Testing stationarity in time series

A. Witt,* J. Kurths, and A. Pikovsky
Department of Physics, Nonlinear Dynamics Group, Universität Potsdam, Postfach 601553, D-14415 Potsdam, Germany
(Received 29 October 1997)

We propose a procedure for testing stationarity of time series by combining a test for time independence of the 1D probability density with one of the spectral density. The potentials and limits of this test procedure are established by its application to different types of numerically generated time series ranging from simple linear stochastic processes to high-dimensional transient chaos as well as to observational data from geophysics and physiology. Problems of practical implementation are discussed, in particular the relation between the lengths of the time series and its maximal relevant time scales. Furthermore, artifacts and counterexamples are presented. [S1063-651X(98)14708-6]

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

I. INTRODUCTION

During the last decade many methods for nonlinear signal processing have been proposed. These include estimations of dimensions, entropies, or mutual information, and modeling or prediction of data series using local linear models, radial basis functions, neuronal networks, and nonlinear stochastic processes [1,2]. Most of them assume (implicitly) stationarity of the time series under study. However, detecting stationarity in a time series is not an obvious task. Especially, observations of natural systems are marked by influences of several external processes, which might lead to nonstationarity or long-range correlations. Therefore, it is important to have a procedure that allows one to check whether a time series is stationary or not and that can additionally detect stationary regions in an observational record.

In Fig. 1 we plot three time series to illustrate the problem: (a) a realization of a first-order autoregressive process, which is by construction stationary, (b) a realization of a fractional Brownian motion as an example of a nonstationary process, and (c) a record of heart rate variability of a human subject. This is measured as RR intervals, i.e., the duration between heart beats. Here, the question of stationarity is open. The goal of this paper is to propose a technique that enables one to answer this question.

There already exist quite a number of statistical tests for stationarity. Several attempts grasp the notion of stationarity from the viewpoint of dynamical systems [3–5]. These investigations have led to improved qualitative descriptions of the data, however, they do not result in quantitative characteristics. Some tests for stationarity have been developed in the frame of mathematical statistics [6]. Due to rather strong assumptions on the time series, which are often difficult to check, they seem to be less suitable for our purposes. Furthermore, it has been taken into account that the detection of stationarity requires an observational length which is large in comparison to the typical time scales of the underlying process. In this sense, the discussion about stationarity is closely connected with the question of correlation length or long-range correlation.

*Electronic address: annette@agnld.uni-potsdam.de

The goal of this paper is to propose a statistical test that can be applied for a time series consisting of a few thousand elements (Sec. II). In constructing it, we employ the notion of stationarity used in both mathematical statistics and the theory of dynamical systems.

The properties of this test procedure are demonstrated in Sec. III A by a comparative application to a broad variety of time series from different physical models, where also such special cases as fractional Brownian motion and high-dimensional transient chaos are considered. Thereby it will be pointed out that our procedure enables us to test for a stronger demand than weak stationarity. In Sec. III B we discuss further properties of this algorithm and especially its prerequisites as well as its limits, artifacts, and counterexamples. Then, in Sec. III C, we apply the procedure to experimental data from geophysics and physiology, which have a lot of the unwelcome properties typical for observational data like trends or measurement noise. Finally, the results are summarized and discussed in Sec. IV.

II. DESCRIPTION OF THE TESTS

In this section mathematical definitions of stationarity are recapitulated and the notions needed for the derivation of the test statistic are introduced.

A. Definitions of stationarity

Roughly speaking, a time series is called stationary if its essential statistical properties do not depend on time. In mathematical statistics two types of stationarity are distinguished—strong (or complete) and weak stationarity. They are defined as follows [7]:

A stochastic process $\{X_t\}$ with $t \in \mathbb{N}$ is called *strongly stationary* if for any set of times $t_1, t_2, \ldots, t_n$ and any integer k the joint probability distributions of $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ and of $\{X_{t_1+k}, X_{t_2+k}, \ldots, X_{t_n+k}\}$ coincide. In the language of dynamical systems this means that in each conceivable embedding space the statistical properties of the phase flows referring to different pieces of the time series are the same.

A less strict demand is *weak stationarity*, where only the first and second moments have to exist and have to be independent of time, i.e.,

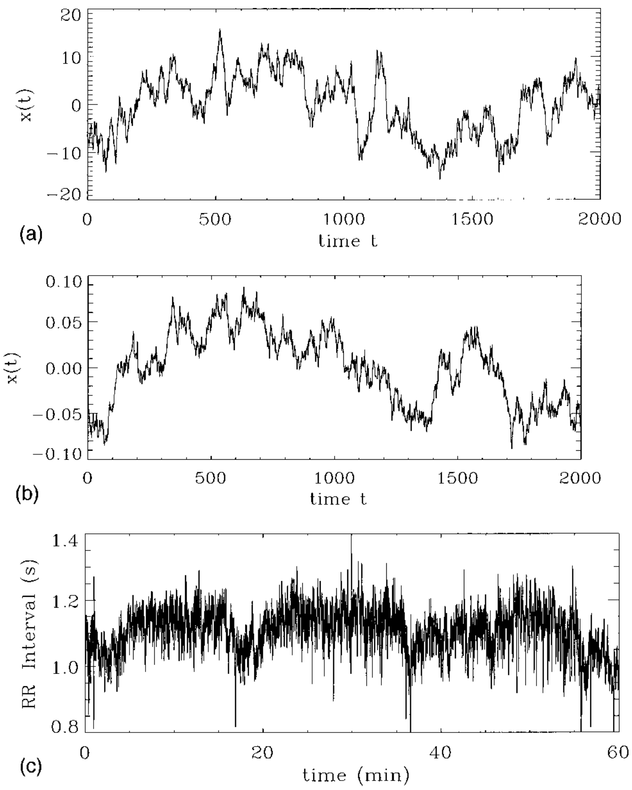

FIG. 1. (a) Realization of a first-order autoregressive process with $a_1=0.99$. (b) Realization of a fractional Brownian motion process with a scaling exponent $\alpha=1.75$. (c) Time series of RR intervals, i.e., the duration between heart beats over 1 h of a healthy subject.

$$\langle X_t\rangle=\mu, \tag{1}$$

$$\langle (X_t-\mu)(X_t-\mu)\rangle=\sigma^2, \tag{2}$$

where $\langle\cdot\rangle$ stands for the ensemble average, μ and σ are constants independent of t. In addition, the autocorrelation function $\rho(t,s)=\langle (X_t-\mu)(X_s-\mu)\rangle/\sigma^2$ has to depend only on the relative time delay $\tau=t-s$, i.e.,

$$\rho(t,s)=\rho(t-s)=\rho(\tau). \tag{3}$$

If the process is weakly stationary, the power spectrum $P(f)$ exists and can be expressed as the Fourier transform of ρ:

$$P(f)=\frac{1}{2\pi}\sum_{r=-\infty}^{\infty}\rho(r)e^{-i2\pi fr}, \quad -\frac{1}{2}\leqslant f\leqslant\frac{1}{2}. \tag{4}$$

Obviously, strongly stationary processes are also weakly stationary.

