Automatic algorithm for monotone trend removal

Călin Vamoş

"T. Popoviciu" Institute of Numerical Analysis, Romanian Academy, P.O. Box 68, 400110 Cluj-Napoca, Romania
(Received 6 October 2005; revised manuscript received 29 November 2006; published 16 March 2007)

The available numerical algorithms for trend removal require a direct subjective intervention in choosing critical parameters. In this paper an algorithm is presented, which needs no initial subjective assumptions. Monotone trends are approximated by piecewise linear curves obtained by dividing into subintervals the signal values interval, not the time interval. The slope of each linear segment of the estimated trend is proportional to the average one-step displacement of the signal values included into the corresponding subinterval. The evaluation of the trend removal is performed on statistical ensembles of artificial time series with the random component given by realizations of autoregressive of order one stochastic processes or by fractional Brownian motions. The accuracy of the algorithm is compared with that of two well-tested methods: polynomial fitting and a nonparametric method based on moving average. For stationary noise the results of the algorithm are slightly better, but for nonstationary noise the preliminary results indicate that the polynomial fitting has the best accuracy. As a verification on a real time series, the time periods with monotone variation of global average temperature over the last 1800 years are established. The removal of a nonmonotone trend is also briefly discussed.

I. INTRODUCTION

In many practical situations processing experimental observations of superposed phenomena having different time scales is needed. Statistical methods can be used to separate these components from the global signal. For example, trend removal is performed by different methods from the simpler ones (least-squares fit of a parametric family of functions or smoothing by moving average) to more complex as discrete differencing [1] or wavelet analysis [2]. More sophisticated methods should be used only if their results are significantly better than the results of the simpler ones. A common drawback of the available methods is a direct subjective intervention in choosing critical characteristics such as functional form of the trend, number of averagings or differentiations of the signal, length of the averaging interval, type of the basis functions, etc.

In this paper a trend removal algorithm, which needs no initial assumptions, is proposed. It is based on a statistical method, which approximates the trend by a piecewise linear curve obtained by dividing into subintervals the signal values interval, not the time interval. The slope of each linear segment of the estimated trend is proportional to the average one-step displacement of the signal values included into the corresponding subinterval, therefore the method is referred to as average conditional displacement (ACD). Here only the case of the monotone trend is considered. The extension of the ACD method to a nonmonotone trend is discussed in the last section and will be presented in a separate future paper.

In practice there are situations when the trend monotony is specially required, for example, ascertainment of the time periods with monotone variation of the global atmospheric temperature, problem discussed in Sec. V. But even when the trend monotony is not significant, the initial separation of a monotone component from the global nonmonotone trend may be useful. Even if initially there is no indication that a persistent phenomenon has a contribution to the generation of the time series, however, rendering evident a significant monotone component is a reason to test the existence of such a phenomenon. Obviously, the main interest in real applications is to estimate the whole trend, including the turning points.

As a rule, the trend removal methods do not take into account whether the estimated trend is or is not monotone. The exceptions are the methods in which the trend is explicitly looked for as a monotone function, usually a linear, exponential, or logarithmic function. The weakness of this approach consists in the limited number of available monotone functional forms. If an enrichment of the functional forms is attempted, for example, using polynomials of order greater than one, then the monotony property is lost. From this point of view, the advantage of the ACD method is that it can describe a much richer set of monotone trends as piecewise linear functions. Furthermore, if the ACD method does not succeed to remove a monotone trend, this is useful information to characterize the time series properties because it shows that no monotone component could be associated to that series.

Besides the request that the estimated trend should be monotone, another essential feature of the ACD method is its capacity to support the development of an automatic algorithm. The automatic choice of the values of the ACD parameters is performed by means of a numerical estimation of the standard deviation of the signal random component. The quality of the estimated standard deviation and the accuracy of the automatic ACD algorithm are tested on artificial signals with stationary noises of type AR(1) (autoregressive of order one).

Even without a thorough analysis, the ACD method is also applied on signals with nonstationary noise. Such signals occur in detrended fluctuation analysis (DFA) [3], extensively used for the detection of long-range correlations in time series with deterministic trend and stationary $1/f$ noise. A detailed analysis of DFA and a review of its applications are presented in [4]. The initial signal is summed up, result-

ing in a secondary one with nonstationary noise. An essential step in DFA is the repeated removal of polynomial trends by means of the least-squares method from the secondary signal and DFA could be significantly improved if the trends were removed by means of an automatic algorithm like the extension of the ACD method for the nonmonotone trend.

Since ACD is a method which does not develop any of the trend removal methods used at present, there is no obvious term of comparison for its performance evaluation. Therefore I have chosen two reference methods based on secondary criteria. The first one is the polynomial fitting with different degrees used in DFA. The second one is "the jump process" presented in [5] which has some similarities with the method used in the ACD algorithm to dampen the fluctuations. Thus I have chosen for comparison one representative from each of the two major types of trend removal methods (parametric and nonparametric) extensively reviewed in [5]. The performance of the ACD algorithm is slightly better than that of the reference methods, but its main quality is the possibility to apply it automatically, without any subjective intervention.

