Detecting differences between delay vector distributions

C. Diks
Department of Physiology, Leiden University, P.O. Box 9604, 2300 RC Leiden, The Netherlands

W. R. van Zwet
Department of Mathematics and Computer Science, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands

F. Takens
Department of Mathematics, Groningen University, P.O. Box 800, 9700 AV Groningen, The Netherlands

J. DeGoede
Department of Physiology, Leiden University, P.O. Box 9604, 2300 RC Leiden, The Netherlands

(Received 21 August 1995)

We propose a test for the null hypothesis that two sets of vectors are drawn from the same multidimensional probability distribution. The application to delay vector distributions provides a test for the null hypothesis that two time series have been generated by the same mechanism.

PACS number(s): 02.50.−r, 05.40.+j, 05.45.+b

I. INTRODUCTION

In time series analysis, the question of whether two sets of delay vectors have the same underlying multidimensional probability distribution is relevant in many situations that are encountered. For example, if a model time series is to be compared with the original time series, objective criteria for the quality of the model are desirable, if not necessary. Also, in examining the stationarity of a time series it is important to be able to detect slow changes in its behavior. These changes may not be manifest in the linear properties of the time series, usually determined with power spectra or estimated parameters of linear models. Usually, in the analysis of time series from spatiotemporal dynamics systems that may be nonhomogenous in space, the first step consists of comparing time series measured at different sites. With these applications in mind, the aim of this paper is to propose a test for the hypothesis that two time series have the same delay vector distributions, a necessary condition for two stationary time series to be generated by the same mechanism.

The characterization of time series by their delay vector distributions is by now standard in the linear and nonlinear time series literature; see, e.g., Grassberger, Schreiber, and Schaffrath [1] for a review. Most of the methods are based on the assumption that the time series can be described by a member from a specific class of models such as the class of linear models or low-dimensional chaotic models. Two relatively new methods that do not explicitly depend on the mechanisms generating the time series but use delay vector distributions only are described by Wright and Schult [2] and Kantz [3]. Wright and Schult have proposed a method for identifying the presence of a time series with known delay vector distribution within a given time series. Kantz has introduced a method for comparing the delay vector distributions of two time series. The remark that one of his statistics under certain conditions and in a certain limit could be interpreted as a distance between two delay vector distributions has motivated the present paper.

Our method is based on a general distance concept between multidimensional distributions. A consistent estimator of the square of this distance is constructed, and the variance of the estimator is calculated conditionally on the set of observed vectors, assuming independence of all vectors. This provides a consistent test for the null hypothesis that two sets of independent vectors are drawn from the same probability distribution.

The paper is organized as follows. In Sec. II, the squared distance Q is defined. It contains a parameter d that may be identified as a bandwidth and is the length scale at which the two sets are compared. A small value of d will take into account differences between the distributions at small scales. However, taking d too small will lead to poor statistics. A natural unbiased estimator $\hat{Q}$ for Q is given, and the variance $V_c(\hat{Q})$ of $\hat{Q}$ under the null hypothesis is calculated conditionally on the observed vectors given their independence. The distribution of $S=\hat{Q}/\sqrt{V_c(\hat{Q})}$ is examined numerically for a fixed value of the bandwidth d. Section III describes the application of the test to delay vector distributions. For the comparison of delay vector distributions obtained from time series, some heuristic modifications have to be made in order to reduce the effect of dependence present among the vectors. The choice of the bandwidth parameter d is discussed in Sec. IV. The results are summarized and discussed in Sec. V.

II. A DISTANCE BETWEEN MULTIDIMENSIONAL DISTRIBUTIONS

In this section a distance between two probability distributions, $\rho_1(\vec{r})$ and $\rho_2(\vec{r})$, in $\mathbb{R}^m$ is defined. It is shown that the square of this distance can be estimated in an unbiased manner from two sets of vectors sampled independently from $\rho_1(\vec{r})$ and $\rho_2(\vec{r})$, respectively. We will consider the case of N_1 vectors, $\{\vec{X}_i\}_{i=1}^{N_1}$, with probability distribution $\rho_1(\vec{X}_i)$, and N_2 vectors, $\{\vec{Y}_i\}_{i=1}^{N_2}$, with probability distribution $\rho_2(\vec{Y}_i)$. Realizations of these vectors will be denoted by their lowercase analogs, i.e., $\vec{x}_i$ and $\vec{y}_i$.

We define the smoothed distributions ρ_k' as the convolution of ρ_k and κ, viz.,

$$\rho_k'(\vec{r})=\int d\vec{s}\,\rho_k(\vec{s})\kappa(\vec{r},\vec{s}) \quad \text{for } k\in\{1,2\} , \tag{1}$$

where $\kappa(\vec{r},\vec{s})$ is a Gaussian kernel defined by

$$\kappa(\vec{r},\vec{s})=(\sqrt{2\pi}d)^{-m}e^{-|\vec{r}-\vec{s}|^2/(2d^2)} , \tag{2}$$

with $|\;|$ denoting the Euclidean distance in $\mathbb{R}^m$ and $d>0$ being a fixed length scale, or bandwidth. The motivation for introducing the smoothed distributions $\rho_k'(\vec{r})$ is that unbiased estimators in terms of the sampled vectors can be easily given for $\rho_k'(\vec{r})$ at each point in $\mathbb{R}^m$. An unbiased estimator $\widehat{\rho_1'(\vec{r})}$ of $\rho_1'(\vec{r})$ is

$$\widehat{\rho_1'(\vec{r})}=\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\vec{r},\vec{x}_i) , \tag{3}$$

since the expected value of $\kappa(\vec{r},\vec{X}_i)$ is

$$\int d\vec{X}_i\rho_1(\vec{X}_i)\kappa(\vec{r},\vec{X}_i)=\rho_1'(\vec{r}) \tag{4}$$

for all i.