Examples of strongly stationary processes are Gaussian distributed white noise processes, autoregressive or moving average processes. Simple counterexamples are white noise with nonconstant (time-dependent) standard deviation or autoregressive processes with varying coefficients. Periodic fluctuations between stationary processes, e.g.,

$$Z_t=\begin{cases} X_{T/2} & T \quad \text{even},\\ Y_{(T-1)/2} & \text{if} \quad T \quad \text{odd},\end{cases} \tag{5}$$

where X_t and Y_t are distinct stationary processes, and are nonstationary in general as well. The main emphasis of this paper, however, is put on testing long-time nonstationarity.

From the physical point of view systems are called stationary if their main physical properties do not change with time. The purely deterministic character of many physical systems does not conflict with the contemplation of their signals as realizations of stochastic processes. If an observational series is analyzed, it is often *a priori* unknown whether the underlying process has a stochastic or a deterministic character or is a mixing of both. In the language of dynamical systems the equivalent for (strong) stationarity is the existence of an invariant ergodic measure [8,9]. Hence, there are close links between statistical and physical views about stationarity.

B. The tests

In this paragraph the test statistics are introduced. An important problem in testing stationarity in observations of natural systems is that only realizations of the system under study are known instead of the system itself. Stationarity, however, is a property of the process. Each test assumes implicitly that the time series is typical for the system and can only give an upper limit for the degree of stationarity violation.

A further difficulty is that usually only a *single* realization $\{x_t\}_{t=1}^{n}$ of the system under study is available. Subdividing the sequence $\{x_t\}$ into several parts $\{x_*^i\}$, we artificially produce a *set* of data series that can be compared with statistical tools for correspondence.

It is important to note that it is never possible to truly establish strong stationarity in experimental data, since either the time independence of all central and noncentral moments $M_k^{t_1t_2\cdots t_k}(t)=\langle (X_{t+t_1}-\mu)(X_{t+t_2}-\mu)\cdots(X_{t+t_k}-\mu)\rangle$ or the time independence of the probability density of any projection into lower dimensional spaces $(t_1,t_2,\ldots,t_n)$ had to be tested. Therefore, we propose to test the following hypothesis whose requirements include those of weak stationarity but are weaker than strong stationarity: (A) The one dimensional (1D) probability density is independent of time, and (B) the power spectral density is independent of time.

1. Test A: Time independence of probability distributions

We take up an idea of Isliker and Kurths [10] who have proposed to test the time independence of the probability distribution of a single time series $\{x_t\}_{t=1}^{n}$ as a necessary assumption for strong stationarity and hence for the existence of an invariant ergodic measure:

(1) The time series $\{x_t\}_{t=1}^{n}$ is divided into l parts (windows) $\{x_\tau^j\}$ of equal length n_w:

$$x_\tau^j=x_{(j-1)n_w+\tau}, j=1\cdots l, \tau=1\cdots n_w. \tag{6}$$

The mean value and the variance of the jth window are denoted by μ^j and σ^j. The problem of finding an appropriate choice of the subsequence lengths l (window length) is discussed in Sec. II C.

(2) For the comparison of the probability distribution of the ith and the jth window a modified χ^2 test is applied. As

usual for χ^2 tests, the elements of both windows are coarse grained with the same binning, say in r bins, such that the kth bin of the ith window contains the elements $X_k^i = \{x_{\rho_1}, x_{\rho_2}, \ldots, x_{\rho_{R_k^i}}\}$. The only condition imposed on the binning is that the number of elements R_k^i in each bin is greater than 20. If the time series elements were uncorrelated, R_k^i could be understood as a realization of a binomial random variable which has a variance σ^2 that is of the same magnitude as R_k^i and even this property is used in the construction of the χ^2 test variable. In the general case of correlated time series, however, the R_k^i's are not binomially distributed. This requires a direct estimation of this variance from the time series.

In the following we call the variance of the number of occurrences R_k^i of the time series in the kth bin with respect to the ith window $\sigma^2(R_k^i)$. We suggest to estimate this quantity $\sigma^2(R_k^i)$ from the variance $\sigma^2[D_k^i(m)_{m=1}^{R_k^i-1}]$ of the index number distances $\{D_k^i(1), D_k^i(2) \ldots D_k^i(R_k^i-1)\} = \{\rho_2-\rho_1, \rho_3-\rho_2, \ldots, \rho_{R_k^i}-\rho_{R_k^i-1}\}$ with respect to the elements of the kth bin inside the ith window as

$$\sigma^2(R_k^i) = c\ \sigma^2[D_k^i(m)_{m=1}^{R_k^i-1}]. \tag{7}$$

The variable c depends on the window length n_w and on the number of elements in the bin R_k^i. In the case of an uncorrelated time series, we can determine c analytically: The number of trials falling in the kth bin is a realization of the binomially distributed random number X_R characterized by the parameters (n_w, p). From a realization this probability p can be estimated by $\hat{p} = R_k^i/n_w$. The variance $\sigma^2(X_R)$ of this random variable reads $\sigma^2(X_R) = n_w p(1-p)$. The random number X_D of the distances of trials of X_R falling in the kth bin is geometrically distributed with the same parameters (n_w, p). This leads to a variance of $\sigma^2(X_D) = (1-p)/p^2$. Consequently the fraction of variances of both processes equals

$$c = \frac{\sigma^2(X_R)}{\sigma^2(X_D)} = \frac{n_w p(1-p)}{(1-p)/p^2} = n_w p^3.$$

This variable c can be estimated from a realization as

$$\hat{c} = (R_k^i)^3/n_w^2. \tag{8}$$

This equation (8) holds further in the case of simple periodic behavior, because both variances $\sigma^2[D_k^i(m)_{m=1}^{R_k^i-1}]$ and $\sigma^2(R_k^i)$ vanish. We have empirically checked that Eq. (8) is valid for several types of correlated time series as well (cf. Sec. III). The test statistic includes exclusively means and variances of index number distances. This restricts its applicability to time series whose distributions of the index number distances can be completely described by their first and second moments. This condition is often equivalent with either pure linear correlations or a significant influence of noise to the system.

The modified χ^2 test statistic reads

$$t_{A,2} = \sum_{k=1}^{r} \frac{(R_k^i - R_k^j)^2}{\sigma^2(R_k^i) + \sigma^2(R_k^j)} \tag{9}$$

$$= n_w^2 \sum_{k=1}^{r} \frac{(R_k^i - R_k^j)^2}{(R_k^i)^3 \sigma^2[D_k^i(m)_{m=1}^{R_k^i-1}] + (R_k^j)^3 \sigma^2[D_k^j(m)_{m=1}^{R_k^i-1}]} \tag{10}$$

and is asymptotically $(n_w \to \infty, r \to \infty)$ χ^2 distributed with r degrees of freedom.

