Testing for general dynamical stationarity with a symbolic data compression technique

Matthew B. Kennel
Institute for Nonlinear Science, University of California, San Diego, La Jolla, California 92093-0402

Alistair I. Mees*
Isaac Newton Institute, Cambridge University, Cambridge CB3 0EH, England

(Received 2 August 1999)

We construct a hypothesis test for examining the *stationarity* of the evolution law for a time series of discrete symbols: whether two data streams appear to originate from the same underlying, but unknown, dynamical system. Based on techniques from the theory of data compression, our method intelligently accounts for the substantial serial correlation and nonlinearity found in realistic dynamical data. We demonstrate the method on a number of realistic experimental datasets.

PACS number(s): 05.45.Tp, 05.10.−a

Symbolic methods have been used in the study of dynamical systems from the earliest days, most notably Kolmogorov and Sinai's [1] use of metric entropy as a dynamical invariant, which spawned a significant mathematical industry in symbolic dynamics. Fraser [2] applied information theoretical concepts to construct useful algorithms and criteria for time-delay embeddings.

Stationarity, the notion that the dynamical law describing the system does not change over long timescales, is a prerequisite for the vast majority of nonlinear data analysis techniques. Only recently have hypothesis tests suitable for realistic chaotic and nonlinear dynamical data been proposed [4]. In this contribution we advocate a symbolic approach that exploits well-studied and powerful techniques of data compression, infrequently applied in the physics literature [3], in order to justify the statistical assumptions in a deeply principled way.

We have a stream of symbols, either quantized from continuous-valued observations or directly measured, $s_1, s_2, s_3, \dots, s_N$, each symbol from some alphabet A reexpressed as integers $s \in \{1,2,\dots,|A|\}$. The distribution of multisymbol words provides information about time-dependent structure and correlation, just as, with continuous nonlinear data, time-delay embedding provides a vector space revealing dynamical information. We emphasize that for our method, the symbolic encoding need not be crafted intelligently or with regard to eludicating topological properties, as, for instance, is typically done in theoretical symbolic dynamics studies when the equations of motion are known. In our tests, symbolization is merely a radical discretization or reduction in precision of the original data, which typically are observed at a much higher digital precision. For example, we divide the observed one-dimensional histogram to a small number of bins, either by equal probability mass or by equal width, and then code each datum by the bin number. Naturally, any reasonable preprocessing strategy deemed suitable for the particular data set, e.g., detrending or symbolizing points projected on an illuminating Poincaré cut instead of at arbitrary time intervals, will only improve the security of the results.

A first attempt at a stationarity test might be to apply the classical χ^2 test to observed counts of distinct multisymbol words observed in, say, the front and back halves of the data. Unfortunately, the assumption underlying this test—that each datum is randomly and *independently* drawn from some distribution—is not true in realistic dynamical data. Short time correlations in physical data strongly couple symbols near in time; thus naive application of such tests fails miserably. Indeed, arbitrary dynamical dependence makes it difficult to construct a proper statistical null test for any hypothesis that allows chaotic or general nonlinear data in the null class, and to our knowledge, few examples of this sort exist.

This work proposes a test procedure that quantifies whether two observed symbol streams have ''the same dynamics,'' even in the presence of serial correlation and dependence. There are two phases to the algorithm: construction of a symbolic predictive model, and the evaluation of a combination of classical statistics, this time on data constructed to be nearly independent. The algorithm is computationally rapid and does not require Monte Carlo simulation.

One traditional definition [5] of a stationary stochastic source is that the joint probability distribution of multisymbol words of random variables X is invariant to global shifts in time:

$$p(X_i, X_{i-1}, \dots, X_{i-k}) = p(X_{i+N}, X_{i-1+N}, \dots, X_{i-k+N}), \forall N, k. \quad (1)$$

The technology we employ does not enable us to easily answer that question, but a slightly different one regarding the time invariance of the predictive conditional distribution, i.e.,

$$p(X_i | X_{i-1}, \dots, X_{i-k}) = p(X_{i+N} | X_{i-1+N}, \dots, X_{i-k+N}), \forall N, k. \quad (2)$$

Physically, this is asking whether the evolution law changes over time. Typically that would occur with slow changes in underlying physical parameters (''drift'') as opposed to the

*Permanent address: Centre for Applied Dynamics and Optimization, The University of Western Australia, Nedlands, Perth 6907, Western Australia.

