihood of reversible dynamics by looking at statistics of short-term segments from the data set, using ergodicity in the usual way so that a single long observed data set provides an ensemble. Our goal includes not merely a number quantifying the amount of time asymmetry, but a statistical *test procedure* with a null hypothesis and p value for rejection of the null. The generic complication is that general dynamics, linear or nonlinear, can possess rather arbitrary serial dependence. We want additionally a general procedure which requires as few assumptions about the structure of the dynamics as possible. The common theme is to try to construct a test out of sufficiently independent elements so that the assumptions of classical statistical test procedures hold.

Daw *et al.* [9] suggested using the observed frequency of symbolic words formed from nearby symbols as seen in the forward and reverse directions. For instance, in a binary alphabet, if a word length of 5 and a time delay of 1 were chosen, then one would accumulate the observed frequency of 11001, and its time reverse 10011, as the word window

*Electronic address: mkennel@ucsd.edu

slid incrementally over the symbolized observed data. The assumption under the null hypothesis of time symmetry is that the observed frequencies came from an equiprobable distribution which could be tested with a simple binomial test. This would be done for all nonpalindromic pairs of words of a fixed length, and the results of tests on all words combined. The difficulty comes in serial correlation which can make the assumption of independent observations in the binomial test incorrect, and the statistical dependence in the combination of results from many pairs. The first was ameliorated with a decorrelation window and additional correlation test, but the second does not have a clear solution. The appropriate word length is also an undesirable free parameter. As usual, short words provide a more accurate estimation of probabilities (high counts) but may improperly average over different dynamics which would be more visible with longer words. This work proposes a different method, adapting techniques from data compression, to rectify all these issues. It provides approximately independent quantities for a statistical test as well as automatic word-length selection.

II. ADAPTIVE DICTIONARY-BASED TIME-SYMMETRY TESTING

The Lempel-Ziv [11] dictionary compression algorithm sequentially parses the input symbol sequence from left to right, at each step finding the *longest* segment in the remaining input which already exists in a dictionary of codewords [21]. Then a new codeword, consisting of the longest existing match concatenated with the next subsequent symbol in the input, is added to the dictionary [12]. An index for the codeword which was originally located and the subsequent symbol are emitted. The input pointer is advanced by the length of the codeword just added plus one. The compressed output is a sequence of pairs of codeword indices and the additional symbol: $(w_1 s_1)(w_2 s_2)\cdots(w_n s_n)$. The dictionary is initialized with A length-one strings, each comprising each unique symbol in the alphabet of size A. Absent *a priori* bounds on the maximum size of the integers, the length, in bits, of the compressed stream is proportional to $n \log_2 n$ with n the number of phrases. This compression is *universal*: the length of the compressed sequence divided by the length of the input will asymptotically approach the Shannon entropy rate (the best possible compression rate) for almost any source, meaning that the method is guaranteed to learn characteristics of the source. Frequently occurring sequences generate longer dictionary entries whose codeword indices (represented as integers) may be transmitted more compactly than their plaintexts.

One may parse a new sequence relative to a given fixed dictionary, for instance, that obtained after compressing another sequence as previously discussed. The longest codeword in the dictionary which is a prefix of the remaining input is identified and emitted. The input pointer is advanced by the length of this codeword. This is like compressing the latter half of a sequence except that the adaptation (adding new phrases to the dictionary) is not performed. Fundamental results in information theory [13,14] imply that when the parsed sequence arises from the same information source which produced the sequence used to train the dictionary, it will nearly always take fewer bits (and phrases) than a parsing using a dictionary trained on a different source. This statement is technically only true asymptotically but in practice exceptions grow exponentially unlikely for mixing sources. This property concerning the relative entropy was recently used to distinguish and categorize natural languages from only representative samples of their texts [15], although there the slightly different algorithm was used and adaptation to the second sequence continued during its parsing, lowering the discrimination power somewhat.

We use this fact to test for time symmetry by comparing the compression performance using dictionaries which were trained on normal, and time-reversed, examples. There are many possible specific ways one could consider using compression to see if there is a difference, for example, parse a test sequence completely by the two different dictionaries and see which dictionary emits the fewest phrases, or, perhaps, looking at the statistical distribution of the lengths of words emitted. The following statistic and test, though, was powerful in detecting irreversibility, the relatively easy task, as well as having a good calibration of the null hypothesis under various diverse instantiations of reversible dynamics, which is the more difficult requirement.

Consider for a moment the generic problem of sequentially parsing a sequence S with respect to *two* dictionaries D_1, D_2 simultaneously. At each step, there is a longest matching codeword for each individual dictionary. Of those two, either the first dictionary provides the longest match, or the second does, or the lengths are tied (both dictionaries provide the same codeword). The input is advanced by the length of the longest match. We define our notation as follows: $n_1 = C_1(S|D_1, D_2)$ is the count of number of times the first dictionary (D_1) provided the codeword, and similarly for $n_2 = C_2(S|D_1, D_2)$; accumulating the counts the second was the best match. The number of ties is discarded. The two counts C_1 and C_2 are computed simultaneously for identical arguments. For our purposes, no actual literal compressed output is necessary, merely the accumulation of these counts. The key notion is that since dictionary-based universal compression attempts to make approximately independent codewords, the "observation" of a parsed phrase is as if it were nearly an independent event in a renewal-type process. This assumption of independence (which will be tested empirically) justifies simple classical statistical tests.

Specializing to the problem at hand, the key idea is to parse a test sequence with respect to dictionaries which were constructed on either forward or backward versions of a different training sequence. If the data are reversible, then either of those dictionaries is as good as the other, statistically, in providing longest matches and hence, on average gives as good compression as the other. Moreover, the assumption is that in time symmetry the distribution of "which dictionary provides a superior match here" is an independent Bernoulli binary random variable with equal probability, and thus the accumulated counts would be distributed like Poisson random variables.