In the next section I describe the ACD algorithm. Then, in Sec. III I present the types of artificial time series used to analyze the ACD algorithm and an estimation of the magnitude order of the standard deviation of a stationary noise superposed on a monotone trend. Section IV contains the analysis of the ACD parameters and the description of the automatic algorithm associated with the ACD method. The performance comparison of the ACD algorithm with the polynomial fit and the jump process is made in Sec. V, where I also analyze the variation of the global temperature anomalies over the last 1800 years. In Sec. VI nonmonotone trends and short time series are analyzed. In the last section I summarize the results and briefly discuss the extension of the ACD method to a nonmonotone trend.

II. ACD METHOD

Consider a time series $\{x_n\}$, with $1 \leqslant n \leqslant N$, generated by the stochastic process

$$X_n = f_n + Z_n, \tag{1}$$

where $\{Z_n\}$ is a discrete stochastic process with $\langle Z_n \rangle = 0$ and f_n are the values of the trend $f(t)$ at the moments $t_n = (n-1)\delta t$, δt being the sampling interval. The time takes values within the finite interval $t \in [0,1]$. We assume that Z_n does not depend on the trend values f_n and $f(t)$ has a slow and monotone variation and a vanishing temporal average $\bar{f} = N^{-1}\Sigma_{n=1}^{N} f_n = 0$.

If the stochastic process $\{Z_n\}$ is stationary, then a mathematical justification of the ACD method can be given. We denote by $p_z(z)$ the probability distribution of Z_n and by $p_x(x,n)$ that of X_n, which depends explicitly on the time index n because $\{X_n\}$ is a nonstationary process. According to Eq. (1), the two distributions are identical with the exception of a translation by f_n, i.e., $p_x(x,n) = p_z(x-f_n)$. The infinitesimal interval about a given real number ξ is denoted by $I_\xi = (\xi - \delta\xi/2, \xi + \delta\xi/2)$. The average number of values of the time series $\{x_n\}$ lying within I_ξ is given by $\langle N_\xi \rangle \delta\xi$, with

$$\langle N_\xi \rangle = \sum_{n=1}^{N} p_z(\xi - f_n). \tag{2}$$

We define the average one-step displacement with the initial value in the neighborhood of ξ

$$g(\xi)\delta\xi = \frac{1}{\langle N_\xi \rangle} \sum_{n=1}^{N-1} p_z(\xi - f_n)\langle \delta X_n | X_n \in I_\xi \rangle, \tag{3}$$

where $\delta X_n = X_{n+1} - X_n$. Using the conditional probability density for successive values of the stationary stochastic process $\{Z_n\}$ denoted by $p_z(z''|z')$, we can write

$$\langle \delta X_n | X_n \in I_\xi \rangle = \delta\xi \int_{-\infty}^{+\infty} (x - \xi)p_z(x - f_{n+1}|\xi - f_n)dx,$$

where relation $p_x(x, n+1 | \xi, n) = p_z(x - f_{n+1} | \xi - f_n)$ is used. From the simple change of variables $z = x - f_{n+1}$ and from the definition of the conditional probability and the consistency condition for the joint probability density $p_z(z'', z')$

$$p_z(\xi - f_n) = \int_{-\infty}^{+\infty} p_z(z, \xi - f_n)dz,$$

it follows that Eq. (3) becomes

$$g(\xi) = \frac{1}{\langle N_\xi \rangle} \sum_{n=1}^{N-1} (f_{n+1} - f_n)p_z(\xi - f_n) + \langle \varepsilon_\xi \rangle. \tag{4}$$

The first term is the average of the one-step variation of the trend within the neighborhood of ξ. A similar relation is obtained if the final value is included in I_ξ.

The second term in Eq. (4)

$$\langle \varepsilon_\xi \rangle = \frac{1}{\langle N_\xi \rangle} \sum_{n=1}^{N-1} \int_{-\infty}^{+\infty} (z + f_n - \xi)p_z(z, \xi - f_n)dz \tag{5}$$

measures the difference between $g(\xi)$ and the trend average slope. Let us investigate when this term vanishes. If the trend variation at one time step is much smaller than that of the noise $[|f'(t_n)\delta t| \ll \sigma_Z$ for all $n]$, then the sum can be approximated by an integral. If, in addition, the trend is linear $f(t) = at + b$ and we make the change of variable $\theta = \xi - f(t)$, then we obtain

$$\langle \varepsilon_\xi \rangle = \frac{a}{\langle N_\xi \rangle \delta t} \int_{\xi - f_1}^{\xi - f_N} d\theta \int_{-\infty}^{+\infty} (z - \theta)p_z(z, \theta)dz.$$

Consider that the noise Z_n has the symmetry property $p_z(z) = p_z(-z)$ and $p_z(z', z'') = p_z(-z', -z'')$. It follows that for $\xi = (f_1 + f_N)/2$ the term $\langle \varepsilon_\xi \rangle$ vanishes and its value increases when ξ approaches the extreme values of the time series. If the noise is moderately asymmetric, then $\langle \varepsilon_\xi \rangle$ vanishes for a different ξ but close to $(f_1 + f_N)/2$.