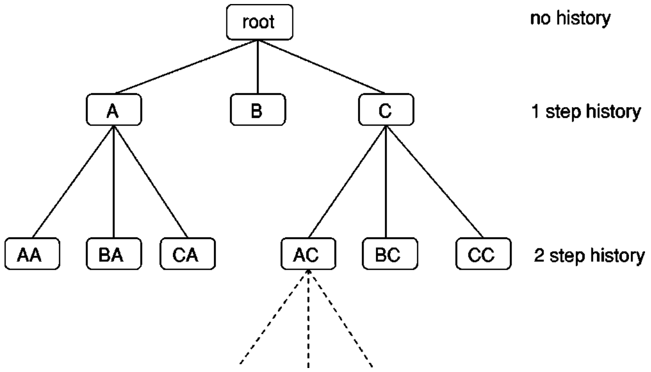

FIG. 1. Example of a small context tree for a three symbol alphabet. Internal nodes (nodes with deeper children) are the root node, A, C, and AC, and terminal nodes, AA, BA, CA, B, BC, CC. Descendents of AC continue off the figure. Each node accumulates counts of future symbols and two internal code lengths.

sort of nonstationarity ascribed to Brownian motion processes or Lévy flights, i.e., long-memory processes. Given a time series with substantial autocorrelation even at very long lags, one might consider applying the test to a differenced version of the observed series to see whether or not the observed data will arise from the integral of a time-invariant process.

Fundamental results of information theory [6,5] require that optimally compressed output data approach independence. This is the central theoretical justification for our subsequent application of classical statistical inference, and we feel, the most useful concept outside the specific application presented here. The strategy is to perform the modeling stage of a modern symbolic data compression algorithm, but instead of subsequently emitting a compressed binary stream, we examine the statistics observed in the model to discern stationarity.

Our symbolic dynamical model is a ‘‘context tree:’’ as shown in Fig. 1, the recent symbols in the stream themselves define the state, known here as the *context*; contexts are analogous to the states reconstructed by time-delay embedding in conventional nonlinear dynamical analysis. We describe elsewhere [7] other applications to nonlinear dynamics but in the present paper we specialize in stationarity testing.

The tree structure accumulates the statistics of observed symbol vectors down to some maximum depth d, with distinct branches corresponding to distinct symbols of alphabet A which occurred at prior times. The top node corresponds to no history, the first $|A|$ nodes correspond to a one-dimensional reconstruction of the most recent symbol, their $|A|^2$ descendents to a two-dimensional reconstruction of the two most recent symbols, and so forth. Naturally one only constructs the nonempty nodes. Each node stores $|A|$ integers that record the occurences c_j of every symbol that occurred immediately after its particular context C, which provides an estimate of the conditional probability of seeing the various symbols after the current context $\hat{P}(s_{t+1}|C)$. Following the data compression literature, we use the Krichevsky-Trofimov [10] estimator,

$$\hat{P}(k)=(c_k+1/2)\bigg/\sum_i (c_i+1/2)$$

for a symbol with value $1\leqslant k\leqslant|A|$. Given a particular history, we have an estimator of the conditional probabilities of the symbols that have followed that history.

More than one node matches the current history. For example, in a three letter alphabet, if the four most recently emitted symbols are $bcaa$, matching nodes have contexts a, aa, caa, and $bcaa$ in addition to the zero-history (root) node λ, which always matches by definition. The critical issue is balancing between the more detailed dynamical reconstruction possible in deeper contexts, and the increased quantity of observations at shallower contexts, which gives greater robustness against statistical fluctuation.