We define the quantity Q as

$$Q=(2d\sqrt{\pi})^m\int d\vec{r}[\rho_1'(\vec{r})-\rho_2'(\vec{r})]^2 . \tag{5}$$

For every $d>0$, the square root of Q defines a distance between the probability distributions ρ_1' and ρ_2', based on the inner product of $(\rho_1'-\rho_2')$ with itself. We thus have $Q=0$ if and only if $\rho_1'(\vec{r})=\rho_2'(\vec{r})$. Since the kernel function κ is bounded, absolutely integrable, and has a Fourier transform that does not vanish on any interval, the stronger result follows that $Q=0$ if and only if $\rho_1(\vec{r})=\rho_2(\vec{r})$; cf. Anderson, Hall, and Titterington [4]. By determining whether a consistent estimator of Q is significantly larger than zero, we have a test for the null hypothesis $\rho_1=\rho_2$ that is consistent against all alternatives $\rho_1\neq\rho_2$.

To find an unbiased estimator $\hat{Q}$ for Q, it is convenient to rewrite Eq. (5) as

$$Q=Q_{11}+Q_{22}-2Q_{12} , \tag{6}$$

where

$$Q_{kl}=(2d\sqrt{\pi})^m\int d\vec{r}\,\rho_k'(\vec{r})\rho_l'(\vec{r}) . \tag{7}$$

By substituting (1) into (7) and performing the outer integral, using the property that the convolution of two Gaussian distributions is Gaussian,

$$(\sqrt{2\pi}d)^{-2m}\int d\vec{r}\,e^{-|\vec{r}-\vec{s}|^2/(2d^2)-|\vec{r}-\vec{t}|^2/(2d^2)} =[\sqrt{2\pi}(\sqrt{2}d)]^{-m}e^{-|\vec{s}-\vec{t}|^2/(4d^2)} , \tag{8}$$

we obtain

$$Q_{kl}=\int\int d\vec{s}\,d\vec{t}\,\rho_k(\vec{s})h(\vec{s},\vec{t})\rho_l(\vec{t}) \tag{9}$$

with

$$h(\vec{s},\vec{t})=e^{-|\vec{s}-\vec{t}|^2/(4d^2)} . \tag{10}$$

We have $N_1(N_1-1)/2$ different pairs of vectors $(\vec{x}_i,\vec{x}_j)$, $N_2(N_2-1)/2$ different pairs $(\vec{y}_i,\vec{y}_j)$, and N_1N_2 different pairs $(\vec{x}_i,\vec{y}_j)$. With the use of (6) and (9), the statistic $\hat{Q}$ defined by

$$\hat{Q}=\frac{1}{\binom{N_1}{2}}\sum\sum_{1\le i<j\le N_1}h(\vec{X}_i,\vec{X}_j) +\frac{1}{\binom{N_2}{2}}\sum\sum_{1\le i<j\le N_2}h(\vec{Y}_i,\vec{Y}_j) -\frac{2}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}h(\vec{X}_i,\vec{Y}_j) \tag{11}$$

can be easily seen to be an unbiased estimator of Q.

The variance $V_c(\hat{Q})$ of $\hat{Q}$ under the null hypothesis and conditionally on the set of N_1+N_2 observed vectors is a function of N_1, N_2 and the set of N_1+N_2 vectors. If we define $N=N_1+N_2$ and

$$\vec{z}_i=\begin{cases}\vec{x}_i & \text{for } 1\le i\le N_1\\ \vec{y}_{i-N_1} & \text{for } N_1<i\le N ,\end{cases} \tag{12}$$

the variance under the null hypothesis conditionally on $\{\vec{z}_i\}_{i=1}^{N}$ is

$$V_c(\hat{Q})=\frac{2(N-1)^2(N-2)}{N_1(N_1-1)N_2(N_2-1)(N-3)} \times\left[\frac{1}{\binom{N}{2}}\sum\sum_{1\le i<j\le N}\psi_{ij}^2\right] \tag{13}$$

in which

$$\psi_{ij}=H_{ij}-g_i-g_j , \tag{14}$$

where

$$H_{ij}=h(\vec{z}_i,\vec{z}_j)-\frac{1}{\binom{N}{2}}\sum\sum_{1\le i<j\le N}h(\vec{z}_i,\vec{z}_j) \tag{15}$$

and

$$g_i=\frac{1}{N-2}\sum_{\substack{j\\ j\neq i}}H_{ij} . \tag{16}$$

We give a derivation of this result in the Appendix. We remark that an expression for the unconditional variance in terms of N_1 and N_2 and ρ_1 and ρ_2 is more easily derived. However, since ρ_1 and ρ_2 are not completely known, we would have to estimate the unconditional variance from the observed vectors. Working conditionally on the set of vectors, however, we obtain an exact value for the variance. The quantity S defined by

$$S=\frac{\hat{Q}}{\sqrt{V_c(\hat{Q})}} \tag{17}$$

is a random variable with zero mean and standard deviation equal to 1 under the null hypothesis.