Divide the input sequence S into its two contiguous halves (S_1 and S_2), create literal time-reversed versions of them

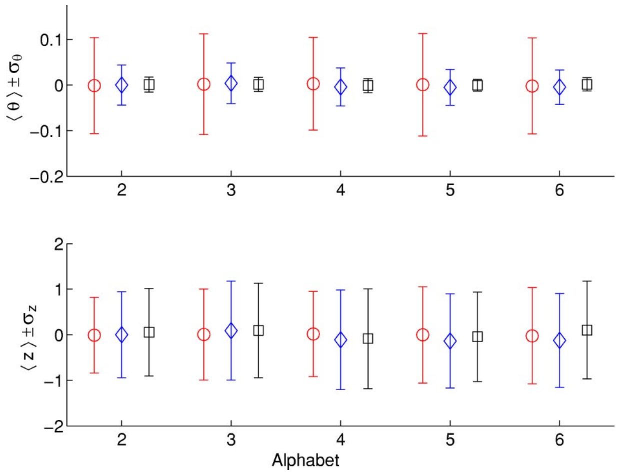

FIG. 1. (Color online) Summary statistics for white equiprobable symbols. There were 200 data sets drawn for each data set size, $N=200, 2500, 25\,000$ (red circle, blue diamond, black square), and the reversibility statistics θ and z were evaluated for each. Top: $\langle\hat{\theta}\rangle$, the ensemble average (arb. units), and its standard deviation. Bottom: $\langle z\rangle$ (arb. units), and its standard deviation.

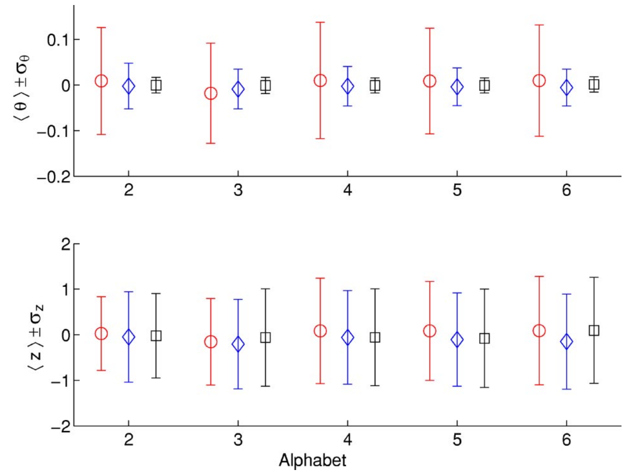

FIG. 2. (Color online) Top: summary statistics for a reversible mixture of logistic map dynamics. Symbolization was by equal-probability bins with $|A|$ from 2 to 6. There were 200 data sets drawn for each data set size, $N=200, 2500, 25\,000$ (red circle, blue diamond, black square), and the reversibility statistics θ and z were evaluated for each. Top: $\langle\hat{\theta}\rangle$, the ensemble average (arb. units), and its standard deviation. Bottom: $\langle z\rangle$ (arb. units) and its standard deviation. x axis is the size of the alphabet.

(R_1 and R_2), and create four dictionaries (D_{S1}, D_{S2}, D_{R1}, and D_{R2}) using the Lempel-Ziv construction as before. Parse each of the four sequences with respect to the the two dictionaries trained on the other half of the data. Accumulate the total number of same-direction (n_s) matches,

$$n_s = C_1(S_2|D_{S1},D_{R1}) + C_1(R_2|D_{R1},D_{S1}) + C_1(S_1|D_{S2},D_{R2}) + C_1(R_1|D_{R2},D_{S2}), \quad (1)$$

and different-direction (n_d) matches,

$$n_d = C_2(S_2|D_{S1},D_{R1}) + C_2(R_2|D_{R1},D_{S1}) + C_2(S_1|D_{S2},D_{R2}) + C_2(R_1|D_{R2},D_{S2}). \quad (2)$$

With $n=n_s+n_d$, define the time-symmetry statistic

$$\hat{\theta} = \frac{n_s - n_d}{n}. \quad (3)$$

Under the null hypothesis, $\hat{\theta}\to 0$ as $n\to\infty$. For $n\geqslant 25$, the null distribution of

$$z(\hat{\theta},n) = n^{1/2}\left(\hat{\theta} - \frac{1}{2n}\mathrm{sgn}(\hat{\theta})\right) \quad (4)$$

is well approximated by a zero-mean unit-variance Gaussian [16], with an associated upper tail probability $p(z) = \frac{1}{2}\mathrm{erfc}(z/\sqrt{2})$. For smaller n the exact binomial tail probability should be used. When the sequence comes from an irreversible source, there will typically be a larger fraction of same-direction matches, hence positive $\hat{\theta}$. Observing $\hat{\theta}>0$ with corresponding $p(z)<\alpha$ implies a rejection of time symmetry with the given level of significance. This test is one sided since irreversibility should [17] increase n_s relative to n_d.

III. PERFORMANCE ON VARIOUS DATA SETS

The quality of any statistical test is governed by two issues: how close the actual distribution matches the assumed null distribution with data from the null class, and how well the test is able to detect violations of that null. In particular, the null hypothesis of the time-symmetry test is flagrantly composite, encompassing a wide variety of reversible symbol streams. The justification for the test procedure is intuitively appealing—that compression automatically yields independent segments—but admittedly not rigorously proven. The success of this assertion is tested empirically by computing the statistic on ensembles of data sets taken from inputs known to be statistically reversible. Take an ensemble of M data sets from a reversible data class and compute $\hat{\theta}_k$ and $p_k=p(z_k)$ for $k=1,\ldots,M$. If the data are reversible and the test assumptions are fulfilled, the p_k ought to be as if drawn from the uniform distribution on $[0,1]$, or equivalently, the empirical cumulative distribution of p_k, $C(p_k)$, ought to converge with increasing M to a straight line, plotting $C(p_k)$ versus p_k. Similarly, over ensembles the standard deviation of z ought to tend towards one in the null class.