Under the conditions of the previous paragraph, since $f_{n+1} - f_n = a\delta t$, from Eqs. (4) and (2) it follows that $g(\xi) = a\delta t$, i.e., we find the exact slope of the linear trend. In the general case of nonlinear trends, $g(\xi)/\delta t$ is only an approximation of the trend slope $f'(t)$. The trend $F(t)$ estimated by the ACD method is calculated from the requirement that its

derivative expressed with respect to the function values (not to the argument) should be proportional to the average one-step displacement

$$g(\xi) = F'[F^{-1}(\xi)]\delta t. \qquad (6)$$

This relation holds only if $F(t)$ is invertible, i.e., if $g(\xi)$ preserves the same sign over all its domain of definition.

Our goal is to find the numerical quantity corresponding to the theoretical one defined in Eq. (3) using the values of a time series $\{x_n\}$ obtained as a realization of the stochastic process $\{X_n\}$. The infinitesimal intervals I_ξ must be replaced with finite ones. We divide the domain of the time series values into disjoint intervals $I_s = (\xi_s, \xi_{s+1}], s = 1, 2, \ldots, S$, so that any value x_n is contained into an interval I_s. Denote by N_s the number of values x_n lying within I_s corresponding to the theoretical quantity $\langle N_\xi \rangle$ defined in Eq. (2). The intervals I_s can be chosen in many ways. The simplest solution is to use homogeneous intervals, i.e., the values of N_s should differ from each other by a unit at the most. Then the total number of intervals S is the single parameter describing the distribution of the series values. For the final form of the ACD algorithm we use a distribution into nonhomogeneous intervals described in Appendix B.

The one-step variation of the time series is $\delta x_n = x_{n+1} - x_n$. For a given s, we compute the sample average of δx_n under the condition that the initial or final values should be included into the interval I_s

$$\hat{g}_s = \frac{1}{2N_s}\left(\sum_{x_n \in I_s} \delta x_n + \sum_{x_{n+1} \in I_s} \delta x_n\right), \qquad (7)$$

which corresponds to the definition in Eq. (3). If all the values $\hat{g}_s$ have the same sign, then we can use them to determine a numerical approximation of the trend by a piecewise linear curve denoted $\hat{F}(t)$ and corresponding to the theoretical one in Eq. (6). Since the initial f_1 and the final f_N values of the trend are unknown, we use instead the extreme values of the time series $\{x_n\}$.

The domain of definition $t \in [0, T]$ of the function $\hat{F}(t)$ is different from that of the real trend $f(t)$. Therefore the numerical values of the estimated trend are obtained either by translation $\hat{F}_n = \hat{F}[(\kappa+n)\delta t]$, or by scaling $\hat{F}_n = \hat{F}(nT/N)$. The optimal estimated trend is given by the requirement that the difference $\{x_n - \hat{F}_n\}$ should have a minimum standard deviation.

If the fluctuations described by ε_ξ causes the quantities $\hat{g}_s$ to have different signs, then the estimated trend $\hat{F}_n$ cannot be determined and the fluctuations are smoothed by means of a moving average. Therefore the numerical algorithm of trend removal using the ACD method consists of a succession of trend extractions and moving average smoothings. Denote by $\{x_n^{(i)}\}$ the time series obtained after a succession of i extractions and smoothings. Initially $x_n^{(0)} = x_n$. First, at each step i, we try to remove an estimated monotone trend $\hat{F}_n^{(i)}$

$$x_n^{(i+1)} = x_n^{(i)} - \hat{F}_n^{(i)}. \qquad (8)$$

If $\hat{g}_s$ does not have the same sign ($\hat{F}_n^{(i)} = 0$ for all n), then $\{x_n^{(i+1)}\}$ is computed by moving averaging. After i steps, the estimated trend is the sum of the removed components

$$\hat{F}_n = \sum_i \hat{F}_n^{(i)} \qquad (9)$$

and the estimated random component is

$$\hat{z}_n = x_n - \hat{F}_n. \qquad (10)$$

The time series processing is interrupted when the standard deviation of the residual $\{x_n^{(i)}\}$ is ρ times smaller than that of the initial series (ρ is a given positive real number), or if the standard deviation of $\{x_n^{(i+1)}\}$ is larger than that of the previous step $\{x_n^{(i)}\}$, or if by adding the component $\{\hat{F}_n^{(i+1)}\}$ the estimated trend $\{\hat{F}_n\}$ is not monotone.