III. APPLICATION TO DELAY VECTOR DISTRIBUTIONS

In this section we will develop a procedure to test the null hypothesis that the underlying delay vector distributions of two time series are identical. For a scalar time series $\{s_i\}_{i=1}^{M}$ the set of m-dimensional delay vectors with delay τ consists of the $L=M-(m-1)\tau$ vectors

$$\vec{x}_i=[s_i,s_{i+\tau},\ldots,s_{i+(m-1)\tau}] \ . \tag{18}$$

The distribution

$$\rho(\vec{x})=\frac{1}{L}\sum_{i=1}^{L}\delta(\vec{x}-\vec{x}_i) \ , \tag{19}$$

where $\delta(\ )$ is the Dirac delta function, is referred to as the empirical delay vector distribution of the time series. The major problem for the application of virtually all statistical methods in the field of time series analysis is the presence of dependence among delay vectors. Therefore, the effect of dependence will be examined first.

In the previous section, the properties of S under the null hypothesis, i.e., mean zero and standard deviation 1, were derived under the assumption of independence of the delay vectors. This is an oversimplification, even if both time series consist of independent, identically distributed (IID) data. The mere construction of the sets of delay vectors from the two time series introduces dependence among the delay vectors. Dynamical structure in the time series will introduce additional dependence. Using examples, we will demonstrate the deterioration of the test if dependence is not taken into account. A modification of the test then is shown to reduce the effect of dependence to a level at which it can be ignored. Before each analysis, both time series are rescaled by the same factor σ_0/σ. Here σ_0 is the standard deviation of a uniform distribution between -0.5 and 0.5, and σ is the sample standard deviation of the scalar data (i.e., the overall standard deviation of the scalar data of both time series). We choose to normalize the standard deviation rather than the range (i.e., the difference between the largest and the smallest value observed), since for some types of time series the expected range depends on the length of the time series. For example, in the case of IID Gaussian data, the expected range grows as the square root of the time series length. The parameters for the construction of the delay vectors are fixed at $m=3$ and $\tau=1$, and we perform the test with the bandwidth parameter $d=2.5\times10^{-2}$.

The distribution of S under the null hypothesis is not necessarily normal. In general, it depends on N_1 and N_2, the bandwidth d, and the distribution $\rho_1=\rho_2$. Here the null distribution is examined numerically by Monte Carlo simulations. Three different processes are considered. For each process, the calculation of S is performed 10^3 times. Each time, two independently generated time series are used, and from the results the mean $\bar{S}$ and the standard deviation $\sigma(S)$ are calculated. The processes we consider are uniformly distributed random data, logistic map data, and Hénon map data. All time series have a length of $M=200$ samples. The results are summarized in Table I. For independent vectors, the expected value of $\bar{S}$ is 0 and the standard error in $\bar{S}$ is $(10^3)^{-1/2}\approx0.03$. The expected value of $\sigma(S)$ is 1. The standard error in $\sigma(S)$ depends on the fourth moment of the distribution of S. We expect it to be of the order of $2^{1/2}(10^3)^{-1/2}\approx0.04$, the value that is obtained for the standard normal distribution. The values found for the uniformly distributed random data [$\bar{S}=0.05$, $\sigma(S)=1.15$] are close to their expected values. The dependence introduced by the construction of delay vectors appears to be of minor influence on the null distribution. Dynamical dependence, on the other hand, has a severe effect. The results for the two chaotic maps [$\bar{S}=-0.12$ and $\sigma(S)=1.93$ for the logistic map, and $\bar{S}=-0.46$ and $\sigma(S)=2.23$ for the Hénon map] clearly indicate that dependence introduced by the dynamics cannot be ignored.

The problem of dependence among delay vectors has been investigated in a different context by Theiler [5]. In this work it has been shown that a bias due to dependence can be removed from an estimator of the same type as $\hat{Q}$ by leaving out the contributions of (i,j) pairs of indices for which $|i-j|$ is smaller than l. Here l is some fixed number larger than the largest typical time scale of the time series. We note that this procedure still does not allow the use of the expression for the variance of our estimator as derived for independent vectors. Strongly dependent delay vectors will give rise to a different variance in the estimator than independent delay vectors, regardless of whether a bias has been removed or not. In

TABLE I. Mean value $\bar{S}$ and standard deviation $\sigma(S)$ for the distributions of S estimated by performing the test 10^3 times on independent realizations of uniform random data, logistic map data, and Hénon map data. The results are given for the naive method and the segment method.

$\bar{S}\pm\sigma(S)$	Naive method ($l=1$)	Segment method ($l=18$)
Uniform random data	0.05 ± 1.15	0.05 ± 1.06
Logistic map data	-0.12 ± 1.93	-0.01 ± 1.22
Hénon map data	-0.46 ± 2.23	-0.03 ± 1.12

other words, the (i,j) pairs that have $|i-j|\geq l$ are pairwise independent, but the contributions from (i,j) and $(i,j+k)$ for $k<l$ manifest themselves in the estimated variance. A method based on the approach of Theiler, but which also takes into consideration this effect of dependence in the estimated variance, is the following.

We divide the (i,j) plane of indices in squares of size $l\times l$ and use the average value

$$h'(i',j')=\frac{1}{l^2}\sum_{p=1}^{l}\sum_{q=1}^{l}h(\vec{z}_{i'l+p},\vec{z}_{j'l+q}) \qquad (20)$$

in each of these squares rather than the individual values. By using h' instead of h in the definition of $\hat{Q}$ in (11), the conditional variance in terms of h' is analogous to (13), with h' substituted for h. The idea behind this procedure is that for l large enough, nonoverlapping pieces of length l within the multidimensional series of delay vectors can be considered as independent for all practical purposes. Heuristically, in this way one can approximately take into account short-range (within a time scale l) dependence among delay vectors. As in the work of Theiler [5] the parameter l needs to be chosen larger than a typical time scale of decay of dependence in the time series. In the following, the method with $l>1$ will be referred to as the segment method.