We first demonstrate on seemingly trivial data, white independent symbols. Figure 1 shows results of Monte Carlo simulations on these data. As expected, there is no indication of time asymmetry in θ or z, and the standard deviation of z under the null is close to unity.

Next, we consider time-symmetrical dynamical data. These were generated from samples of the logistic map, $x_{n+1}=1-ax_n^2$ in a generic chaotic regime ($a=1.8$). By itself x_i is certainly time-asymmetrical chaotic dynamics. We take two independent samples of length N from the map, $x_{i;1}, x_{i;2}$, and form the mixture

TABLE I. For the ensembles in Fig. 2. Kolmogorov-Smirnov test p values comparing the observed distribution of p_k to the uniform distribution in $[0,1]$. Only the values for $A=3$ and $N=250,2500$ appear to be significant. These apparent rejections are spurious and disappear in a different ensemble, being 0.175 and 0.713, respectively. There is no significant evidence that the p_k are distributed nonuniformly, showing a good calibration of the statistic under this instantiation of the null hypothesis.

Alphabet	$N=250$	$N=2500$	$N=25\,000$
2	0.218	0.457	0.0569
3	0.00335	0.0103	0.645
4	0.0326	0.522	0.303
5	0.332	0.349	0.386
6	0.0383	0.148	0.407

$$y_i = x_{i;1} + \alpha x_{N-i;2}. \tag{5}$$

When $\alpha=1$ the time series y_i is statistically reversible by construction; lower values of α give increasingly irreversible data. Figure 2 shows results over ensembles of $M=200$ samples of the reversible time series, each symbolized with varying small alphabets with equal-probability histograms. The statistic shows no time asymmetry, and the distribution of p_k is statistically close to uniform (see Table I), which is desirable for a correct null test.

Figure 3 shows a sample of a time series and its power spectrum from an arbitrarily constructed linear, Gaussian, and hence time-symmetrical [10], stochastic process. The top panel of Fig. 4 shows summary results on ensembles measuring reversibility on sample time series of varying size, analogously to Fig. 2. For the larger data sets the standard deviation of z is near unity and distribution of p_k is uniform, but for the shortest data sets, $N=250$, the standard deviation of z is less than 1, i.e., there is somewhat of a central tendency in the p_k. What is happening here is that the training sets are so short (each 125 symbols) that the dictionary built from observations is not sufficiently good to remove visible correlation. This is not unexpected as dictionary compression learns with increasing data. The total number of phrase matches $n=n_s+n_d$ used in the statistic is very small, even being as low as 10–20 for some of the samples. Nevertheless, the test is only slightly conservative, and data from system would not be characterized incorrectly as irreversible. The lower panel shows results on the square of the same process. The stochastic time series, which has mean zero, is

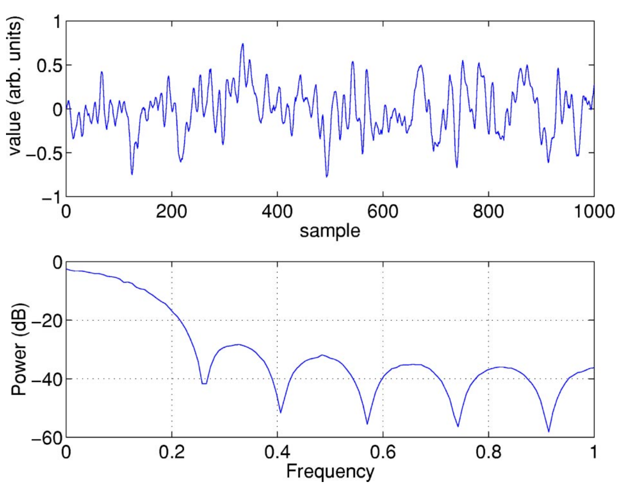

FIG. 3. (Color online) Top: sample time series from a discrete linear Gaussian process, constructed by a bandpass filter of an independent random Gaussian process. y axis is signal value (arb. units), x axis is sample number in integer-valued time. Bottom: power spectral density vs frequency (in units of the sampling frequency).

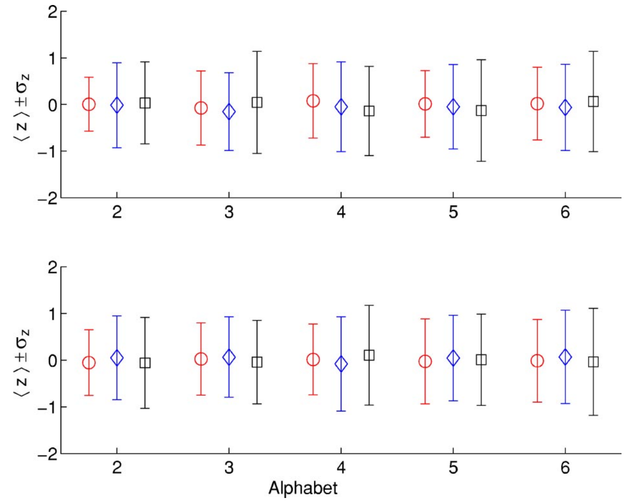

FIG. 4. (Color online) Summary statistics for linear Gaussian process, and square of that process. Top: $\langle z\rangle$ (arb. units) ± standard deviation for $N=250,2500,25\,000$ on linear process. Bottom: $\langle z\rangle$ (arb. units) ± standard deviation for square of that process, i.e., a nonmonotonic static nonlinear transformation of a reversible process.