For the interior values two-sided moving averages are performed over an interval of length $2K+1$. If $n \leq K$ ($n > N - K$), then the average is taken over the first $n+K$ (the last $N-n+K+1$) values. This asymmetric average forces the values near the time series boundaries to follow the variations of the interior values. The initial averages are performed on small intervals such that the trend should be deformed as little as possible. For the first smoothing we use $K=1$. If the quantities $\hat{g}_s$ do not acquire the same sign, then K is gradually increased by a unit for each new smoothing up to a maximum value K_f and then the next smoothings are computed keeping the same value for K. In this way a compromise is made between the computing efficiency and the requirement that the trend should not be distorted when the noise is small.

A numerical problem of the ACD algorithm described above is the possibility that some values $\hat{g}_s^{(i)}$ could be very close to zero and then the segments of the estimated trend in the corresponding intervals I_s would be almost parallel to the time axis. In such a case the estimated trend would be artificially deformed and its length would be much longer than the length of the initial series ($T \gg 1$). In order to eliminate this possibility we impose the additional condition that the absolute value of the slope $\hat{g}_s^{(i)}$ should have an inferior bound. Denote by $n_s^{\min}$ ($n_s^{\max}$) the time step when the signal takes for the first time (the last time) a value in the interval I_s. The difference $\Delta n_s = n_s^{\min} - n_s^{\max}$ is the number of time steps during which the time series takes all the values in I_s. Then we choose the minimum value of the estimated trend slope as $|\hat{g}_s^{(i)}| \geq (\xi_{s+1} - \xi_s)/\Delta n_s$.

Since the ACD algorithm is rather complex, it is affected by many error sources. First, there is the theoretical error given by Eq. (5) which is larger when the trend nonlinearity is greater, the time step is larger, and the coordinate ξ is closer to the extreme values of the time series. Then, the numerical implementation induces other errors such as the replacement of the infinitesimal intervals I_ξ with the finite ones I_s, the use of the extreme values of the time series instead of those of the real trend, and the trend distortion by moving average. In spite of the error diversity, the evaluation
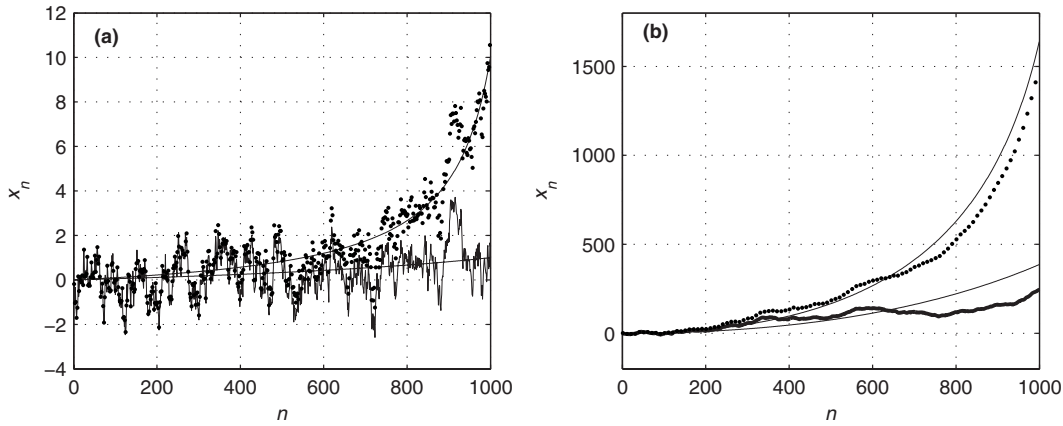
FIG. 1. Examples of signals used to analyze the performance of the ACD method. The noise is superposed on the trend $f(t)=t/(a-t)$, $t\in[0,1]$ represented with a thin line. The values of the trend parameter are $a=2$ (thick line) and $a=1.1$ (point markers). (a) Stationary AR(1) noise with $\phi=0.9$ and $\sigma_Z=1$. (b) Nonstationary noise given by a realization of a fractional Brownian motion obtained by summing up a $1/f^\gamma$ noise with $\gamma=0.4$.

presented in Secs. V and VI shows that the final error of the ACD algorithm is not larger than the errors of the usual trend estimation methods. This result could be explained by the fact that the errors of the successive estimations of the trend components $\hat{F}_n^{(i)}$ compensate each other.