For time series of length $M=200$, and $\tau=1$, $m=3$, there are $L=M-(m-1)\tau=198$ delay vectors for each time series. For the discrete time dynamical systems discussed in this paper, the dependence within the time series has a typical decay time of a few iterations. The results obtained by the segment method with $l=18$, which is sufficiently large because of the short-range dependence, are given in Table I. The values of $\bar{S}$ and $\sigma(S)$ found with the segment method for the three processes considered are close to the values expected for independent delay vectors. We found $\bar{S}=0.05$ and $\sigma(S)=1.06$ for the uniform random data. For the logistic map the values are $\bar{S}=-0.01$ and $\sigma(S)=1.22$, and for the Hénon map they are $\bar{S}=-0.03$ and $\sigma(S)=1.12$. For the three models, the segment method gives better results than the naive ($l=1$) method.

In applications of the test, one will need a rejection criterion. As argued in the previous section, the null hypothesis can be rejected when estimated values of S are large. We choose to reject the null hypothesis at values of S larger than 3. Assuming a zero mean, a standard deviation of 1, and a unimodal distribution for S, the probability of finding a value of S larger than 3 is smaller than 0.05; see Pukelsheim [6]. Since the latter inequality applies to the two-sided test case, this inequality will be conservative when used in the one-sided case. In the numerical simulations with the segment method, the number of rejections of the null hypothesis were 12, 22, and 11 out of 10^3 for the uniform random data, the logistic data, and the Hénon data, respectively. These values indeed indicate a probability of rejecting the null hypothesis well below 0.05 if the null hypothesis holds.

Example 1. As an example for which the null hypothesis does not hold, we compare the delay vector distributions of time series generated with Hénon's model, $X_{n+1}=1-aX_n^2+bX_{n-1}$, at slightly different model parameters. For each first time series we choose $a=1.35$ and $b=0.31$; for each second series we take the standard parameter values $a=1.4$ and $b=0.3$; cf. Kantz [3]. The results found for $d=2.5\times10^{-2}$ with the segment method are $\bar{S}=3.59$ and $\sigma(S)=1.17$. Using the rejection criterion $S>3.0$, we found that there were 667 rejections of the null hypothesis out of the 10^3 simulations.

IV. BANDWIDTH

In this section, we briefly discuss the choice of the bandwidth parameter d, which sets the length scale of the smoothing. By choosing it relatively small, $\hat{Q}$ will pick up local differences in ρ_1 and ρ_2. Taking it too small, however, leads to poor statistics, as can be seen from the behavior of $\hat{Q}$ in the limit $d\to0$. In this limit, the sums in (11) are dominated by one term only: the term corresponding to the smallest distance in the set of vectors. For large d, the delay vector distributions $\rho_1'(\vec{r})$ and $\rho_2'(\vec{r})$ are smoothed to such an extent that they become almost indistinguishable. We expect to find an optimal value for the bandwidth at the tradeoff of these two effects. It is known that the optimal bandwidth for density estimators depends on the number of observations and decreases for larger numbers of data. The problem we address here is not that of optimally estimating densities but rather that of finding the bandwidth for which the test is most powerful. Suppose the two distributions ρ_1 and ρ_2 are different. We are interested in the question of how the optimal value of the bandwidth parameter d depends on the number of observations N. In general, for a fixed value of the bandwidth d, the expected value $\langle S\rangle$ will depend on $p=N_1/N$ and N as

$$\langle S\rangle\sim p(1-p)N \quad \text{for } N \text{ large} . \qquad (21)$$

This result is obtained by considering the large N behavior of $\hat{Q}$ and $\sqrt{V_c(\hat{Q})}$. The estimator $\hat{Q}$ for large N becomes sharply distributed around the true value Q, which depends on ρ_1, ρ_2, and d, and by (13) we have

$$1/\sqrt{V_c(\hat{Q})}\sim p(1-p)N \quad \text{for } N \text{ large} . \qquad (22)$$

Taking into account the dependence on d, we have

$$\langle S\rangle\sim p(1-p)NF_{\rho_1,\rho_2}(d) \quad \text{for } N \text{ large} , \qquad (23)$$

where $F_{\rho_1,\rho_2}(d)$ is some function of d that depends on ρ_1 and ρ_2. Thus for large N and given ρ_1, ρ_2, and p, the value of d for which the optimum expected value $\langle S\rangle$ occurs is independent of N.

Next we would like to consider the d dependence of the three different terms in the estimator $\hat{Q}$ defined by (11). The first term contains distances within the delay vector distribution of the first time series; the second term is a function of the second time series only; and the third term contains the cross terms, which involve both time series. For two time series generated with Hénon's model at different parameter values (cf. example 1 in the previous section), we calculated (with the segment method,

$l=18$) the separate contributions

$$Q_{11}=\frac{1}{\binom{N_1}{2}}\sum\sum_{1\le i<j\le N_1} h(\vec{X}_i,\vec{X}_j)\ , \tag{24}$$

$$Q_{22}=\frac{1}{\binom{N_2}{2}}\sum\sum_{1\le i<j\le N_2} h(\vec{Y}_i,\vec{Y}_j)\ , \tag{25}$$

and

$$Q_{12}=\frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} h(\vec{X}_i,\vec{Y}_j) \tag{26}$$

as a function of d.