Estimating $\hat{P}$ at fixed depth contexts, independent of symbol history, is a fixed-order Markov model, which is a poor performer in general. Our algorithm selects a single node to use based on a sequential minimum description length criterion [8] selecting those nodes that in the past have encoded their particular future symbols using few bits. In a context tree, the number of past symbols that contribute to predicting the future is not uniform. Some past histories need to be deeply examined because there, long-past symbols influence the future significantly, whereas for other past histories, there is less need, either because future evolution is more unpredictable or there has been little information previously observed regarding those histories. In this respect, contexts are like ‘‘variable embeddings’’ [11].

At time t we attempt to code symbol $s(t+1)$, using the prior symbols $s(t)$. The algorithm has four phases, which must be performed in this order: selection of the special encoding node, creation of new child nodes, update of local code lengths, and the increment of the conditional counts. In addition to this set of counts c_k, each node accumulates two additional internal code lengths, L_s and L_p, L_s representing the cumulative cost to encode symbols whose contexts matched the current node, employing the current node's counts, c_k, L_p representing the code length encoding the same data but estimating the future using the c_k in the parent node. A sequential data compressor must not use any information whatsoever about the identity of $s(t+1)$ to choose a probability estimate to encode it, otherwise, a causal decompressor is impossible. In our application, it would mean that the code length cost representing the complexity of the model would not be properly included, and we would be susceptible to spurious overfitting. Batch encoders for tree sources also exist and could be applied to our problem, but those require explicit penalty terms for the model complexity that we have found to be difficult to analyze consistently.

At each time step, we traverse the nodes that match the current context from the root node down. If the tree is as Fig. 1, and the recently seen symbols are $ABBBC$ (C being the most recent), the nodes in the tree that match the current context are the root node (matches everything), C and BC. The current node n is the ‘‘encoding node’’ for the next symbol (in the example, coming after C) if it is a terminal node (i.e., if it has no deeper children whatsoever), or barring that, if

$$\sum_c L_p(c)\leqslant\sum_c L_s(c), \tag{3}$$

with c summing over all extant children of n. Otherwise, one descends one level deeper to the matching child and repeats. (A blank node is created here if necessary.) The notion is that we wish to find nodes at the level where, heretofore, the current counts better predicted (via a smaller code length) the future vs descending to a deeper level.

If we were literally compressing the data, we now feed the best predictive estimate of the next symbol $\hat{P}$ (estimated using the current encoding node), and additionally the actual symbol $s(t+1)$ to a standard algorithm called an ‘‘arithmetic coder’’ that emits compressed bits. The decompressor sequentially reconstructs the same estimator by an identical tree method, and given the compressed bits, the inverse of the arithmetic coder reproduces the actual symbol $s(t+1)$. This final stage is not relevant for our present needs, but we mention it here to show the relationship between on-line modeling of the symbolic dynamics and data compression.

We now add new blank nodes to the tree if necessary to match the current context of symbols as deeply as possible. Then, for all matching nodes, including possibly newly created blank nodes, the local quantities L_s and L_p are incremented by $-\log_2 \hat{P}[s(t+1)]$, using the node's own counts for L_s and the parents' counts for L_p. The parent of the root node is defined to always have a uniform distribution $\hat{P} = 1/|A|$. Finally, at all matching nodes, the appropriate count c_k, corresponding to the value of $s(t+1)$, is incremented by one.

After encoding all the symbols using the combined set we carry out the stationarity test. Answering the question, do two data sets appear to arise from the same underlying dynamical system, translates to combining hypothesis tests performed at each encoding node regarding whether the distribution of future symbols actually encoded—whether from the first set or the second—could have come from a single underlying probability distribution, and if any apparent difference is statistically significant. At encoding contexts, we may use standard tests because these events ought to be nearly independent using a good compression algorithm.