III. ARTIFICIAL TIME SERIES

The ACD algorithm performance is analyzed using statistical ensembles of time series with as various as possible characteristics. We generate the time series according to Eq. (1) superposing realizations of a given stochastic process $\{Z_n\}$ on a monotone trend. Since in Sec. V the ACD algorithm is compared with the polynomial fit, we do not analyze the algorithm for a polynomial trend, but for $f(t)=t/(a-t)$, $t\in[0,1]$, with $a>1$. If the trend were of the same functional form as the function used in the least-squares-fit method, then the ACD method would be disadvantaged. We choose this functional form because its slope $f'(t)=a/(a-t)^2$ has a nonhomogeneous distribution with its extreme values near the boundaries of the definition domain: the minimum at $t=0$ and the maximum at $t=1$. It is more difficult to remove such a trend than one with the maximum slope in the interior of the definition domain because numerical algorithms are more difficult to implement near the boundaries of the time series. In our case the ratio between the maximum and the minimum slope is $f'(1)/f'(0)=(1-1/a)^{-2}$. We give to the parameter a values within the interval $a\in[1.1,2.0]$, so that the ratio of the extreme values of the slope varies between 121 and 4. The resolution of the time series is varied with one order of magnitude choosing the number of values N within the interval $N\in[100,1000]$.

The majority of the numerical tests are performed on signals with a stationary random component $\{z_n\}$ so that they obey the theoretical considerations in Sec. II. We numerically generate realizations of a stationary stochastic process $\{Z_n\}$ of type AR(1) (autoregressive of order one) defined by $Z_n=\phi Z_{n-1}+G_n$, where $0\leqslant\phi<1$ and $\{G_n\}$ is a Gaussian noise with null mean and standard deviation σ_G. The properties of

the AR(1) stochastic process are well-known [6]. Its probability distribution is Gaussian with null mean and standard deviation $\sigma_Z^2=\sigma_G^2(1-\phi^2)^{-1}$. When the value of the parameter ϕ increases, the successive values of the noise become more correlated. In the numerical tests ϕ takes values within the interval $\phi\in[0,0.99]$, white noise being obtained for $\phi=0$. By means of the noise standard deviation σ_Z we control the ratio between the variation due to the noise and that due to the trend. In order to generate signals of both types, the values of σ_Z are chosen within the interval $\sigma_Z\in[0.1,10]$. Figure 1(a) contains two signals with AR(1) noise, one dominated by noise, the other by trend.

Taking into account the intention that the extension of the ACD method for a nonmonotone trend should be used in DFA, I have chosen for preliminary tests a $1/f$ noise, i.e., $\langle Z_n Z_{n+h}\rangle\sim h^\gamma$, where $0<\gamma<1$. Such stationary time series with unit standard deviation are generated applying the numerical method in [7] for a spatially correlated noise on a one-dimensional time lattice. An important step in DFA is the polynomial trend removal from a series obtained by summing up the initial one $y_n=\sum_{k=1}^n x_k$, which is a nonstationary fractional Brownian motion. Figure 1(b) shows two signals $\{y_n\}$ obtained by summing up a time series $\{x_n\}$ with a $1/f^\gamma$ noise with $\gamma=0.4$. Although the theory presented in Sec. II does not hold for nonstationary noise, we remove the trend replacing in Eqs. (7)–(10) x_n and z_n with y_n and $\sum_{k=1}^n z_k$, respectively.

In time series processing a critical parameter is the ratio of the overall trend variation to the amplitude of noise fluctuations quantitatively expressed as

$$\eta_0 = \sigma_f/\sigma_Z, \qquad (11)$$

where σ_f is the trend standard deviation and σ_Z is the noise standard deviation. Figure 2 shows the dependence of η_0 on the noise standard deviation σ_Z for the extreme values of the trend parameter a. The global variation of η_0 ranges over three magnitude orders providing a sufficient diversity of the time series characteristics on which the ACD algorithm is tested. The signals in Fig. 1(a) correspond to the two possible
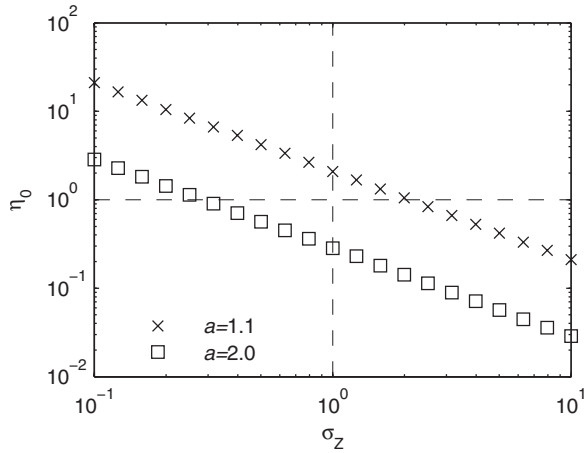
FIG. 2. Average ratio of the trend standard deviation to the noise standard deviation in terms of standard deviation of the noise for the maximum value ($a=1.1$) and the minimum one ($a=2.0$) of the trend parameter for time series with AR(1) noise.
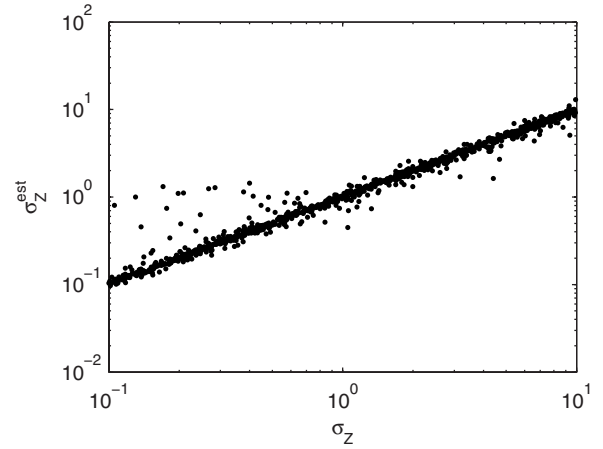