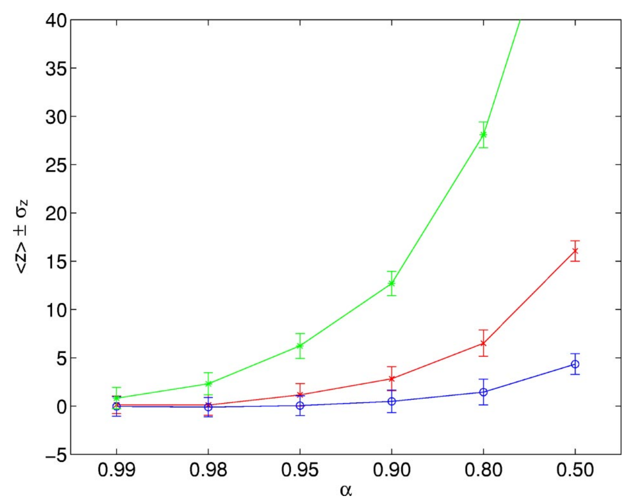

FIG. 5. (Color online) Time-asymmetry statistic z on $M=200$ sets of points from a mixture of logistic map time series. The x axis shows the mixing coefficient α ($\alpha=1$ is reversible) and y axis is $\langle z\rangle$ (arb. units) with bars displaying the sample standard deviation on the ensemble. Curves from bottom to top show $N=250$, $N=2500$, $N=25\,000$. Each data set was partitioned at $A=3$ with equal probability histograms.

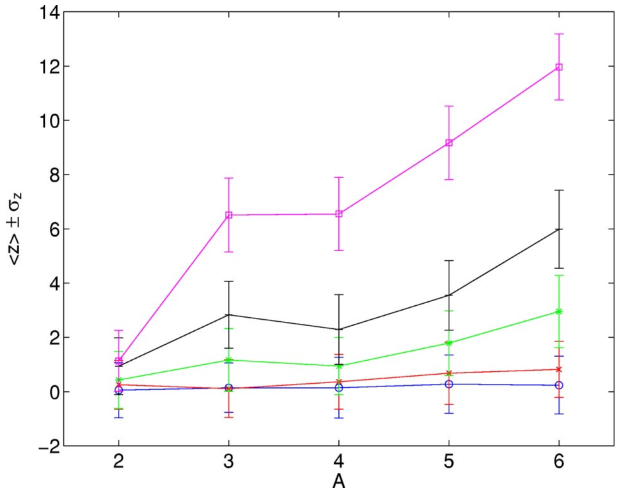

FIG. 6. (Color online) Time-asymmetry statistic z on $M=200$ sets of points of size $N=2500$ from a mixture of logistic map time series. The x axis shows the symbolization alphabet A and y axis is $\langle z\rangle$ (arb. units) with bars displaying the sample standard deviation on the ensemble. Curves from bottom to top show $\alpha=0.99,0.98,0.95,0.90,0.80$.

squared, and then symbolized with equal probability histograms. This is still a statistically time-symmetrical data set and thus in the null class. Surrogate data methods to detect nonlinearity typically can cope with only monotonic transformations of the observed variable, as they typically estimate the transformation to a Gaussian marginal distribution. As squaring is a nonmonotonic transformation, these data would reject the null with this sort of surrogate data method, but here the reversibility test correctly recognizes the data as being in the null class. The bottom panel of Fig. 4 shows no trend in z and standard deviation of unity across most parameters and data set sizes.

When examining data for signs of irreversibility it is often illuminating to plot z instead of θ as statistical significance can be seen easily with larger z. Figure 5 shows the detection of statistically significant reversibility with the mixture of

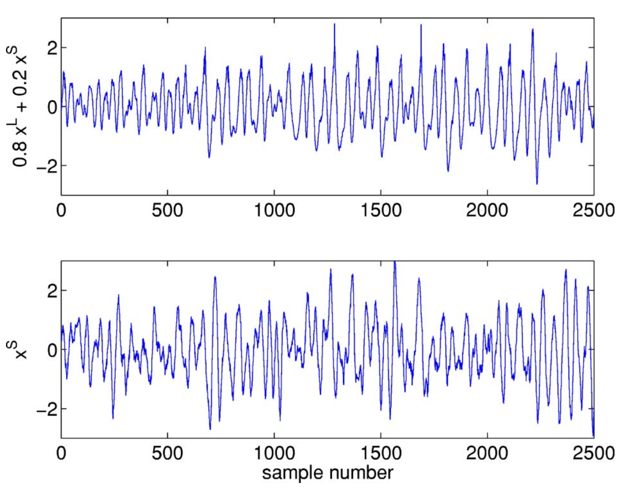

FIG. 7. (Color online) Sample of irreversible Lorenz mixture time series with $\alpha=0.2$ (top); sample of symmetrical surrogate data set (bottom). y axes are the signal value (arb. units) and x axes are discrete time.