Figure 1 shows a log-log plot of the estimates versus the bandwidth parameter d. For large d the curves of the three estimators practically coincide, while for small d the curves diverge from one another. We emphasize that the smoothness of the curves is the result of using the same two sets of delay vectors for all d values. In Fig. 2, S is given as a function of d for the same two time series. The function clearly has a pronounced maximum value near $d\approx 7.5\times 10^{-3}$, where the value of S indicates a deviation from the expected zero value of about 6.2 standard deviations. It cannot, however, be concluded that the curve differs from the expected curve by 6.2 standard deviations. By considering only the optimum value, one is selecting the largest value in a realization of a random function. One way of dealing with this problem is to develop a theory concerning the fluctuations of S. However, there is a simple method to overcome this problem. The function $F_{\rho_1,\rho_2}(d)$ appearing in (23) can be estimated using a relatively small number of the available vectors, for example, those obtained from two small initial segments of both time series. This estimated function can be used together with (23) as a guide for the determination of a value of d for which the method is most powerful given the type of data under consideration. The last segments of both time series can then be used to obtain an independent value of S for this estimated optimal length scale parameter d.

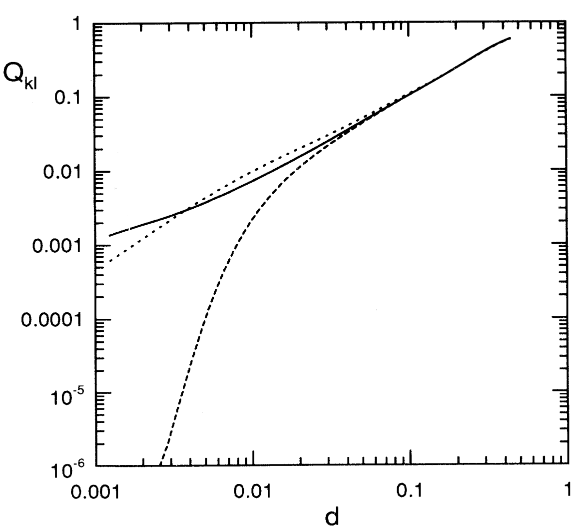

FIG. 1. Estimated values (dimensionless) of Q_{11} (solid line), Q_{22} (dotted line), and Q_{12} (dashed line) as a function of the bandwidth parameter d (dimensionless) for two time series of length $M=200$, generated by slightly different Hénon models.

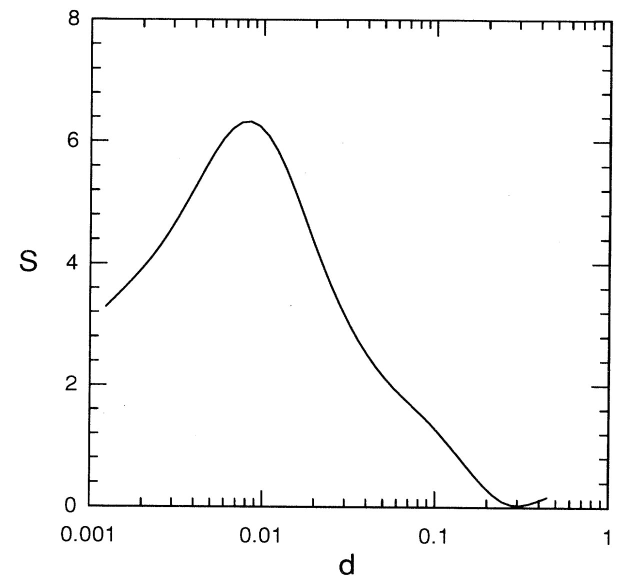

FIG. 2. Number of standard deviations S (dimensionless) as a function of d (dimensionless) corresonding with Fig. 1.

Example 2. Again, we compare the delay vector distributions of time series generated with Hénon's model with different parameters. The model parameters used are identical to those in example 1 in the previous section. The test parameters are the same, except for the bandwidth. The estimated optimal bandwidth $d=7.5\times 10^{-3}$ is used. The results with the segment method are $\bar{S}=6.00$ and $\sigma(S)=0.70$. With the rejection criterion $S>3.0$, there were 10^3 rejections of the null hypothesis out of the 10^3 simulations. This is an improvement over the first example, where the test was performed with $d=2.5\times 10^{-2}$ and where the null hypothesis could be rejected only 667 out of 1000 simulations.

We have discussed the bandwidth dependence of the estimators here in the context of the optimal choice for statistical inference. However, the bandwidth dependence of Q_{11}, Q_{22}, and Q_{12} may contain some additional information. As done in Ref. [3], one may search for the bandwidth above which the two delay vector distributions are the same within a given accuracy ϵ. This bandwidth d_ϵ is defined as the minimum bandwidth above which one has both $\log Q_{11}-\log Q_{12}<\epsilon$ and $\log Q_{22}-\log Q_{12}<\epsilon$. The physical interpretation of this bandwidth is clear in some cases. For example, in comparing clean and noisy time series, d_ϵ has a close correspondence to the noise level.