(3) If the probability distribution of *several* windows should be compared, then one can employ the test statistic

$$t_{A,l} = l n_w^2 \sum_{k=1}^{r} \sum_{i=1}^{n_w} \frac{(R_k^i - R_k/li)^2}{(R_k+1)^3 \sigma^2[D_k(m)_{m=1}^{R_k^i-1}]}. \tag{11}$$

R_k denotes the number of elements (of all l windows) in the kth bin and $\sigma^2(D_k)$ is the variance of their index number distances. $t_{A,l}$ is χ^2 distributed with $r(n_w-1)$ degrees of freedom.

By comparing the probability densities of the windows in principle the time independence of *all* central statistical moments is tested. Due to the coarse graining and further finite-size effects, however, it does not do so in practice.

2. Test B: Time independence of power spectra

Now the time independence of the second order noncentral moments, which are represented in the spectral distribution, is checked. If additionally the mean value is a constant, especially if the series has passed test A, the time series appears as a realization of a weakly stationary process by definition.

The most important step of this test procedure is to transform the data into samples of the spectral distribution densities, i.e., with respect to each window to get a set of data $\{f_i\}_{i=1}^{k}$ that are identically and independently distributed (i.i.d.) with the spectral distribution.

(1) For the subseries obtained for test A, the autocorrelation functions are estimated:

$$\rho^j(k) = \frac{(n_w-k)^{-1} \sum_{\tau=1}^{n_w-r} (x_\tau^j - \mu^j)(x_{\tau+k}^j - \mu^j)}{\sigma^j} \quad \text{with} \quad k = 0, \ldots, n_\rho. \tag{12}$$

To characterize their properties it is transformed into a realization of the power spectral density: Taking a set of uniformly distributed uncorrelated random variables $\{\zeta_k\}_{k=1}^{n_\rho}$ the solutions $\{f_k^j\}_{k=1}^{n_\rho}$ of

$$\int_0^{f_k^j} P^j(f)\, df = \zeta_k \tag{13}$$

are a finite number of random variables. P^j is the Fourier-transformed autocorrelation function ρ^j of the jth window point [see point (4)]:

$$P^j(f)=\sum_{r=-n_\rho}^{n_\rho}\rho^j(r)\cos(2\pi fr). \tag{14}$$

The choice of uniformly distributed $\{\zeta_k\}_{k=1}^{n_\rho}$ leads to statistical independence of the $\{f_k^j\}_{k=1,\dots,n_\rho}$. [It seems more obvious to use a set of equidistant ζ_k, $k=1,\dots,n_\rho$, instead of uniformly distributed ones. However, in case of $n_\rho \ll n_w$ equidistant ζ_k's lead to values of $\{f_k^j\}$ that are near the exact solution ($n_\rho \to \infty$), they cannot be regarded as a realization of a continuous random number. By choosing the ζ_k's uniformly distributed exact these effects are suppressed.] Having different sets f_*^i and f_*^j for different windows i and j we want to check whether they belong to the same power spectral distributions. By considering these power spectral distributions as probability distributions we can interpret f_*^i and f_*^j as two samples and can compare them by statistical tests as a χ^2 test of homogeneity or a Kolmogorov-Smirnov test.

(2) Similar to test A we apply a χ^2 statistic. If the spectral density of the ith and the jth window are compared, then one uses the test statistic

$$t_{B,2}=\sum_{k=1}^{r}\frac{(R_k^i-R_k^j)^2}{R_k^i+R_k^j} \tag{15}$$

with respect to the chosen binning, where R_k^i denotes the number of f_*^i's in the kth bin. This test statistic $t_{B,2}$ is χ^2 distributed with $r-1$ degrees of freedom.

(3) If it is hypothesized that $\{f_j^i\}_{j=1,\dots,n_\rho}$ represent the same distribution for all $i, 1\leq i\leq l$, then the test statistic t_{B,n_w} reads

$$t_{B,n_w}=l\sum_{i=1}^{l}\sum_{k=1}^{r}\frac{(R_k^i-R_k/l)^2}{R_k}, \tag{16}$$

which is χ^2 distributed with $(r-1)(l-1)$ degrees of freedom if $R_k=\Sigma_i R_k^i$.

If the hypothesis is rejected, i.e., the sequence is nonstationary, we can compare mutually the samples $\{f_i^j\}_{i=1,\dots,n_\rho}$ with respect to different windows. In this way we can detect whether the structure of the data series is generally inhomogeneous (as in transient states) or whether there are only some parts (windows) with a special structure (e.g., due to bursts).

3. Combined test

The procedures A and B explained above can be applied separately. Sometimes, it might be preferable to have only one test statistic that contains the results of both procedures. Since the χ^2 statistic is used, a combined test is obtained by simple addition. Consequently, if two subseries are compared the final test statistic is

$$t_2=t_{A,2}+t_{B,2}, \tag{17}$$

and in the case of comparison of n_w windows one uses

$$t_l=t_{A,l}+t_{B,l}. \tag{18}$$

The new statistics are again χ^2 distributed, and the degrees of freedom are the sum of those of the summands. In the next section it is explained how a minimal number of sample elements can be determined.

C. Width of the windows

One essential assumption of the tests explained above is that each of the subseries can be regarded as stationary. Therefore, they should be chosen to be long enough so that each of them presents all essential (especially, the long-range) properties of the time series. On the contrary, we would like to have *at least several* subseries so that a comparison between them is possible. Thus a compromise between a good representation of the long-frequency shares in each window and the largest possible number of windows has to be attained. In the Appendix, we present a method that semiempirically derives a minimal possible window length for a given time series.

Due to strong long-scale influences, it may be impossible to choose such an acceptable window length (e.g., in case of fractional Brownian motion). In some cases one has to accept that the time scales of the underlying physical process are too large in comparison with the length of the data series, sometimes, however, this can be diminished by filtering, which is discussed in the next subsection.

D. Filtering

For reducing trends and long-scale influences in data series we apply two types of filters [11]: (i) Locally linear detrending: This simple manipulation reduces long-range properties very effectively. The filter length is chosen on a case by case basis. Butterworth filter: Here all waves that are longer than a fixed wavelength vanish. The result is similar to the above one, but the changes in the power spectrum are more fundamental.

The resulting series include short- and intermediate-scale components. It is also possible that they have nonstationary properties. Therefore, it is necessary to check the filtered data for a hidden nonstationarity.

III. APPLICATIONS

To study the potentials and limits of the proposed test, we first apply it to series obtained from known dynamical systems. Later we test observational data series for stationarity.

A. Numerically generated time series

The validity of the test statistic introduced above is checked by computing the test statistic t_2 or t_l [Eqs. (17), (18)] for 1000 different realizations of the special system. This way the observed distribution of the test statistic can be compared with the hypothesized one. Further we consider different window lengths, bin sizes, etc.