FIG. 3. Correlation between the standard deviation of an AR(1) noise superposed on a monotone trend and the estimated standard deviation of the noise.

cases: $\eta_0 > 1$ for $a=1.1$ when the signal is dominated by trend and $\eta_0 < 1$ for $a=2.0$ when the signal is dominated by noise. Thus a time series with AR(1) noise is characterized by four parameters: the series length N (related to the resolution of the series), the trend parameter a, the parameter ϕ describing the correlation between the successive terms of the noise, and the noise standard deviation σ_Z.

In building an automatic algorithm for time series processing it is useful to have a simple criterion to estimate the order of magnitude of the noise standard deviation σ_Z using only the values of the original series $\{x_n\}$. The random component $\{z_n\}$ is actually one of the unknown quantities which has to be determined by numerical processing. Denote by $\delta_m x_n = x_{m+n} - x_n$ the variation of order $m < N$ of the time series $\{x_n\}$. According to the justification given in Appendix A, a workable estimation of the magnitude order of σ_Z is given by

$$\sigma_Z^{\text{est}} = \frac{1}{\sqrt{2}}\hat{\sigma}(\delta_{m_0} x_n), \tag{12}$$

where $\hat{\sigma}(\cdot)$ is the sample standard deviation and m_0 is the smallest integer number $m < N/2$ for which

$$\|\delta_{m_0+1} x_n\| < \|\delta_{m_0} x_n\|, \tag{13}$$

where $\|\cdot\|$ is the usual quadratic norm. I have tested the accuracy of the estimation (12) on 1000 time series with AR(1) noise, each of them with parameters N, a, ϕ, and σ_Z randomly chosen within their maximum range of variation. In Fig. 3 a remarkable correlation between the actual value of σ_Z and the estimated one from Eq. (12) is noticeable. Only for $\sigma_Z < 1$ there are some estimated values almost one order of magnitude greater than the real value. A percentage of 2.8% time series for which there is no value m_0 satisfying Eq. (13) is missing from the figure. Such situations occur when the noise standard deviation is very small (see Appendix A) and then we consider $\sigma_Z^{\text{est}} = 0$. In the next section we describe a method to obtain a nonvanishing estimation even in these cases. In Appendix A we also prove that estimation

(12) is meaningful even if the trend is not monotone, such that it can be applied without additional restrictions on an arbitrary signal.

IV. ACD PARAMETERS

In order to apply the ACD method there are two things to be done: (1) establish the values of two parameters (the maximum value ρ of the ratio between the initial signal standard deviation and the final residual standard deviation and the maximum value K_f of the semilength of the averaging interval); and (2) distribute the signal values into disjoint intervals I_s. In this section we show how these tasks can be automatically accomplished.

First we analyze the influence of the parameter ρ on the ACD performance. Since no criterion to choose the values of the other parameter N_f has been established, its values are chosen randomly according to a homogeneous probability distribution over its variation range. The parameter N_f controls the speed of the damping of the signal fluctuations by the moving average. Its value is calculated from the value of the semilength K_f of the averaging interval ($N_f = 2K_f + 1$) which takes values in the interval $K_f/N \in [0.01, 0.1]$. For the shortest signals ($N=100$) the minimum value of K_f is $K_f = 1$.

In this stage of the analysis we use homogeneous intervals I_s, such that they are completely described by a single parameter, i.e., the total number of intervals S. The minimum value of this parameter is $S_{\min} = 2$, and the maximum one $S_{\max} = [N/N_{\min}]$, where $[\cdot]$ is the integer part function and $N_{\min} = 14$ is the minimum number of the signal values necessary to recognize a Gaussian white noise, which is determined in Appendix A. In the end we shall introduce a nonhomogeneous distribution of the data values.