V. DISCUSSION

In this paper we have proposed a test for the null hypothesis that two multidimensional probability distributions are identical. The test is based on the statistic $\hat{Q}$, which is an unbiased estimator of the square of a distance between the two probability distributions. This approach is similar to that followed in [7], where a distance based test statistic has been used to detect differences in forward and time-reversed delay vector distributions. The standard deviation of $\hat{Q}$ under the null hypothesis, and under the condition of statistical independence of the vectors, is calculated conditionally on the set of vectors. The effect of dependence in empirical delay vector distributions is examined. For applications to delay vector distributions, a procedure is proposed that turns out to be relatively robust with respect to dependence. The generalization of the method to multivariate time series is straightforward. Once delay vectors have been constructed from the multivariate time series, the test statistic can be calculated.

The test statistic contains a characteristic length scale parameter d, which can be used to detect differences between the distributions at a length scale d. The power of the test depends on this parameter. A two-step method is suggested in which an optimal value for d is estimated by using only a part of the available data before the test is performed on the remaining part of the data. Examples demonstrate that only a few hundred data points are necessary to distinguish two Hénon time series with slightly different parameters (the attractors of which look the same as judged by the naked eye; see [3]).

The properties of the kernel h appearing in Eq. (9) is determined by the choice made for κ. In [3], the starting point has been (9) with the kernel $h=\Theta(d-|\vec{r}-\vec{s}|)$, where $\Theta(\ )$ is the Heaviside step function. In this case, we can find the conditional variance in a way completely analogous to the method used for our estimator. However, this kernel does not define as estimator with an expected value that may be interpreted as the square of a distance between two distributions. This is the case only in the limit $d\to 0$, assuming smoothness of the distributions. Therefore, this kernel cannot be used for one-sided testing, and the test based on this kernel is not consistent against all alternatives $\rho_1\neq\rho_2$. There is, however, a qualitative correspondence between the behavior of the curves in Fig. 1 and of the estimators of the correlation integrals

$$C_{kk}=\int\int d\vec{s}\,d\vec{t}\,\rho_k(\vec{s})\Theta(d-|\vec{s}-\vec{t}|)\rho_k(\vec{t}) \quad \text{for } k\in\{1,2\} \tag{27}$$

and the cross-correlation integral [5]

$$C_{12}=\int\int d\vec{s}\,d\vec{t}\,\rho_1(\vec{s})\rho_2(\vec{t})\Theta(d-|\vec{s}-\vec{t}|)\ . \tag{28}$$

The scaling behavior of correlation integrals of the delay vector distribution of low-dimensional dynamical systems is also present for Q_{11} and Q_{22}. It can be proved that if $\rho_k(\vec{r})$ has a correlation dimension D_2, i.e.,

$$C_{kk}\sim d^{D_2} \quad \text{for } d\to 0\ , \tag{29}$$

then Q_{kk} obeys

$$Q_{kk}\sim d^{D_2} \quad \text{for } d\to 0\ . \tag{30}$$

By construction, our test statistic is unbiased. The test statistic discussed by Anderson *et al.* [4] is

$$T=\int d\vec{x}(\hat{f}_1(\vec{x})-\hat{f}_2(\vec{x}))^2\ , \tag{31}$$

where $\hat{f}_1$ and $\hat{f}_2$ are kernel density estimates of ρ_1 and ρ_2, respectively. Because of this definition, this test statistic is biased if the number of observations is finite. The problem of estimating the bias is overcome by using bootstrap techniques [4].

The test can possibly be used to compare time series with their surrogates, e.g., phase randomized time series [8]. However, this will require further investigations. So far, we have applied our test only to pairs of time series that were mutually independent. It should be checked whether the method can still be applied if the series depend on each other. For example, a phase randomized surrogate time series and the corresponding original time series are strongly dependent since they are constructed to have their sample power spectra identical. Another possible application is a test for the stationarity of a time series. We expect that the comparison of different segments of a given time series with the proposed test is promising for the study of the stationarity of a time series.

We end with a few remarks on the choice of the parameters m and τ used for the construction of the delay vectors. For the delay τ we expect good results using the optimality criteria proposed in the nonlinear time series analysis literature. For example, the first minimum in the mutual information criterion of Fraser [9] can be used. Unfortunately, there is not always a minimum in the mutual information function. If so, we suggest an examination of two-dimensional delay vector distributions for a range of delays. Too small delays will give a phase plot with all points close to the diagonal line, whereas structure within a time series will not be seen in phase plots with too large delays. The optimal choice of m, usually referred to as embedding dimension, will probably be strongly coupled with the choice of d. For small m there will be more small distances than for large m. We thus expect a relatively high performance at small values of d for small m. Until the effect of dependence as a function of m and d has been investigated in more detail, it appears safe to use small m together with small d. The independent distance hypothesis of Theiler [10] implies that the delay vectors in the limit $d\to 0$ may be treated as if they are independent.

APPENDIX

In this appendix we will derive expressions for the unconditional and the conditional variance of $\hat{Q}$. The derivation of the unconditional variance (which is not used in the test) is included for completeness. We will closely follow the methodology of Van Zwet [11].