Every node records the frequency with which symbol k was encoded there, $e_{k;1}$ in the first set and $e_{k;2}$ in the second. (Note that $e_k \neq c_k$, the latter accumulating frequencies whenever a context was excited.) Assuming independence, the statistic

$$\chi^2 = \sum_{k=1}^{|A|} \frac{(R^{1/2} e_{k;1} - R^{-1/2} e_{k;2})^2}{e_{k;1} + e_{k;2}} \tag{4}$$

with $R = \Sigma e_{k;2} / \Sigma e_{k;1}$ follows the standard χ^2 distribution with $|A| - 1$ degrees of freedom under the null hypothesis that both empirical probability distributions came from the same underlying distribution [12]. Given the value of χ^2 and the degrees of freedom, standard numerical algorithms provide a likelihood L asymptotically uniform $L \in (0,1)$ under the null. Small values of L reject the null at the given significance level, e.g., Ł<0.01.

It is known that the analytic approximation used for the asymptotic distribution of the χ^2 statistic becomes increasingly inaccurate as the number of observations decreases. Thus for $\Sigma e_k < 75$ (a somewhat arbitrary cutoff) we switch over to a combinatorial test for differences in proportions, called *Fisher's exact test*. The calculations for this test are easy only in the 2×2 case. We coalesce bins by keeping the observation for the most frequent symbol [bin m which achieves $\max(e_{m;1} + e_{m;2})$] and merging the others into $e_{o;1}, e_{o;2}$, resulting in four quantities conventionally expressed in a ‘‘contingency table,’’ with cumulative row and column sums:

$e_{m;1}$	$e_{o;1}$	n_1
$e_{m;2}$	$e_{o;2}$	n_2
n_m	n_o	N

Under the null that the difference in proportions between m and o counts is independent of being in set 1 and 2, the probability for seeing any particular table with the given marginal sums is

$$p_T = n_m! n_o! n_1! n_2! / (e_{m;1}! e_{m;2}! e_{o;1}! e_{o;2}! N!).$$

One directly enumerates all tables with the given observed marginals (only a one-degree of freedom for a 2×2 table) and sums p_T for every table with a difference in proportions at least as great as that observed [13], resulting in a likelihood L for accepting the null hypothesis at this node.

We combine these M likelihoods, each measuring some aspect of of the same null hypothesis, into a single overall test. Under the null, the quantity

$$X^2 = \sum_{k=1}^{M} (-2 \ln L_k) \tag{5}$$

is χ^2 distributed with $2M$ degrees of freedom, from which we compute our final $\mathcal{L}$, again uniform in (0,1) under the null. Especially small values of $\mathcal{L}$ imply a small likelihood that this level of difference would have been observed had the two symbol datasets been generated by the same underlying dynamical process. This completes our desired test procedure.

We first test the accuracy of the statistic under the null. We produced an ensemble of 1000 time series from the x coordinate of the ‘‘Lorenz 84’’ attractor: a tiny geophysical model with attractor dimension $d \approx 2.5$ [14]. This system is higher dimensional and more complex (see Fig. 2) than the traditional Lorenz dataset, and is thus a somewhat more stringent test. Figure 3 shows the distribution of $\mathcal{L}$ comparing the first and second halves of each set, demonstrating $\mathcal{L}$ is close to uniform $\in (0,1)$. This is a stringent requirement and shows the success of our independence assumption, as it is difficult to get a high-quality null distribution with complicated, arbitrarily correlated, chaotic data in the null class. With this number of data, the test is also quite powerful.

We demonstrate discrimination power with a set of pressure data from an experimental model of a ‘‘fluidized bed reactor’’ [15]. This experimental system consists of a vertical cylindrical tube of granular particles excited from below by an externally input gaseous flow. In some regimes (‘‘slugging’’), the particles exhibit a combination of collective low-dimensional bulk dynamics and small-scale high-dimensional turbulence of the individual particles [15]. The observed variable was an azimuthally averaged pressure difference between two vertically separated taps. Figure 4 shows portions of time-delay embedding of orbit sections of