The optimum value of the parameter ρ is found by means of the estimation (12) of the noise standard deviation. It is known that the sample mean of a random variable with a Gaussian distribution has the standard deviation $\sqrt{N_r}$ times smaller than that of the random variable, where N_r is the number of realizations. Considering that the reduction by
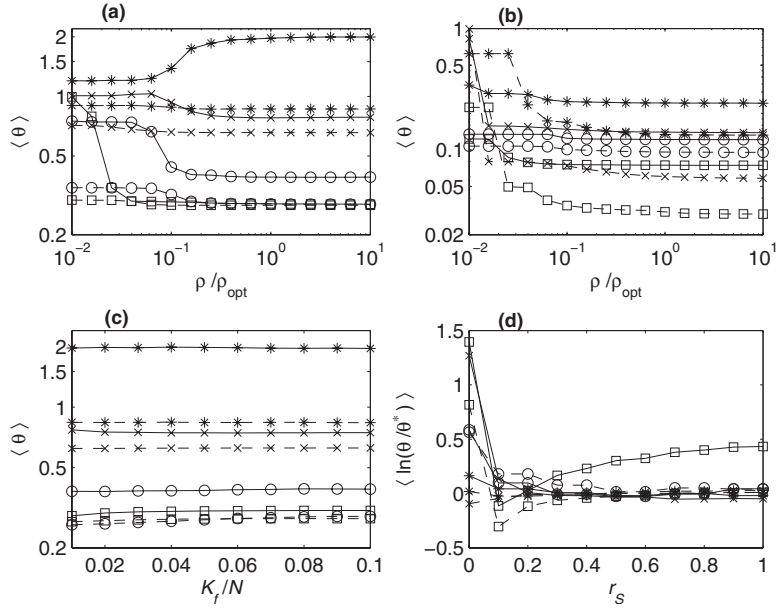
FIG. 4. Dependence of the average evaluation index of the ACD algorithm on the ACD parameters. The time series are characterized by different values of their length N and of the parameter a in the trend formula $f(t)=t/(a-t)$, $t \in [0,1]$. The couple (N,a) takes the values $(100,2.0)$ (asterisk marker $*$), $(100,1.1)$ (cross marker $\times$), $(1000,2.0)$ (circle marker $\bigcirc$), and $(1000,1.1)$ (square marker $\square$). In panels (a), (c), and (d) the random component of the signals is a stationary AR(1) noise with $\sigma_Z=1$ and $\phi=0.9$ (continuous lines) and a nonstationary fractional Brownian motion with $\gamma=1.4$ (dashed lines). In panel (b) the random component is a stationary AR(1) noise with $\sigma_Z=0.1$ and $\phi=0.9$ (continuous lines), and $\phi=0$ (dashed lines). (a) and (b) The variable parameter is the maximum value ρ of the ratio between the standard deviation of the initial signal and that of the final residual, while the other two ACD parameters have random values. A log-log scale is used. (c) The variable parameter is the maximum value K_f of the semilength of the averaging interval while $\rho=\rho_{opt}$ and S takes random values. A log-linear scale is used. (d) The evaluation index θ for S homogeneous intervals divided to the evaluation index θ^* obtained using nonhomogeneous intervals in terms of the ratio $r_S=(S-S_{min})/(S_{max}-S_{min})$ while $\rho=\rho_{opt}$ and $K_f=0.1N$. A standard error for $\langle\theta\rangle$ smaller than 5% is obtained with statistical ensembles containing 500 (a), 3000 (b), and 300 [(c) and (d)] time series.

$\sqrt{N_r}$ times of the standard deviation is a measure of the maximum effectiveness which a statistical method can have, we assume that in the case of the ACD method the standard deviation of the final residual can be at best $\sigma_Z/\sqrt{N}$ as well. Then we assume that the optimum value of ρ is

$$\rho_{opt} = \frac{\sqrt{N}\hat{\sigma}_x}{\sigma_Z^{est}}, \tag{14}$$

where $\hat{\sigma}_x$ is the sample standard deviation of the original signal. As noticed in the previous section, there are situations when Eq. (12) does not allow the determination of σ_Z^{est} and then ρ_{opt} cannot be calculated. Since such situations occur if σ_Z is small with respect to the variation due to the trend, then first the trend is removed from the time series $\{x_n\}$ and the estimation (12) is applied to the residual obtained. The trend removal is performed by means of the ACD algorithm with $S=S_{max}$ homogeneous intervals and $K_f=0.1N$. If the trend removal is not possible, then we apply Eq. (12) to the difference between the initial signal and its moving average performed with $K_f=0.01N$. In this way, we always obtain a value σ_Z^{est} by means of which ρ_{opt} in Eq. (14) can be calculated.