1. Unconditional variance

The random vectors $\{\vec{X}_1\}_{i=1}^{N_1}$ and $\{\vec{Y}_i\}_{i=1}^{N_2}$ are assumed to be independent and to have identical probability distributions $\rho(\vec{X}_i)=\rho(\vec{Y}_i)$. We have

$$\hat{Q}=\frac{1}{\binom{N_1}{2}}\sum\sum_{1\le i<j\le N_1} h(\vec{X}_i,\vec{X}_j)+\frac{1}{\binom{N_2}{2}}\sum\sum_{1\le i<j\le N_2} h(\vec{Y}_i,\vec{Y}_j)-\frac{2}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} h(\vec{X}_i,\vec{Y}_j)\ . \quad (A1)$$

Denoting expected values by $E(\ )$, we define the functions

$$H(\vec{x},\vec{y})=h(\vec{x},\vec{y})-E(h(\vec{X}_1,\vec{X}_2)) \quad (A2)$$

and

$$g(\vec{x})=E(H(\vec{X}_1,\vec{X}_2)|\vec{X}_1=\vec{x})\ . \quad (A3)$$

Now $\hat{Q}$ can be expressed in terms of

$$\psi(\vec{x},\vec{y})=H(\vec{x},\vec{y})-g(\vec{x})-g(\vec{y}) \quad (A4)$$

by substituting

$$h(\vec{x},\vec{y})=\psi(\vec{x},\vec{y})+g(\vec{x})+g(\vec{y})+E(h(\vec{x},\vec{y})) \quad (A5)$$

into (A1). Since the contributions $g(\ )$ and $E(h(\ ,\ ))$ sum to zero, we obtain

$$\hat{Q}=\frac{1}{\binom{N_1}{2}}\sum\sum_{1\le i<j\le N_1} \psi(\vec{X}_i,\vec{X}_j)+\frac{1}{\binom{N_2}{2}}\sum\sum_{1\le i<j\le N_2} \psi(\vec{Y}_i,\vec{Y}_j)-\frac{2}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \psi(\vec{X}_i,\vec{Y}_j)\ . \quad (A6)$$

Now, by construction we have

$$E(\psi(\vec{X}_i,\vec{X}_j)|\vec{X}_i=\vec{x})=0 \quad \text{for all } \vec{x}\ , \quad (A7)$$

so that the right-hand side of (A6) consists of a sum of terms that are all uncorrelated. By denoting variances by $V(\ )$, the variance of $\hat{Q}$ can be written as

$$V(\hat{Q})=\left[\frac{2}{N_1(N_1-1)}+\frac{2}{N_2(N_2-1)}+\frac{4}{N_1N_2}\right]\times V(\psi(\vec{X}_i,\vec{X}_2))\ . \quad (A8)$$

We can express $V(\psi(\vec{X}_1,\vec{X}_2))$ as

$$V(\psi(\vec{X}_1,\vec{X}_2))=V(h(\vec{X}_1,\vec{X}_2))-2V(g(\vec{X}_1))\ , \quad (A9)$$

using the definitions (A2)–(A4).

2. Conditional variance

We introduce $N=N_1+N_2$ and define

$$\vec{z}_i=\begin{cases}\vec{x}_i & \text{for } 1\le i\le N_1\\ \vec{y}_{i-N_1} & \text{for } N_1<i\le N\ .\end{cases} \quad (A10)$$

Given the vectors $\{\vec{z}_i\}_{i=1}^{N}$, we consider random divisions of the set of vectors into two groups of sizes, N_1 and N_2. We obtain

$$\hat{Q}=\frac{1}{\binom{N_1}{2}}\sum\sum_{\substack{i,j\in D\\ i<j}}h(\vec{z}_i,\vec{z}_j)+\frac{1}{\binom{N_2}{2}}\sum\sum_{\substack{i,j\in D^c\\ i<j}}h(\vec{z}_i,\vec{z}_j)-\frac{2}{N_1N_2}\sum_{i\in D}\sum_{j\in D^c}h(\vec{z}_i,\vec{z}_j)\ , \quad (A11)$$

where D is a set of N_1 indices randomly selected without replacement from $\{1,2,\ldots,N\}$ and D^c is the complementary set of indices. We write (A11) as

$$\hat{Q}=\sum\sum_{i<j}h(\vec{z}_i,\vec{z}_j)A_{ij}\ , \quad (A12)$$

with

$$A_{ij}=\frac{1}{\binom{N_1}{2}}1_D(i)1_D(j)+\frac{1}{\binom{N_2}{2}}1_{D^c}(i)1_{D^c}(j)-\frac{2}{N_1N_2}1_D(i)1_{D^c}(j)-\frac{2}{N_1N_2}1_{D^c}(i)1_D(j), \quad (A13)$$

where $1_D(i)$ is the indicator function defined by

$$1_D(i)=\begin{cases}1 & \text{if } i\in D\\ 0 & \text{if } i\notin D\ .\end{cases} \quad (A14)$$

The A_{ij} in (A12) can be identified as random variables, while the $h(\vec{z}_i,\vec{z}_j)$ are constants. The conditional expected value $E_c(\hat{Q})$ of $\hat{Q}$ is zero since

$$E_c(A_{ij}|1_D(j))=E_c(A_{ij}|1_D(i))=0 \quad (A15)$$

for all i and j. We define the constants

$$H_{ij}=h(\vec{z}_i,\vec{z}_j)-\frac{1}{\binom{N}{2}}\sum\sum_{i<j}h(\vec{z}_i,\vec{z}_j)\ , \quad (A16)$$

$$g_i=\frac{1}{N-2}\sum_{\substack{j\\ j\ne i}}H_{ij}\ , \quad (A17)$$

and

$$\psi_{ij}=H_{ij}-g_i-g_j\ . \quad (A18)$$

Since A_{ij} is symmetric in i and j and

$$\sum_{\substack{i \\ i\neq j}} A_{ij}=0 \quad \text{for all } j \ , \tag{A19}$$

the statistic $\hat{Q}$ remains unchanged by replacing $h(\vec{z}_i,\vec{z}_j)$ with ψ_{ij}, and we can express $\hat{Q}$ as