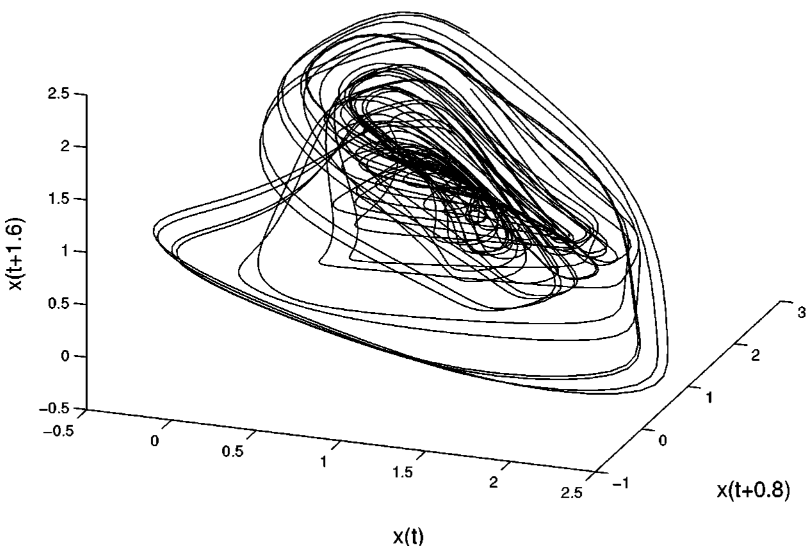

FIG. 2. Sample orbit of Lorenz dynamical system in reconstructed state space.

the dataset taken at the same experimental parameters, and one when the flow was boosted by 5%. The change in the attractor is rather subtle and difficult to reliably diagnose by eye. Figure 5 shows $\mathcal{L}$ on a data set whose flow was increased at the midpoint. As the alphabet size increased and the hypothesized breakpoint approached the true value of 50%, the strength of the rejection increased, $\mathcal{L}\rightarrow 0$. Even the binary alphabet case showed a significant rejection of the null. On data taken in stationary conditions $\mathcal{L}$ fluctuates randomly in $(0,1)$, as expected. We performed the same statistical test, with qualitatively identical results, on an experimental system whose flow rate was ramped slowly by the same amount. Even with this sort of data, the greatest rejection tends to occur in the middle of the dataset because the statistical discrimination power is greatest when there are equal numbers in the two sets considered in the test. We saw qualitatively identical results on a dataset whose flow was adiabatically boosted by the same degree during the run. This is not surprising, even though the test assumes a discrete breakpoint in order to lump the symbols into one class or another, as even given a smooth change, front and back sets will have different characteristics, and the ability to detect this (strongest rejection) will peak with approximately equal quantity of data in each set. Using a breakpoint test on smoothly changing dynamics might result in a small loss of statistical power compared to an ideal test, but in our experience with the proposed method on experimental datasets of at least a thousand points, adequate power to detect physically significant nonstationarity is rarely a concern with this method in our experience.

The southern oscillation index (SOI), the normalized pressure difference between Tahiti and Darwin, is a proxy for the El Niño southern oscillation, as ocean temperature influences atmospheric dynamics. The period from mid 1990 to 1995 exhibited an anomalously sustained period of El Niño-like conditions (Fig. 6), perhaps indicative of global climate change. One statistical analysis [16] found such an anomaly quite unlikely assuming stationarity, but another group [17] used a different analysis and found it significantly more likely to be a chance fluctuation. Both papers used traditional linear forecasting models, with the difference centered around an autocorrelation based correction for serial correlation to arbitrarily reduce the degrees of freedom. We applied our algorithm to the three-month moving average SOI (binary symbolized) testing the 5.4-yr period in question against the rest of the series (starting from 1900), with a resulting $\mathcal{L}\approx 0.01$, meaning that one would expect to see a region this anomalous by chance every 540 yrs. The result is closer to that of [17] than [16] but we certainly do not want to take any particular position regarding climate; rather, we wish to point out an application for our method where correcting for serial correlation automatically is useful.