1. Stochastic processes

a. Autoregressive processes. Autoregressive (AR) processes are standard examples of (strongly) stationary stochastic processes [6]. An AR process of order k is defined as

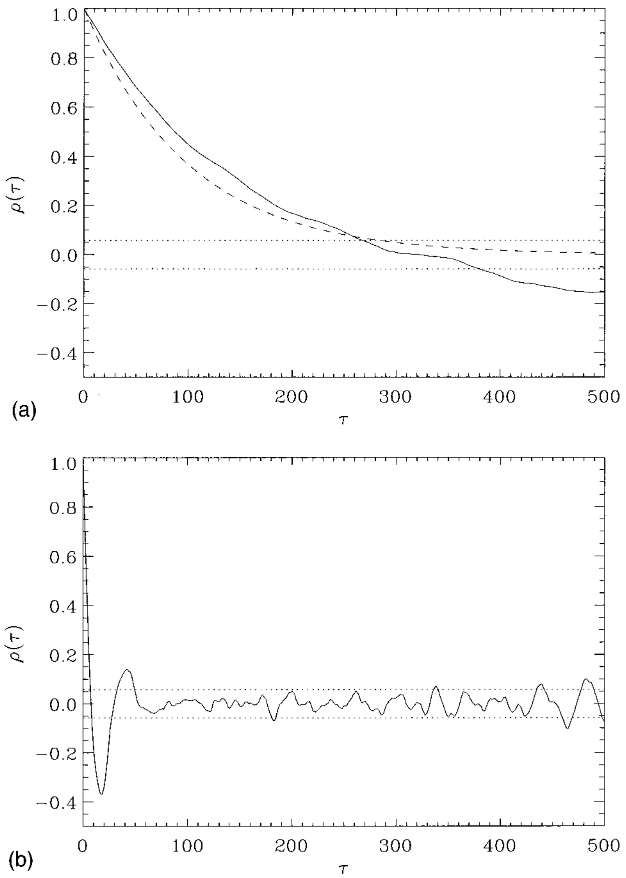

FIG. 2. (a) Autocorrelation function of a first order AR process with $a_1=0.99$ (solid line, special realization; dashed line, theoretical value) with significance level f_{SL} (dotted lines). (b) Autocorrelation function of the same but filtered record.

$$X_t=\sum_{i=1}^{k} a_i X_{t-i}+\xi_t, \tag{19}$$

where ξ_t is a Gaussian i.i.d. process. The power spectrum P_X has the form

$$P_X(f)=(2\pi)^{-1}\frac{1}{\left|1+\sum_{j=1}^{k} a_j e^{-ijf}\right|^2}. \tag{20}$$

Note that it is independent of t.

We have simulated the distribution of t_2 for several AR processes as described. If the correlation time is small enough, then the test statistic converges to the target distribution. As simplest models we use first order AR models ($n=2000$, $n_w=500$, $r=10$, comparison of first and last window). The hypothesized statistics of t_2 is a χ^2 distribution with 19 degrees of freedom. It is reached for $|a_1|<0.7$. Larger absolute values of a_1 lead to a right shift of the distribution, i.e., the test hypothesis is rejected with higher than error probability. This effect is caused by the window length of 500 data points, which is too small compared to the correlation length.

The problem becomes worse for a first order AR process with a coefficient near 1. In Fig. 2(a) the estimated autocorrelation function for a first order AR process with $a_1=0.99$ (i.e., the correlation decay is about 100) and a data length of 5000 points is plotted. It is clearly seen that the strong correlations lead to a ρ that lays outside the significance level SL(5000,700) (for more details cf. the Appendix). That means that for a window length of $n_w=700$ the test provides with high probability a negative result, i.e., it rejects stationarity. This is due to both the short data set and the long correlation length. The structure of such a time series calls for a very long window length; the procedure in the Appendix recommends a minimal window length of $n_w=3500$ time steps. Since this is often not realistic, filtering can be an alternative: In Fig. 2(b) the autocorrelation function for the Butterworth-filtered sequence is given, where all components longer than 40 time steps are reduced. This allows a window length of 500 points. For these filtered data the test statistic is distributed as hypothesized. The correlation time, however, which is an essential property of that process, is diminished.

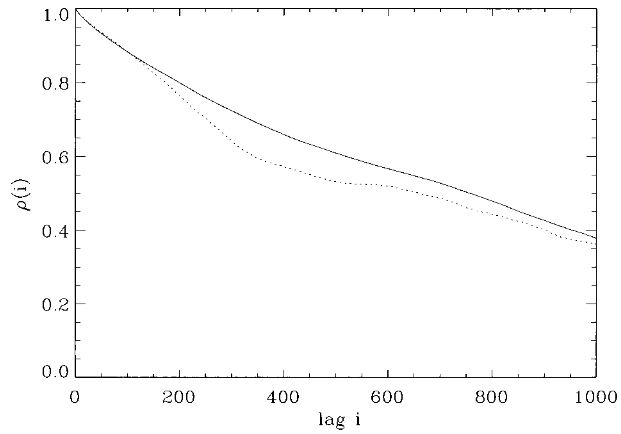

FIG. 3. Autocorrelation function for FBM with $\alpha=1.75$; full line: ρ of the complete series (10 000 data points), dashed line: ρ of the first half series.

b. Fractional Brownian motion. Several natural processes such as heart rate variability or languages [12,13] are characterized by an $1/f$-like power spectrum

$$P(f)\sim f^{-\alpha} \tag{21}$$

with a scaling exponent α. Osborne and Provenzale [14] proposed a simple procedure to generate signals, called fractional Brownian motion (FBM), whose power spectra have such a power-law dependence. For such FBM the standard deviation depends on the window length $n_w:\sigma^2(n_w)\sim n_w^{\beta}$, i.e., FBM is not weakly stationary similar to the behavior of random walks. Furthermore, the FBM has by construction long-term correlations, which are amplified by an increasing coefficient α.

A realization of FBM with $\alpha=1.75$ is shown in Fig. 1(b). The singularity of its power spectrum at $\omega=0$ is characterized by the exponent α. This singularity induces correlation lengths in the magnitude of sequence length, the resulting time series are nonstationary. So one finds the test statistic t_A quite larger than expected: for $\alpha=1.75$, $n=1000$, $n_w=500, r=8$ the mean value of $t_{A,2}$ is 50 instead of 8. This is strengthened by a growing parameter α.

It is important to notice that the autocorrelation functions (normalized to have variance 1) for subseries with different length coincide (cf. Fig. 3). In this sense, FBM exhibits a self-similar behavior in the linear correlation structure. Consequently, the spectral densities of different windows coincide and the test statistic t_B is χ^2 distributed with the expected degrees of freedom.

This example shows impressively that different properties of stationarity need to be tested. A geophysical observation with a similar structure is discussed by Kurths *et al.* [15].