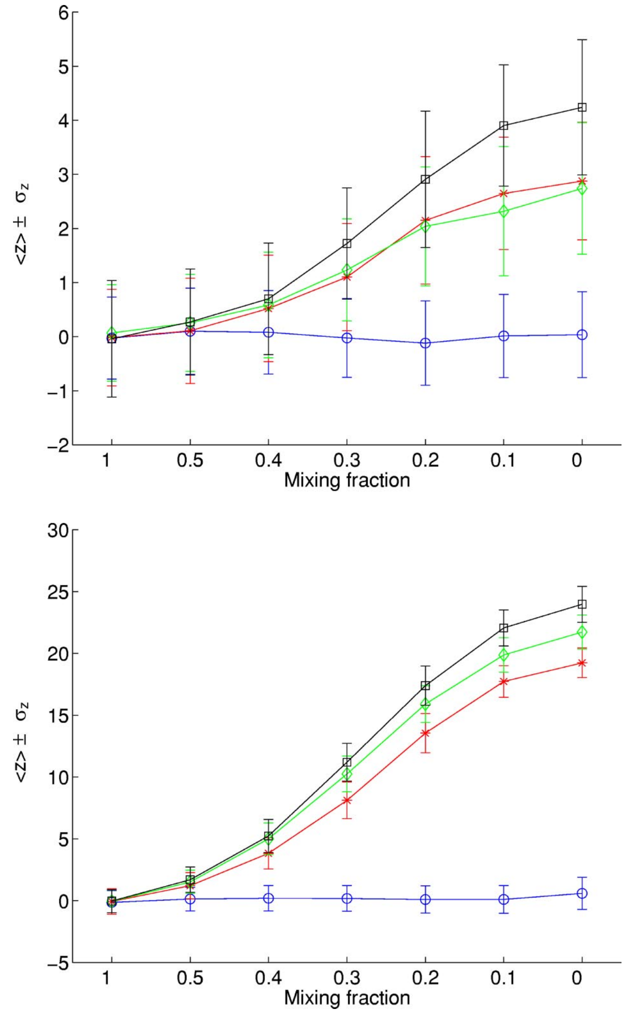

FIG. 8. (Color online) y axis: average z statistic (arb. units). x axis: mixing coefficient α as per Eq. (6), averaged over 200 replicas of size $N=2500$ (top) and $N=25\,000$ (bottom). Curves are for $A=2,3,4,5$ (blue circles, red stars, green diamonds, black squares).

forward and backwards logistic maps, Eq. (5). As expected, power to detect irreversibility increases with sample size and the degree of irreversibility. Figure 6 shows the effect of changing alphabets: with significant irreversibility, increasing alphabet size improved detecting it, but if irreversibility were minimal, the alphabet size was unimportant.

Now on to a more complicated system, the "Lorenz 1984" attractor: a tiny geophysical model with attractor dimension $d\approx 2.5$ [18]. The model is $dx/dt=-y^2-z^2-a(x-F)$, $dy/dt=xy-bxz-y+1$, $dz/dt=bxy+xz-z$, $a=1/4$, $b=4$, $F=8$. The x coordinate is sampled rather finely, every $\delta t=0.08$. To these samples were added white Gaussian noise of amplitude 5% of its standard deviation. There is substantial nontrivial autocorrelation of an arbitrary form. The sets tested for reversibility are mixtures of the dynamical data with surrogate data with identical power spectrum created in the ordinary way, by randomizing the phases of the discrete frequency-space representation and untransforming. Since that process by itself produces data with a typically Gaussian marginal density, the dynamical data [$x(t)$ sampled from the Lorenz model] are prewarped to have a Gaussian density as well.

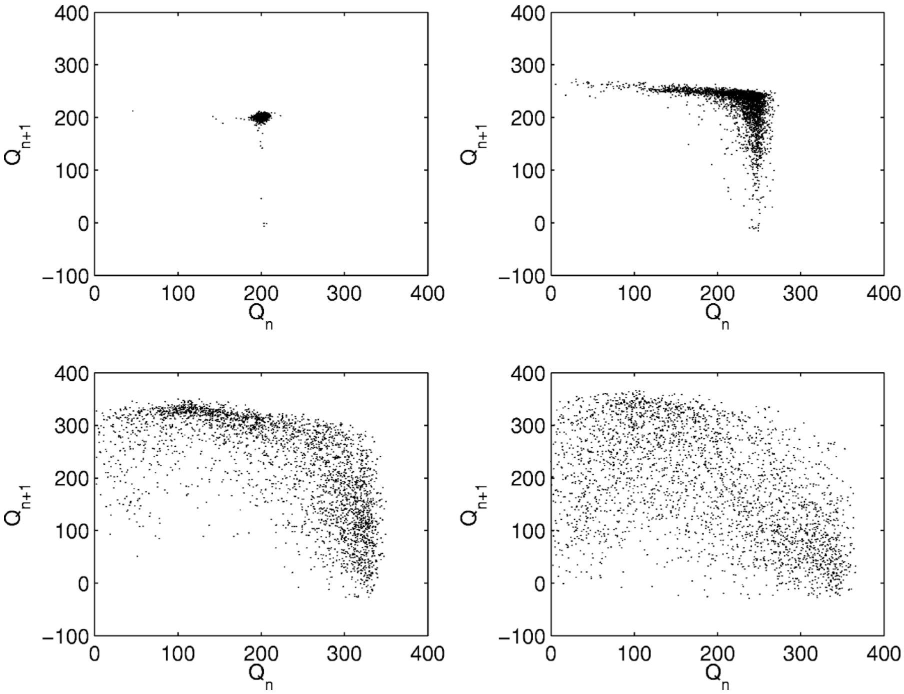

FIG. 9. (Color online) Time-return plots for spark-ignition data at four conditions of exhaust gas recirculation (EGF). x axis, heat released at cycle n (arb. units); y axis, heat released at cycle $n+1$ (arb. units). Upper left to lower right, EGR (%) $=0,16.0,22.2,24.7$.

The Gaussianized dynamical data are used both for the mixture and as the input for the surrogate to make the procedure the fairest. The mixture data sets, parametrized by α, are

$$y_i = (1-\alpha)x_i^L + \alpha x_i^S, \tag{6}$$

with x^L the Gaussianized, noised, Lorenz series, and x^S one of its surrogates. Figure 7 illustrates two example time series: a mixture ($\alpha=0.2$) and a symmetrical surrogate. Figure 8 shows the increasing detection of significant reversibility with decreasing α. Note that for $\alpha=1$ (all surrogate), the average z statistic (like θ, not shown) is zero with unit standard deviation, meaning that it is consistent with the assumed null distribution, as expected. Also, $A=2$ consistently permits no clear detection of irreversibility, which may reflect a particular symmetry in the system.