We point out that the proposed method is not exclusively limited to testing or finding a single breakpoint—all that is needed is a sensible, *a priori* hypothesized division of the dataset into discrete multiple classes. For instance, one might want to test for the presence of cyclostationarity, that the dynamics are externally modulated at some slow frequency Ω. In this case, one could choose elements of set 1 and set 2 depending on whether $\sin(\Omega t+\theta)$ is positive or negative,

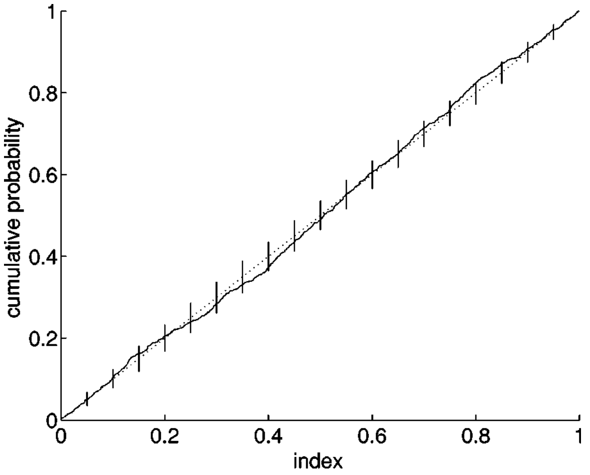

FIG. 3. Quantile-quantile plot of $\mathcal{L}$ under the null hypothesis. The observed values of $\mathcal{L}$ are sorted and plotted vs their normalized index $(i+1)/1001$. Asymptotically the curve should approach the diagonal under the null. Bars are ± two standard deviations for 100 samples of 1000 uniform deviates $\in[0,1]$ processed similarly.

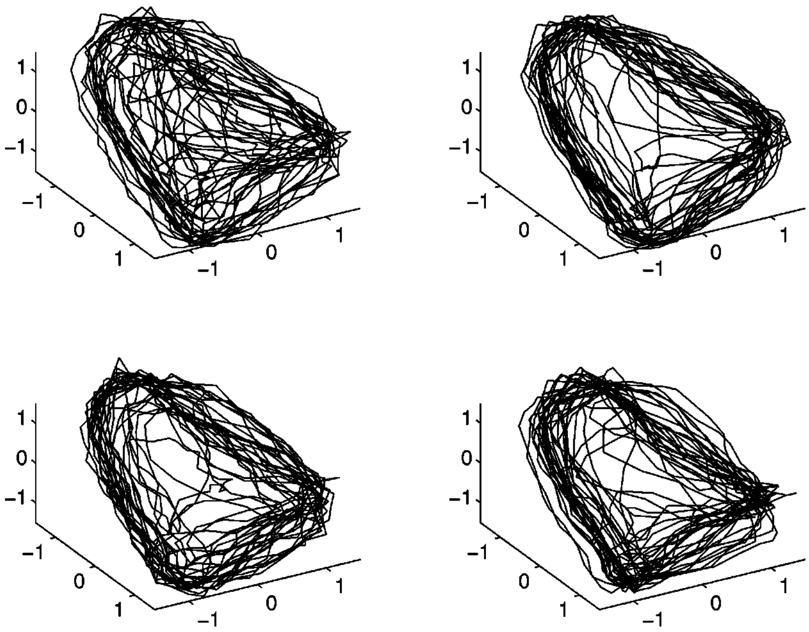

FIG. 4. Phase space plots of the differential pressure signal from a fluidized bed reactor. Three are from the same parameters, one is different.

given fixed Ω and θ. Here the hypothesis is that the dynamics are significantly different when trying to predict symbols on one part of the cycle compared to the other. Testing against three or more classes would require an upgrade of the χ^2 or Fisher test of proportions procedures; there exist conventional methods in the statistical literature.

Recent work has successfully used a distance in the symbolic space to fit unknown parameters of a physically motivated continuous model to observed data, including substantial observational and dynamic noise all in one framework, a challenge for traditional regression. Tang *et al.* [18] first proposed minimizing over free parameters the difference between an observed distribution of symbol words and that produced by discretizing some proposed model's output. Daw *et al.* [19] successfully employed this technique to fit experimental internal combustion engine measurements to a low-dimensional model. The optimization target was a Euclidean norm in [18] and a χ^2 distance in [19]. Due serial correlation, a true hypothesis test confirming the apparent compatibility of observed data to a well-fitting model was not possible in those works. We feel our current method could provide a less *ad hoc* optimization goal, e.g., maximizing average $\mathcal{L}$ or minimizing the code length of the physical model's output, encoded using the symbolic model learned from the observed data.