2. Low-dimensional nonlinear systems

In this section deterministic systems in rather complex states are investigated. As mentioned in the Secs. II A and II B, it is useful to test stationarity also in deterministic systems. Especially, in the case of chaotic regimes, the exponential growth of the distance between trajectories that are initially nearby can be interpreted as a production of information, i.e., they have a strong correlation decay.

a. Skew tent map. As a paradigmatic model of discrete chaotic systems we analyze the dynamics of the simple skew tent map:

$$x_{n+1}=\begin{cases}\dfrac{x_n}{a} & \text{if} \quad x_n<a\\[2ex] \dfrac{x}{a-1}-\dfrac{1}{a-1} & \text{if} \quad x_n\geqslant a,\end{cases} \tag{22}$$

with $x_0\in(0,1)$ and $a\in(0,1)$. It is well known that the natural measure of this system is the uniform distribution on the unit interval and the autocorrelation function decreases exponentially [16]. If the correlation length is small, e.g., the control parameter satisfies $0.25<a<0.75$, the test statistic t_2 and t_l ($n_w=500, r=10$) are distributed as expected. Other values of a induce stronger correlation, which leads to the same problems as discussed for AR models.

b. Logistic map and time continuous nonlinear systems. Another popular example of a nonlinear system is the logistic map

$$x_{n+1}=rx_n(1-x_n). \tag{23}$$

Chaotic behavior of the x_n occurs for many of the control parameters when $3<r<4$. The results of the stationarity test are completely different from that for the skew tent map. The simulated distribution of $t_{A,2}$ does not coincide with the expected distribution (cf. Fig. 4). The main reason is that the nonlinear behavior here is caused by a quadratic nonlinearity, whereas the skew tent map is piecewise linear. Due to this nonlinearity the distribution of index number distances for the elements of a bin is structured very complicatedly, it can have gaps or singularities. In particular, it cannot be completely described by the mean value and standard deviation. Therefore, the distribution of $t_{A,2}$ may be far from the expected one (cf. Fig. 4).

The test for the independence of the power spectrum, however, works, i.e., the statistic $t_{B,2}$ is distributed as hypothesized.

For typical chaotic time-continuous systems such as the Lorenz or the Duffing oscillator we get analogous results—the nonlinear deterministic character of the system leads to distributions of the test statistic t_A and consequently t, which differ significantly from the expected ones.

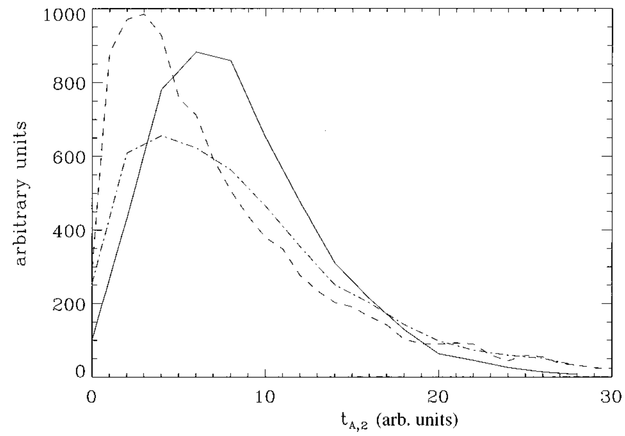

FIG. 4. Distributions of $t_{A,2}$ for time series of the logistic map with $r=3.7$ (dashed) and $r=3.58$ (dotted), for windows of 500 points and a coarse graining of 10 bins. The full line represents the target distribution.

c. Strange nonchaotic attractors. As an example of a deterministic process at the border between regularity and chaos, we analyze the trajectories of strange nonchaotic attractors [17] of the system

$$x_{n+1}=f(x_n,\theta_n)=3(\tanh x_n)\cos(2\pi\theta_n), \tag{24}$$

$$\theta_{n+1}=(\theta_n+\omega)\bmod 1, \tag{25}$$

where $\omega=(\sqrt{5}-1)/2$ is the inverse golden mean. These trajectories approach an attractor which is not chaotic (negative Lyapunov exponents), but has a fractal geometry. Furthermore, the spectrum has a fractal structure, i.e., it is singular continuous [18]. Even for such degenerated spectral distributions the test statistic t_B attains its expected distribution. The results with respect to test A are similar to those of the logistic map in the regime of 2- or 4-band attractors (cf. Fig 4, $r=3.58$). The strong periodic component in the behavior of the trajectories is reflected in a left shift of the distribution of t_A.

3. Kuramoto-Sivashinsky equation

As an example of high-dimensional systems we investigate solutions of the one-dimensional Kuramoto-Sivashinsky equation (KS) [19]

$$\frac{\partial u}{\partial t}+4\frac{\partial^4 u}{\partial x^4}+\alpha\left(\frac{\partial^2 u}{\partial x^2}+u\frac{\partial u}{\partial x}\right)=0 \tag{26}$$

subject to periodic boundary conditions $0<x<2\pi$. Here it is convenient to study the norm (or energy) $s(t)=\int u^2(x)dx$ of solutions u in dependence on (sampled) time. The KS equation possesses a rich bifurcation scenario; different periodic as well as chaotic branches are known. Moreover, transient states are typically found.

a. Chaotic regime. For the parameter value $\alpha=134.0$, a chaotic solution of Eq. (26) exists. As in the examples of low-dimensional chaos, the test statistic t_B indicates stationarity, i.e., it is χ^2 distributed with the correct number of degrees of freedom. This shows that the procedure is also

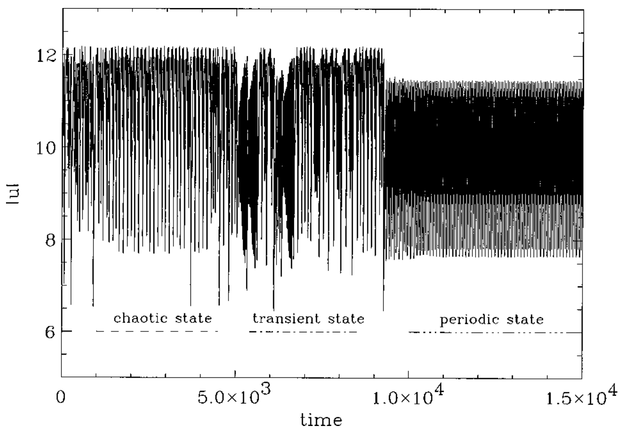

FIG. 5. Norm of a solution of the Kuramoto-Sivashinsky equation Eq. (26) over time. The three different states (quasichaotic, transient, periodic) are recognizable.