IV. EXPERIMENTAL EXAMPLE

We apply the reversibility tests to data from two combustion engines, one spark ignition, the other a Diesel cycle. The observed time series is the total heat released per cycle measured in a single fixed cylinder. The experimental apparatus maintained a constant speed and hence periodicity of engine. Fluctuations in the heat released may reflect turbulence in the cylinder, variations in air-fuel ratio due to residual gas ef-

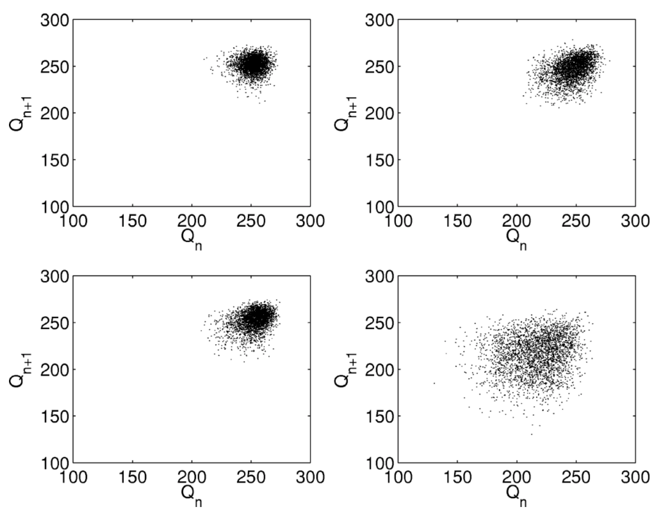

FIG. 10. (Color online) Time-return plots for spark-ignition data at four conditions of exhaust gas residual fraction (ERF). x axis, heat released at cycle n (arb. units); y axis, heat released at cycle $n+1$ (arb. units). Upper left to lower right, ERF (%) $=0,35,45,50$.

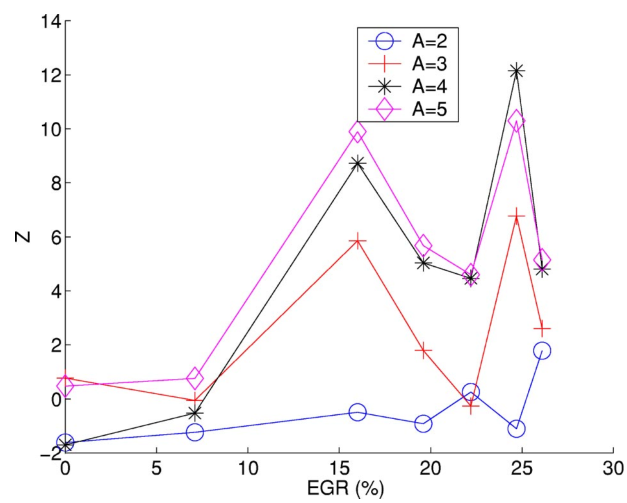

FIG. 11. (Color online) Time-asymmetry statistic for a time series of heat releases from a spark-ignition internal combustion engine. x axis, exhaust gas recirculation (%); y axis, reversibility statistic (arb. units).

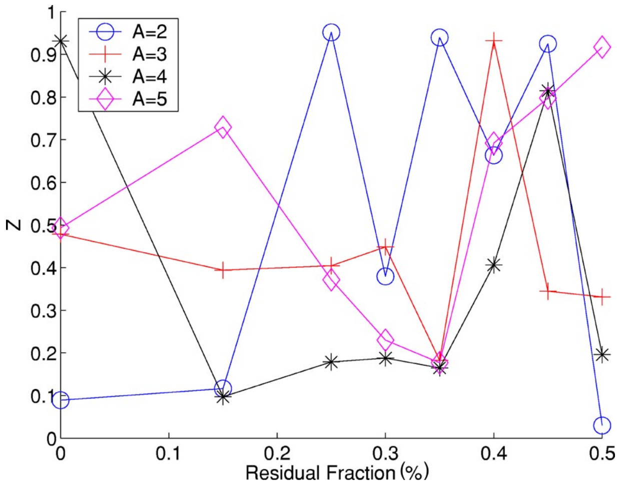

FIG. 12. (Color online) Time-asymmetry statistic for a time series of heat releases from a Diesel-cycle internal combustion engine. x axis, residual gas fraction (%); y axis, reversibliity statisic (arb. units).

fects, variation in intake air dynamics, among other reasons. For the spark-ignition engine, the input air-fuel ratio was maintained in stoichiometric conditions, but the proportion of exhaust gas recirculation (EGR) was altered for various runs and was the principal experimental parameter. Figure 9 shows time-return plots of example data in various conditions of EGR. For the Diesel data, the fraction of residual gas remaining from one combustion cycle to the next was estimated with changes in experimental parameters and is the effective experimental parameter. Figure 10 shows some time-return plots.

In all cases, the data were symbolized with equal-weight histograms of varying small alphabets. The reversibility results for the spark ignition data are shown in Fig. 11. There is a clear trend toward highly statistically significant irreversibility with EGR above 15%. Despite a large amount of noise, some form of deterministic nonlinear dynamics is a plausible explanation for the cycle-to-cycle variability. This is consistent with previous observations [9,20] where very similar irreversibility and bifurcations were observed with changing input air-fuel ratios. Diesel reversibility data are shown in Fig. 12. By contrast here, there is no statistically significant evidence of irreversibility over the entire parameter range, and thus one may conclude that the data could likely be generated by an effectively high-dimensional linear stochastic process. (Symmetrical low-dimensional chaos is unlikely given the high apparent noise level.) Physically what is most likely is that this dynamics is dominated by sufficiently high-dimensional turbulent fluctuations that globally averaged quantities such as the one considered here are effectively indistinguishable from linear processes by some kind of central limit theorem effect.