$$\hat{Q}=\sum_{i<j}\sum \psi_{ij} A_{ij} \ . \tag{A20}$$

The calculation of the conditional variance $V_c(\hat{Q})$ that is the conditional expected value of $\hat{Q}^2$, $E_c(\hat{Q}^2)$, is straightforward. By counting the terms in $\hat{Q}^2$ we obtain

$$\begin{aligned} V_c(\hat{Q})=&\tfrac{1}{2}N(N-1)E_c(A_{12}^2)\overline{\psi_{ij}^2} \\ &+N(N-1)(N-2)E_c(A_{12}A_{23})\overline{\psi_{ij}\psi_{jk}} \\ &+\tfrac{1}{4}N(N-1)(N-2)(N-3)E_c(A_{12}A_{34})\overline{\psi_{ij}\psi_{kl}} \ , \end{aligned} \tag{A21}$$

where the bars denote taking averages (with i, j, k, and l all different). For $\overline{\psi_{ij}\psi_{jk}}$ we obtain

$$\begin{aligned} \overline{\psi_{ij}\psi_{jk}}&=\frac{1}{N(N-1)(N-2)}\sum_{\substack{i \ j \ k \\ i\neq j, i\neq k, j\neq k}}\sum\sum \psi_{ij}\psi_{jk} \\ &=\frac{1}{N(N-1)(N-2)} \\ &\quad\times\left\{\sum_{\substack{i \ j \ k \\ i\neq j, j\neq k}}\sum\sum \psi_{ij}\psi_{jk}-\sum_{\substack{j \ k \\ j\neq k}}\sum \psi_{kj}\psi_{jk}\right\} . \end{aligned} \tag{A22}$$

It can be readily checked that the ψ_{ij} have the property

$$\sum_{\substack{i \\ i\neq j}} \psi_{ij}=0, \quad \text{for all } j \ , \tag{A23}$$

Using this relation, we can verify that the sum over k in the first term within the curly braces in (A22) equals zero. We have

$$\overline{\psi_{ij}\psi_{jk}}=-\frac{1}{N-2}\overline{\psi_{ij}^2} \ . \tag{A24}$$

Similarly, we can derive the expression

$$\begin{aligned} \overline{\psi_{ij}\psi_{kl}}&=-\frac{2}{N-3}\overline{\psi_{ij}\psi_{jk}} \\ &=\frac{2}{(N-2)(N-3)}\overline{\psi_{ij}^2} \ . \end{aligned} \tag{A25}$$

The conditional expected values of the products of random variables (i.e., their averages with respect to the possible choices of D) can be found by straightforward calculation, involving the relative number of times the indices in the products are within D and D^c. The results are

$$E_c(A_{12}^2)=\frac{2}{N(N-1)}\frac{2(N-1)(N-2)}{N_1N_2(N_1-1)(N_2-1)} \ , \tag{A26}$$

$$\begin{aligned} E_c(A_{12}A_{23})=&-\frac{2}{N(N-1)(N-2)} \\ &\times\frac{2(N-1)(N-2)}{N_1N_2(N_1-1)(N_2-1)} \ , \end{aligned} \tag{A27}$$

and

$$\begin{aligned} E_c(A_{12}A_{34})=&\frac{4}{N(N-1)(N-2)(N-3)} \\ &\times\frac{2(N-1)(N-2)}{N_1N_2(N_1-1)(N_2-1)} \ . \end{aligned} \tag{A28}$$

After substitution of these expected values into (A21), the final expression for the conditional variance reads

$$\begin{aligned} V_c(\hat{Q})=&\left[1+\frac{2}{(N-2)}+\frac{2}{(N-2)(N-3)}\right] \\ &\times\frac{2(N-1)(N-2)}{N_1N_2(N_1-1)(N_2-1)}\overline{\psi_{ij}^2} \\ =&\frac{2(N-1)^2(N-2)}{N_1N_2(N_1-1)(N_2-1)(N-3)}\overline{\psi_{ij}^2} \ . \end{aligned} \tag{A29}$$

We can write $\overline{\psi_{ij}^2}$ as

$$\overline{\psi_{ij}^2}=\overline{H_{ij}^2}-2\frac{N-2}{(N-1)}\overline{g_i^2} \ , \tag{A30}$$

using (A16)–(A18).

[1] P. Grassberger, T. Schreiber, and C. Schaffrath, Int. J. Bifurcation Chaos **1**, 521 (1991).
[2] J. Wright and R. L. Schult, Chaos **3**, 295 (1995).
[3] H. Kantz, Phys. Rev. E **49**, 5091 (1994).
[4] N. H. Anderson, P. Hall, and D. M. Titterington, J. Multivar. Anal. **50**, 41 (1994).
[5] J. Theiler, Phys. Rev. A **34**, 2427 (1986).
[6] F. Pukelsheim, Am. Stat. **48**, 88 (1994).
[7] C. Diks, J. C. van Houwelingen, F. Takens, and J. DeGoede, Phys. Lett. A **201**, 221 (1995).
[8] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank, and J. D. Farmer, in *Nonlinear Modeling and Forecasting,* edited by M. Casdagli and S. Eubank (Addison-Wesley, Reading, MA, 1992).
[9] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).
[10] J. Theiler, Phys. Rev. A **41**, 3038 (1990).
[11] W. R. van Zwet, Z. Wahrscheinlichkeitstheorie verw. Gebiete **66**, 425 (1984).