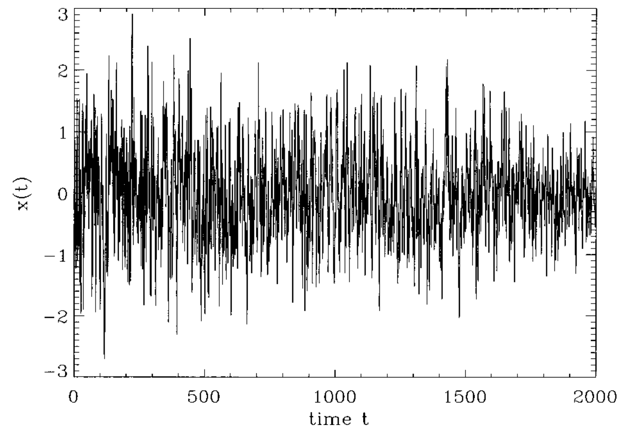

FIG. 6. Example of a standard-deviation normalized first order AR process with moving coefficient $a_1 = 0.6, \ldots, 0.9$

appropriate for analyzing high-dimensional systems if a suitable reduction to a one-dimensional subspace is found, as the norm in this case.

b. Transient chaos. A quite different behavior is observed for the parameter value $\alpha = 137.0$, where transient chaos occurs [19]. After a finite chaotic phase, the system comes via a transient state into the final periodic regime. As expected, the test indeed finds non-stationarity for the whole series, as plotted in Fig. 5. Comparing mutually the power spectra of the windows with test *B*, the two different dynamical regimes are identified as stationary subsequences, in particular, the periodic regime where stationarity is found independently of the window length as well as the initial nonstable chaotic regime, which is recognized if the window length is larger than 500 points. For the piece of the series between both the test rejects stationarity independently of the window length, i.e., instead of suddenly jumping from the chaotic into the periodic regime a transient state is passed. It should be mentioned that this method allows not only the detection but the localization of the different dynamical regimes. Lyapunov exponents that applied usually lead to more rough approximations.

4. Examples of nonstationary processes

The transient phase of the Kuramoto-Sivashinsky equation gives an example of a nonstationary time series. Now we study processes that are in general nonstationary.

a. Autoregressive processes with varying coefficients. As a generalization of autoregressive processes, we consider here such processes of first order with varying coefficients:

$$X_t = a(t) X_{t-1} + \xi_t \quad \text{with} \quad -1 < a(t) < 1. \tag{27}$$

These processes are nonstationary by construction if $a(t)$ is nonstationary.

Regarding such processes, where a linearly depends on t ($n = 2000$, a increases from 0.5 to 0.7, the first 800 data points compared with the last 800 points), we have found for the test statistic t_2 instead of a χ^2 distribution with a mean of 15 a distribution with a mean of about 23. This is equivalent to a rate of about 45% finding the series nonstationary. If the parameter a changes only between 0.5 and 0.6, then the distribution of t_2 has a mean of 17.4; i.e., the sequence seems to be a realization of a stationary process with high probability.

If longer sequences (e.g., $n = 4000$ as given with the procedure of the Appendix) are used, the distribution of t_2 is more distant from the target distribution, and, therefore, the nonstationarity is detected with higher probability.

This means that only sequences with either a strong nonstationarity or with a sufficient length can be recognized as nonstationary. Or, in other words, the time scales on which the process varies have to be small in comparison with process length.

b. Tent maps with varying skewness. As a deterministic counterpart, tent maps with a growing parameter a, $0 < a < 0.75$, are considered. Test *A* indicates stationarity, since the probability density is a uniform-distributed one, i.e., it is independent of a. Only Test *B* is sensitive to the structural changes.

This example proves that the test for the time independence of the probability distribution *alone* is not sufficient even for testing weak stationarity.

B. Artifacts and apparent counterexamples

In this section we demonstrate some cases where special effects of the test procedures are underlined and their limits are shown.

The first example is a standard-deviation normalized first-order AR process with varying coefficients:

$$X_t = \sqrt{1 - a(t)^2}\, [a(t) X_{t-1} + \xi_t], \tag{28}$$

whereby a depends linearly on t (cf. Fig. 6). The probability density of X_t is by construction Gaussian-normal for all t. The time invariance of this probability distribution is confirmed by test *A*. But test *B* finds that there are structural changes in the time series—the power spectrum depends on t. The characteristics of $a(t)$ and the window length are the same as for Eq. (27). Hence it seems to be that this is an example for a strongly but not weakly stationary process—in contradiction to the definitions. The explanation for this phenomenon is that the test for the time independence of the probability density is only a necessary condition for strong stationarity, but not a sufficient one. Analogous results hold for the above mentioned tent maps with varying skewness.

TABLE I. Overview of the results: The different analyzed processes are given in the first column, their properties with respect to the statistical definitions of stationarity are displayed in the second and third columns, the last two columns contain the results of the test.

Process	Weakly stationary	Strongly stationary	Test *A*	Test *B*
AR	+	+	passed	passed
FBM	-	-	not passed	passed
Skew tent map	+	+	passed	passed
Logistic map	+	+	not passed	passed
Nonlinear time-continuous systems	+	+	not passed	passed
Strange nonchaotic attractors	+	+	not passed	passed
Kuramoto-Sivashinsky equation				
Chaotic regime	+	+	not passed	passed
Transient state	-	-	not passed	not passed
AR with varying coefficients	-	-	not passed	not passed
Skew tent map with varying skewness	-	-	passed	not passed

The opposite behavior occurs if we compose a series whose first part is Gaussian distributed white noise and the second part is uniformly distributed noise (with the same mean value and standard deviation). Here the noncentral second moments are independent of time (delta functions), whereas the probability density is time dependent. Similar structures are exhibited by a series that is composed of a nonlinear trajectory and its phase-randomized surrogate [20]. Both are examples of weakly but not strongly stationary processes.

Another limit of the test procedure is explained by the following example: If a concatenation of uniform distributed white noise and a trajectory of a skew tent map with $a=0.5$ is considered, the first and second moments as well as the probability density coincide. The tests give rise to stationarity. But, the series is only stationary in the weak sense, since the skew tent map is characterized by nonzero higher noncentral moments as opposed to a white noise process. This example emphasizes that only the central moments and the noncentral moments of second order are tested. All other noncentral moments are not taken into consideration.

The results of the test with respect to all processes discussed above are summarized in Table I. The statements done there assume a subsequence length as determined with the technique proposed in the Appendix.

C. Observational data

Finally, we apply these techniques to outdoor data, which are observations from natural processes. Opposite to times series, which are obtained by laboratory experiments, these observations often cannot be repeated or easily manipulated. Furthermore, several (nonstationary) measurement errors have to be diminished by filtering.

1. Geophysical observations

The changes in the atmospheric radiocarbon isotope ^{14}C as given by the decay-corrected $\Delta^{14}C$ activity in tree rings provide essential information about long-term solar variations [21,22]. The influence of the geomagnetic field to the ^{14}C production leads to a trend in the data series that is extracted by a local linear underground subtraction (cf. Fig. 7). The solar cycle of about 200 years is clearly recognized in the power spectrum [23]. Stationarity is found in the (filtered) data if windows longer than $n_w = 200-1000$ years are used. The only exception are the last 150 years (30 data points). If these data points are included the power spectrum of the last window is essentially different from all other ones. This seems to be in accordance with the industrial revolution (human impact).