V. REVERSIBILITY AND PRESENTATION OF PROCESSES

We demonstrate an interesting and somewhat surprising phenomenon by looking at two simple symbolic dynamical systems, both first order Markov chains on a ternary alphabet. A first-order Markov chain on discrete symbols can be represented by a transition matrix M_{ij} for the transition probability from state i to j. Assuming that it is irreducible, its stationary probability μ is the left eigenvector with unit eigenvalue, $\mu=\mu M$. The transition matrices are

$$\mathbf{M}_1=\begin{bmatrix}0 & 1/3 & 2/3\\ 1/5 & 4/5 & 0\\ 1/10 & 0 & 9/10\end{bmatrix},\quad \mathbf{M}_2=\begin{bmatrix}0 & 1/3 & 2/3\\ 1/4 & 0 & 3/4\\ 2/10 & 0 & 8/10\end{bmatrix}$$

The transition matrices here were chosen arbitrarily; the only substantial difference between the systems is that $\mathbf{M}_1$ is statistically time-symmetrical and $\mathbf{M}_2$ is not. A Markov chain is time symmetrical if and only if the matrix $Q_{ij}=\mu_i M_{ij}$ is symmetric, i.e., $Q_{ij}=Q_{ji}$.

Consider the symbolic process where the index of the new state (in the ternary alphabet) is emitted for each transition. As expected, the distinction between the two chains is reflected in the empirical reversibility statistic. Ensembles of data from chain $\mathbf{M}_1$ accept the null, and those from $\mathbf{M}_2$ emphatically reject the null (see Table II).

TABLE II. Reversibility statistics on ensembles of size 200 of ternary $A=3$ symbols from Markov chains. $\mathbf{M}_1$ is symmetrical, and the results are consistent with the null distribution (though the variance is a bit too small for $N=250$). Data from the irreversible $\mathbf{M}_2$ emphatically reject the null, shown by substantial $\langle\theta\rangle,\langle z\rangle$ and the count of rejections at the $p<0.01$ level out of the ensemble of 200 data sets.

System	$\langle\theta\rangle\pm\sigma_\theta$	$\langle z\rangle\pm\sigma_z$	Rejections
$\mathbf{M}_1$, $N=250$	0.0022 ± 0.107	0.017 ± 0.65	0
$\mathbf{M}_1$, $N=2500$	-0.0022 ± 0.049	-0.035 ± 0.89	2
$\mathbf{M}_1$, $N=25000$	-0.00092 ± 0.020	0.045 ± 0.96	0
$\mathbf{M}_2$, $N=250$	0.26 ± 0.150	2.15 ± 1.22	90
$\mathbf{M}_2$, $N=2500$	0.40 ± 0.055	8.89 ± 1.26	200
$\mathbf{M}_2$, $N=25000$	0.53 ± 0.021	31.7 ± 1.24	200

TABLE III. As in Table II, except that now the alphabet is $A=2$, with a zero symbol emitted when the first allowable transition from each state is taken and a one when the other transition is taken. What was reversible in the explicit Markov representation ($\mathbf{M}_1$) now shows increasing evidence of irreversibility, whereas the previously patently irreversible process $\mathbf{M}_2$ is now apparently reversible.

System	$\langle\theta\rangle\pm\sigma_\theta$	$\langle z\rangle\pm\sigma_z$	Rejections
$\mathbf{M}_1$, $N=250$	0.0073 ± 0.109	0.033 ± 0.68	0
$\mathbf{M}_1$, $N=2500$	0.019 ± 0.051	0.34 ± 0.92	4
$\mathbf{M}_1$, $N=25000$	0.059 ± 0.022	2.86 ± 1.08	139
$\mathbf{M}_2$, $N=250$	0.0052 ± 0.11	0.034 ± 0.81	0
$\mathbf{M}_2$, $N=2500$	-0.0069 ± 0.050	-0.011 ± 0.96	3
$\mathbf{M}_2$, $N=25000$	0.00023 ± 0.018	0.013 ± 1.01	1

Now consider a hidden Markov-chain presentation of this same process. At each time step, instead of emitting the index of the new state, emit a zero or a one, depending on which transition of nonzero probability has taken place. In the language of theoretical symbolic dynamics [19], the explicit-state version is a presentation of a "vertex shift," as a symbol is emitted corresponding to each new vertex of the transition graph which is visited, and hence explicitly a shift of finite type (with memory 1) on a three-symbol alphabet. The implicit-state version is a sofic shift with an associated graph and labeling: a distinct binary symbol is emitted depending on which edge is taken on the transition. (These particular shifts are also of finite type, but not all sofic shifts are finite type).

The Shannon entropy rates $[h_S(\mathbf{M}_1) \approx 0.5623 \text{ bits/symbol}, h_S(\mathbf{M}_2) \approx 0.7602 \text{ bit/symbol}]$ of the two representations are identical, as there is the same amount of uncertainty about the next state and the same invariant density. Moreover, their topological entropies (1 bit/sym) and minimum periods are also all identical, implying an "almost conjugacy" between them (Ref. [19], Chap. 9). Roughly, this means that bi-infinite sequences in the shift spaces may be mapped one-to-one into each other except for sequences of vanishing probability.