We conclude by reminding the reader that stationarity is a property of the *model* deemed useful for the dataset. As a result, any test for stationarity or time-invariance depends on

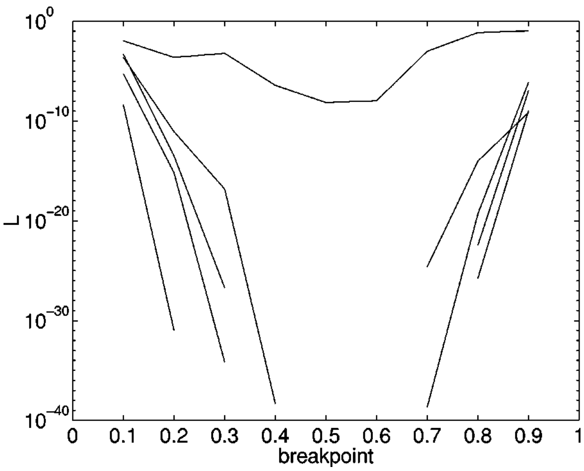

FIG. 5. Nonstationary fluidized bed results with air flow altered at the 50% mark. Plotted statistic $\mathcal{L}$ as a function of hypothesized breakpoint in time series and symbolic alphabet precision. Results for $|A|>2$ numerically underflowed to $\mathcal{L}=0$ toward the center and are not plotted. Null hypothesis emphatically rejected due to the very small values of $\mathcal{L}$.

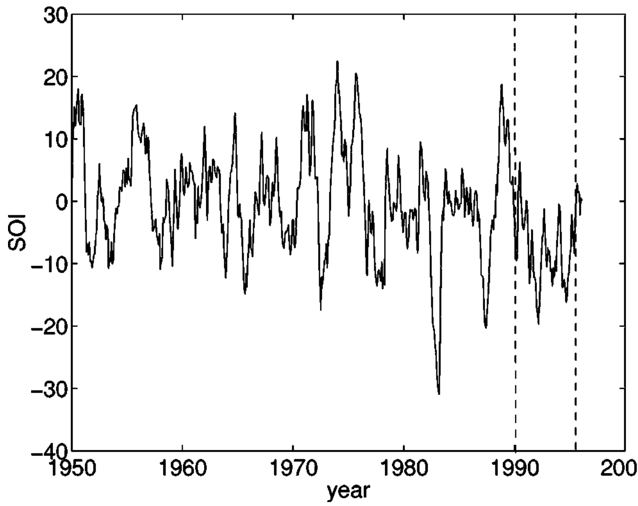

FIG. 6. Three month moving average of the southern oscillation index, the normalized pressure difference between Tahiti and Darwin, Australia. Strongly negative values correspond to El Niño events. It is debated whether the extended negative period from mid 1990 through 1995 is especially anomalous.

the appropriateness of the hypothesized model applied to the data. In our case, the symbolic methods are closely related to universal data compression algorithms [20]. The performance on finite-length sequences will, of course, depend on a good representation such that the predictabilility of the future symbol is best determined by the first context symbol, next best by the second context symbol, etc. If the method has to descend many symbols to find a useful predicting context, power will be lost. Because the predictive code length is a ‘‘fair’’ minimum description length measure, one can optimize over reconstruction parameters (e.g., time delay or symbolization breakpoints) to find the lowest code length, at which point the maximum serial dependence has been learned by the context. Note that altering the alphabet or symbolization parameters will change the no-memory entropy as well, thus the compression ratio of compressed bits per symbol vs no-memory bits per symbol is the optimization target.

If one is interested in the stationarity of some narrow aspects or of exclusively rare events in the data, our sort of general compression model might not be appropriate and could give misleading results or low statistical power. In simulation and experiment, however, the present method is useful and sensitive to many sorts of parametric drift or change.