2. Physiological data

As mentioned in the Introduction, in this section we deal with heart rate variability (HRV). Such records are extracted from long-time ECG records (over 60 min at rest and 24 h during a normal day with a transportable ECG device). They describe the time differences between two adjacent heart beats. The underlying dynamical behavior is understood only partially (see [24] and references therein). Moreover, this dynamical process might be influenced by several external perturbations such as exercise, the circadian rhythm, or acoustics.

There have been attempts to characterize the dynamics of HRV quantitatively [24,25]. But only seldomly the problem

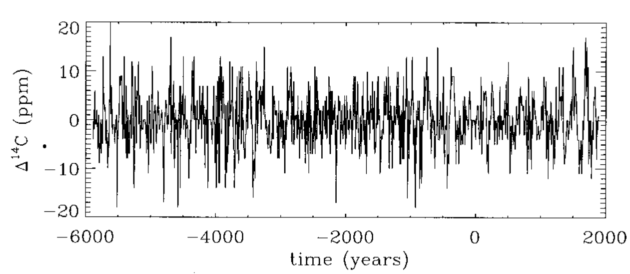

FIG. 7. The filtered $\Delta^{14}C$ record.

of stationarity was taken into account, though this is an assumption of most techniques. As is well known, the HRV contains strong long-range correlations: So $1/f$ behavior in the power spectra was reported in [26] and Bigger *et al.* [27] analyzed the power in the frequency band between 0 and 0.0033 Hz as a signature of some temperature regulating processes. Even such long-range processes pretend low dimensional behavior.

We were interested in getting the longest parts of such an HRV series, which could be assumed to be stationary. The sleeping phase in the 24 h records seems to be the most promising one.

The trend of the data was subtracted by the Butterworth filter. In the 24-h records all frequencies <0.0033 Hz and in the 1-h records all frequencies <0.015 Hz were deleted. Furthermore, the standard deviation was locally normalized. Applying the tests we found in the 1-h records stationary parts of 15–45 min, for the 24-h records such parts have a length of maximal 70 min.

In the analysis of HRV series we meet again the problem of the relation between the time scales of the signal and the observational length. As described above, records of 1 or 24 h are definitely nonstationary due to long-range correlations. If we had, however, a very long HRV sequence, say of 50 days, there is a good chance of finding stationarity over longer scales too.

IV. SUMMARY AND DISCUSSION

In this article we have proposed to test stationarity of data series with a combination of a test for the time independence of the probability distribution and a test based on the time independence of the power spectra. By applying these tests to several types of time series their potentials and limits have been demonstrated.

Each data series that passes both tests can be regarded at least as weakly stationary, because the time independence of the first and all second moments is examined. The analysis of several examples demonstrates clearly the necessity of both tests for testing stationarity. Since the tests do not include the time independence of all central and noncentral moments, strong stationarity is not tested.

We have presented an example of a weakly but not strongly stationary process that has passed both tests. This refers to the fact that our hypothesis is the coincidence of the 1D probability distribution and the spectral distribution, which both demand less restrictive structural properties than strong stationarity does. Thus, a limit of the procedure proposed is obvious. Another limit is that the detection of nonstationarity requires an expressed variation of the structure, as discussed for some autoregressive processes with varying coefficients (cf. Sec. III B).

It has been shown that the methods cannot be used as black-box algorithms: In particular, the window length must be in accordance with the correlation length, i.e., in the case of data with strong correlations the window length has to be chosen in such a manner that the long-frequency shares are sufficiently represented. On the contrary, it might happen to be correct that small parts of a stationary series cannot be considered as stationary.

In the case of deterministic systems it is recommended to use only the test for the coincidence of the spectral distribution. The other test statistic may be strongly influenced by the deterministic nonlinearity [28].

If the data series is a concatenation of different stationary subsequences, these could be found again by mutual comparison of the probability distribution and the spectral densities of the windows. So a regime of transient behavior has been found (cf. Sec. III A 3). In this way it is already possible to detect short interruptions in the structure of the data, e.g., if the heart rhythm during the sleeping phase is shortly changed by movements.

Some important consequences for data analysis are as follows: (i) A lot of data-analytical methods assume stationarity. If this prerequisite is not checked disastrous artifacts may occur. (ii) Methods that assume stationarity can be applied only if the main time scales of the process considered are small in comparison with the observational length. (iii) Further, the test can be used for picking out structures as Thiessenhusen *et al.* [29] have done for Saturn's rings.

We have applied the proposed procedure to geophysical observations as well as to physiological data. In both cases we have found stationary subseries. Both types of data need filtering due to long-range influences of the reversals of the Earth's magnetic field and temperature-regulating processes. If we find stationarity for filtered data this means that the dynamics of the series that have passed the filter (bands with higher frequencies) are time independent.

If stationarity is checked for deciding whether an attractor dimension estimation is possible we recommend the method of Schreiber [5], which is especially constructed for this problem and possesses a more powerful approach for solving it.

In this article only 1D observations of systems are considered. It is necessary to extend these tests to multivariate time series. Moreover, it should be emphasized that we are not able to give a strong mathematical proof about the correctness of the modifications Eq. (8) and Eq. (13) of the χ^2 test procedures, but the results of the simulations support our considerations.

ACKNOWLEDGMENTS

The data series of heart rate variability have been put at our disposal by A. Voss (Max-Delbrück-Center, Berlin Buch). For helpful hints and explanations we are indebted to our colleagues U. Feudel (strange nonchaotic attractors), F. Feudel (Kuramoto-Sivashinsky equation), and U. Schwarz (geophysical data). We thank D. Kaplan, H. R. Künsch, M. Muldoon, L. Smith, J. Timmer, and H. Voss for detailed discussions.

APPENDIX

This section deals with an algorithm that gives a possible window length for an arbitrary time series. It is obvious that for characterizing the first and second moments of a time series with a large correlation time a longer realization is needed than for an uncorrelated process. The minimal length of that realization can be determined in the following way:

The information about the (linear) memory range of the time series under consideration is expressed in the autocorrelation function ρ. The depth of the memory is given by the largest value of the lag k for that $\rho(k)$ is not vanishing. Due to the finite length of the data series the autocorrelation function $\rho(l)$ for $l>k$ reaches instead of the exact value ''0'' which has to be determined firstly. We introduce this ''numerical 0,'' as the significance level $[f_{\rm SL}=f_{\rm SL}(n,n_\rho)]$ depending on the sequence length n and the length of the autocorrelation function n_ρ. For an uncorrelated time series with the same probability distribution as the original data series we define $f_{\rm SL}$ by the equation

$$p\left(\max_{0<|i|\leq n_\rho}[|\mathrm{acf(i)}|]<f_{\rm SL}(n,n_\rho)\right)=0.95. \qquad \text{(A1)}$$

Since this definition requires that *each* element of ρ is with 95% probability smaller than $f_{\rm SL}$, $f_{\rm SL}$ is bounded from below by confidence limit $1.96/\sqrt{n}$ [6] for a single element of ρ.