Despite the topological equivalence, the presence or absence of probabilistic reversibility in the symbolic sequences becomes reversed by the change in presentation. Table III shows the evidence. The Markov process described $\mathbf{M}_1$, which is *prima facie* reversible in the explicit representation, now shows increasing statistical evidence of irreversibility with larger data set sizes in the implicit-state presentation. $\mathbf{M}_2$, which was patently irreversible explicitly, now shows no evidence for irreversibility whatsoever. One conclusion is that reversibility or irreversibility depends on the nature of the observed variable: although invariant to *static* functional transformations of an observable, it is not invariant to a change from an explicit to hidden Markov chain. In this case, the transformation from explicit to hidden Markov representation would require successive *pairs* of observed states in successive times (a sliding-block code) in order to generate the 0/1 symbol of the hidden version time series, i.e., the transformation is not a static function of the current state. In continuous space, there could be an analogous effect: for example, the time series formed from successive differences of a reversible—but non-Gaussian—process could display irreversibility. Another conclusion is that topological equivalences of the shift spaces generated by the symbolic processes do not necessarily carry over to probabilistic metric quantities such as statistical irreversibility as considered here.

VI. CONCLUSIONS

We have demonstrated a statistic to distinguish between statistically time-symmetrical and time-asymmetrical data-generating processes from an observation of their output, a sufficiently long data set. The data must be symbols of a discrete alphabet, preferably of rather small size—this symbolization could be from a discretization of a continuous valued process. Dictionary-based data compression methods provide the inspiration and technology for a scheme which will adaptively and automatically account for generic forms of dependence. This justifies a classical direct null test. Given the symbol stream, there are no free parameters.

There is one minor caveat. If the entropy of the input symbols is *extremely* low, for instance, very long repeats of identical symbols (say by symbolizing a very oversampled data set), then the statistical calibration of the null may be imperfect. The dictionary compression procedure is known to be suboptimal for those systems: the codelength per symbol of the compressed output would be well above its true entropy rate. The algorithm appends only one symbol at a time to each dictionary entry to form new dictionary entries, thus the phrases it finds are not sufficiently long to have excellent compression. For our purposes the successive matches would not be quite as independent as they should be, and the calibration of the null distribution is imperfect, e.g., the standard deviation (over ensembles) of z may be larger or smaller than one. It does take rather extreme data for this to be an issue, and it is nearly always simple to rectify by using different symbolization or by undersampling the input data set appropriately. Time reversibility will not be influenced by such a change. A typical rule of thumb may be to be suspicious of data whose entropy rate is less than one-tenth the maximum, i.e., $\log_2(A)$ bits per iteration.

Complete source code in C++ for the algorithm is available in the EPAPS archive accompanying this manuscript [22].

ACKNOWLEDGMENTS

The author wishes to thank C. E. A. Finney and R. M. Wagner for making available the combustion engine data sets.

[1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Physica D **58**, 77 (1992).
[2] M. Kennel and S. Isabelle, Phys. Rev. A **46**, 3111 (1992).
[3] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).
[4] D. Kugiumtzis, Int. J. Bifurcation Chaos Appl. Sci. Eng. **11**, 1881 (2001).
[5] P. Rapp, C. Cellucci, T. Watanabe, A. Albano, and T. Schmah, Int. J. Bifurcation Chaos Appl. Sci. Eng. **11**, 983 (2001).
[6] T. Schreiber and A. Schmitz, Physica D **142**, 346 (2000).
[7] D. Kugiumtzis, Phys. Rev. E **60**, 2808 (1999).
[8] C. Diks, J. C. van Houwelingen, F. Takens, and J. DeGoede, Phys. Rev. Lett. **201**, 221 (1995).
[9] C. S. Daw, C. E. A. Finney, and M. B. Kennel, Phys. Rev. E **62**, 1912 (2000).

[10] G. Weiss, J. Appl. Probab. **12**, 831 (1975).

[11] J. Ziv and A. Lempel, IEEE Trans. Inf. Theory **24**, 530 (1978).

[12] It is a “prefix dictionary”: for any codeword $w=s_1 s_2 \dots s_L$ in the dictionary, all prefixes of w, e.g., $s_1 \dots s_j, \forall j < L$ are also in the dictionary. Parsing is greedy: search for increasingly long matches of the next input, and when the first such word of length $L+1$ fails to exist in the dictionary, the most recent match of length L is the codeword. The dictionary can be efficiently implemented as a prefix tree giving high performance.

[13] T. Cover and J. Thomas, *Elements of Information Theory* (Wiley Interscience, New York, 1991).

[14] N. Merhav and J. Ziv, IEEE Trans. Inf. Theory **39**, 1280 (1993).

[15] D. Benedetto, E. Caglioti, and V. Loreto, Phys. Rev. Lett. **88**, 048702 (2002).

[16] $\mathrm{sgn}(\cdot) = -1, 0, 1$ for negative, zero, and positive arguments.

[17] Statistically significant $\hat{\theta} < 0$, i.e., small $1-p$, may imply some kind of unmodeled nonstationarity. In that case, the data set ought to be broken up into more, shorter, interleaved training and test sets, accumulated and repeated.

[18] E. N. Lorenz, Tellus, Ser. A **36A**, 98 (1984).

[19] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding* (Cambridge University Press, Cambridge, England, 1995).

[20] J. B. Green, C. S. Daw, J. S. Armfield, C. E. A. Finney, R. M. Wagner, J. A. Drallmeier, M. B. Kennel, and P. Durbetaki, SAE Paper No. 1999-01-0221 (1999); C. S. Daw, M. B. Kennel, C. E. A. Finney, and F. T. Connolly, Phys. Rev. E **57**, 2811 (1998).

[21] In common computer practice the class of compression software generally known as “zip” uses dictionary-based universal compression methods.

[22] See EPAPS Document No. E-PLEEE8-69-140404 for the complete source code in C++. A direct link to this document may be found in the online article’s HTML reference section. This document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or from ftp.aip.org in the directory /epaps/. See the EPAPS homepage for more information.