We thank members of INLS and CADO for discussions.

[1] A. N. Kolmogorov, Dokl. Akad. Nauk (SSSR) **119**, 861 (1958); Y. Sinai, *ibid.* **124**, 768 (1959).

[2] A. M. Fraser, IEEE Trans. Inf. Theory **35**, 245 (1989); A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).

[3] The most common application has been for entropy estimation, usually via string matching methods inspired by Lempel-Ziv compression. A context method similar to the present one was presented in T. Schurmann and P. Grassberger, Chaos **6**, 414 (1996).

[4] M. B. Kennel, Phys. Rev. E **56**, 316 (1997); T. Schreiber, Phys. Rev. Lett. **78**, 843 (1997); A. Witt, J. Kurths, and A. Pikovsky, Phys. Rev. E **58**, 1800 (1998).

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

[6] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).

[7] L. Mason, A. I. Mees, and K. Judd (unpublished); A. I. Mees and M. B. Kennel (unpublished).

[8] J. Rissanen, IEEE Trans. Inf. Theory **29**, 656 (1983); S. Bunton, in *A Percolating State Selector for Suffix-Tree Context Models*, Proceedings of the Data Compression Conference, 1997, edited by J. A. Storer and M. Cohn (IEEE Computer Society Press, Los Alamitos, CA, 1997), pp. 32-41.

[9] F. M. Willems, Y. M. Shtarkov, and T. J. Tjalkens, IEEE Trans. Inf. Theory **41**, 653 (1995).

[10] Y. M. Shtarkov, T. J. Tjalkens, and F. M. J. Willems, Probl. Inf. Transm. **33**, 17 (1997). Although the Krischevsky-Trofimov (KT) estimator has nice analytically computable asymptotic properties, other estimators may be superior for finite length observed data. We explored a number of ansatzes used by the text compression community but none were consistently superior to the KT estimator with a variable ‘‘ballast’’ $0<\beta<1$ optimization parameter: $\hat{P}(k)=(c_k+\beta)/\Sigma_i(c_i+\beta)$. Often $\beta=1/|A|$ gave results close to the minimum code length.

[11] K. Judd and A. I. Mees, Physica D (to be published).

[12] The χ^2 analytics degrade for small bin counts. For those bins, we merge any symbols whose expected count—for either set one or two—is less than five, reducing the degrees of freedom appropriately. If, after merging, there remain fewer than two symbols passing this criterion, then this node is wholly excluded.

[13] In this discrete case, the sum will encounter tables exactly as likely as the observed one (such as the observed table itself); the summed p_T for these tables is weighted by a uniform random deviate $r\in[0,1)$.

[14] E. N. Lorenz, Tellus, Ser. A **36**, 98 (1984). The model is $dx/dt=-y^2-z^2-a(x-F)$, $dy/dt=xy-bxz-y+1$, $dz/dt=bxy+xz-z$, $a=1/4$, $b=4$, $F=8$. Each set was 5000 points long sampled every $\delta t=0.08$.

[15] C. S. Daw, C. E. A. Finney, M. Vasudevan, N. A. van Goor, K. Nguyen, D. D. Bruns, E. J. Kostelich, C. Grebogi, E. Ott, and J. A. Yorke, Phys. Rev. Lett. **75**, 2308 (1995).

[16] K. E. Trenberth and T. J. Hoar, Geophys. Res. Lett. **23**, 57 (1996).

[17] D. E. Harrison and N. K. Larkin, Geophys. Res. Lett. **24**, 1779 (1997).

[18] X. Z. Tang, E. R. Tracy, A. D. Boozer, A. deBrauw, and R. Brown, Phys. Rev. E **51**, 3871 (1995).

[19] C. S. Daw, M. B. Kennel, C. E. A. Finney, and F. T. Connolly, Phys. Rev. E **57**, 2811 (1998).

[20] Unfortunately the state-selecting context methods that are amenable to rigorous proof of universality are slightly different from those that are most practical to implement and have the best performance, though most of the time the differences are small.