We have estimated $f_{\rm SL}$ in the case of Gaussian white noise via Monte Carlo experiments and found dependences on n and n_ρ that show the following scaling laws: For constant time series length n we got

$$f_{\rm SL}(n_\rho)=a+b\ \ln(n_\rho) \qquad \text{(A2)}$$

and for constant n_ρ the relation

$$f_{\rm SL}(n)=c_0 n^c \qquad \text{(A3)}$$

is fulfilled. For varying n and n_ρ the ansatz

$$f_{\rm SL}(n,n_\rho)=\frac{a(n)+b(n)\ln(n_\rho)}{n^c} \qquad \text{(A4)}$$

is used, whereby

$$a(n)=a_0+a_1\ \ln(n), \qquad \text{(A5)}$$

$$b(n)=b_0 n^{b_1} \qquad \text{(A6)}$$

and the free parameters are estimated by $a_0=1.25$, $a_1=-0.078$, $b_0=0.340$, $b_1=-0.157$, $c=0.36$. In Fig. 8 this

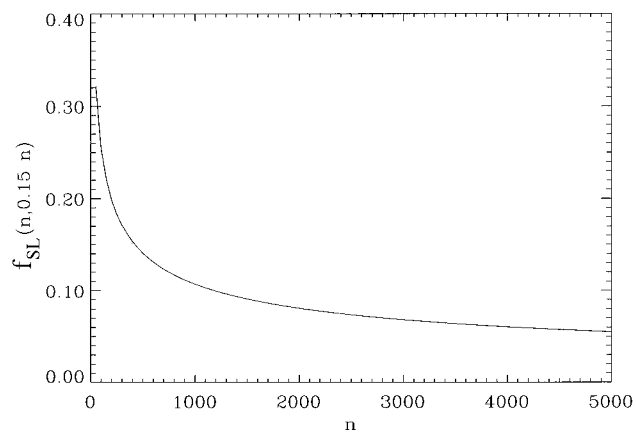

FIG. 8. Level of significance $f_{\rm SL}(n,n_\rho)$ over sequence length n whereby the length of the autocorrelation function is chosen as $n_\rho=0.15n$.

significance level is shown in dependence of the time series length n where the length of the autocorrelation function satisfies $n_\rho=0.15n$.

For applying this level of significance to a time series with a non-Gaussian distribution the time series has to be transformed into a Gaussian-distributed one using the filter of Kaplan [30].

Coming now to the estimation of the minimal window length, we propose to determine it as follows: (1) Transformation of the data series $\{x_t\}$ into a Gaussian distributed one $\{\tilde{x}_t\}$; (2) calculation of the autocorrelation function $\tilde{\rho}$ for $\{\tilde{x}_t\}$; (3) calculation of $f_{\rm SL}$ in according to Eq. (32); (4) determination of the length $n_{\rm ess}$ (memory depth) of the essential part of $\tilde{\rho}$:

$$n_{ess}=\max_{0\leq i<n_\rho}[\rho(i)\geq f_{\rm SL}(n,n_\rho)]. \qquad \text{(A7)}$$

(5) We made good experiences with a minimal window length of $n_w=7n_{\rm ess}$.

If this minimal window length n_w is larger than half of the sequence length, either the data series has to be filtered or the application of the test for stationarity is impossible.

[1] H. D. I. Abarbanel, R. Brown, J. L. Sidorowich, and L.Sh. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).
[2] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997)
[3] R. Manuca and R. Savit, Physica D **99**, 134 (1996).
[4] M. Koebbe and G. Mayer-Kress, in *Nonlinear Modelling and Forecasting, SFI Studies in the Science of Complexity, Proc. Vol. XII*, edited by M. Casdagli and S. Eubank (Addison-Wesley, New York, 1992), pp. 361–378.
[5] T. Schreiber, Phys. Rev. Lett. **78**, 843 (1997).
[6] M. B. Priestley, *Spectral Analysis and Time Series* (Academic Press, London, 1981).
[7] M. B. Priestley, *Non-linear and Non-stationary Time Series Analysis* (Academic Press & Harcourt Brace Jovanovich Publishers, London, 1988).
[8] D. Ruelle, *Chaotic Evolution and Strange Attractors* (Cambridge University Press, Cambridge, 1987).
[9] J. P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[10] H. Isliker and J. Kurths, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 157 (1993).
[11] W. H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C. The Art of Scientific Computating* (Cambridge University Press, Cambridge, 1993).
[12] W. Ebeling and T. Pöschel, Europhys. Lett. **26**, 241 (1994).
[13] J. Feder, *Fractals* (Plenum Press, New York, 1988).
[14] O. R. Osborne and A. Provenzale, Physica D **35**, 357 (1989).
[15] J. Kurths, U. Schwarz, and A. Witt, in *Coronal Magnetic Energy Releases*, Proceedings of the CESRA Workshop, Caputh/Potsdam, Germany, May 16–20, 1994, edited by A. O. Benz and A. Krüger (Springer-Verlag, Berlin, 1995), p. 159.

[16] S. Grossmann and S. Thomae, Z. Naturforsch. A **32**, 1353 (1977).
[17] C. Grebogi, E. Ott, S. Pelikan, and J. A. Yorke, Physica D **13**, 261 (1984).
[18] A. S. Pikovsky and U. Feudel, J. Phys. A **27**, 5209 (1994).
[19] F. Feudel, U. Feudel, and A. Brandenburg, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 1299 (1993).
[20] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Physica D **58**, 77 (1992).
[21] J. Kurths and U. Schwarz, Space Sci. Rev. **68**, 171 (1994).
[22] J. Kurths, U. Schwarz, C. P. Sonett, and U. Parlitz, Nonlinear Processes in Geophys. **1**, 72 (1994).
[23] H. Voss, J. Kurths, and U. Schwarz, J. Geophys. Res. **101**, 15 637 (1996).
[24] L. Glass, P. Hunter, and A. McCulloch, *Theory of Heart* (Springer, New York, 1989).
[25] A. Babloyanz, J. M. Salazar, and C. Nicolis, Phys. Lett. **111A**, 152 (1985).
[26] M. Kobayashi and T. Musha, IEEE Trans. Biomed. Eng. **29**, 456 (1982).
[27] J. T. Bigger, J. L. Fleiss, L. M. Rolnitzky, and R. C. Steinmann, Am. J. Cardiol. **69**, 718 (1992).
[28] A. Provenzale, L. A. Smith, R. Vio, and G. Murante, Physica D **58**, 31 (1992).
[29] K. U. Thiessenhusen, L. W. Esposito, J. Kurths, and F. Spahn, Icarus **113**, 206 (1995).
[30] D. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992).