Figure 4 shows the dependence of the accuracy of the ACD method on the three ACD parameters. We evaluate the accuracy of the trend removal by means of the index

$$\theta = \|\hat{F}_n - f_n\|/\|f_n\|. \tag{15}$$

The smaller θ, the more alike the original and the estimated trends are. If the method ACD does not succeed to remove the trend ($\hat{F}_n=0$), then $\theta=1$. Therefore $\theta>1$ means that the estimated trend differs more from the real one than a vanishing trend and the estimated trend removal would cause a distortion of the information contained by the signal. In such cases the trend removal is not recommended [8] and in Sec. V we describe a method to recognize such situations. The results in Fig. 4 are obtained for statistical ensembles of time series with lengths $N=100$ and 1000 and with trend parameters $a=1.1$ and 2.0. The number of time series in a statistical ensemble has a value between 300 and 3000, such that for each type of time series the standard error for the average evaluation index $\langle\theta\rangle$ is smaller than 5%.

Figure 4(a) shows the average evaluation index $\langle\theta\rangle$ with respect to the parameter ρ for random values of the parameters K_f and S. The time series have the random component either stationary of type AR(1) with $\sigma_Z=1$ and $\phi=0.9$ (continuous lines) or a nonstationary fractional Brownian motion with $\gamma=1.4$ (dashed lines). For $\rho \geq \rho_{opt}$ the average index $\langle\theta\rangle$ reaches a stationary value in all cases, this kind of behavior showing that the value given by Eq. (14) assures the retrieval of all available information from the original signal. As a rule, for $\rho < \rho_{opt}$, $\langle\theta\rangle$ decreases for increasing ρ, showing the

importance of a right choice of ρ_{opt}, because if ρ has too small a value, then the trend is not completely removed. The only situation in which, for $\rho < \rho_{opt}$, $\langle \theta \rangle$ increases for increasing ρ is for time series with $N=100$ and $a=2.0$ (asterisk marker). For such time series, for greater ρ the trend removal leads to worse results, i.e., decreasing the standard deviation of the final residual, the estimated trend becomes more and more different from the real one. This behavior is due to the fact that when the signal is dominated by noise and its resolution is low, the estimated trend follows the noise fluctuations, not the trend. As seen in Fig. 4(a) (signals with the same characteristics as the ones previously discussed but with $N=1000$ and marked with open circles), increasing the signal resolution can render accessible the information on the trend. These results show that the optimum value of the parameter ρ is correctly given by Eq. (14).

Equation (14) implies a special procedure to calculate σ_Z^{est} when σ_Z has small values. The validity of this procedure is verified in Fig. 4(b) for time series with AR(1) noise with the minimum standard deviation $\sigma_Z=0.1$. In all cases the average evaluation index $\langle \theta \rangle$ reaches a stationary value for $\rho=\rho_{opt}$ after significant decreases. This behavior is in accordance with the fact that when the noise is small, the trend can be easily removed. Another observation is that the results depend on the parameter ϕ, i.e., on the correlation between successive values of noise. For $\sigma_Z > 1$ the results are similar with those previously discussed for signals dominated by noise and we do not present them here.

With the optimum value of the first parameter given by Eq. (14), we go on to analyze the parameter K_f controlling efficiency of the moving averaging in the ACD method while the data values are distributed into a random number S of homogeneous intervals. The larger K_f is, the longer the averaging intervals and the more strongly damped the noise fluctuations are, but at the same time the more distorted the trend is. Figure 4(c) contains the results for signals with AR(1) noise with $\sigma_Z=1$ and $\phi=0.9$ (continuous lines) and with nonstationary fractional Brownian motion with $\gamma=1.4$ (dashed lines). Over the entire variation range of K_f the average evaluation index $\langle \theta \rangle$ has approximately the same value. Also for other values of σ_Z and ϕ the behavior of $\langle \theta \rangle$ is similar, therefore we could choose for K_f an arbitrary value. However, the number of averagings decreases with almost two orders of magnitude when K_f varies and to save computing time we use its maximum value $K_f=0.1N$.

Finally we analyze the optimal distribution of the data values into intervals I_s, using for ρ and K_f the optimum values determined above. When we distribute the time series values into disjoint intervals I_s, we have to take into account two opposite requests. If the noise fluctuations are much smaller than the trend variation, then it is recommended to use a large number of intervals I_s in order to describe the trend shape as accurately as possible. On the contrary, if the noise fluctuations are much larger, then it is better to use a smaller number of intervals I_s, each of them containing more signal values so that the noise fluctuations may be smoothed as much as possible in the average slope given by Eq. (7). Therefore we distribute the signal values into a number of disjoint intervals inversely proportional with the estimated

standard deviation of the noise σ_Z^{est} given by Eq. (12),

$$S^{est} = (x_{max} - x_{min})/\sigma_Z^{est}, \qquad (16)$$

where x_{max} and x_{min} are, respectively, the maximum and the minimum value of the time series $\{x_n\}$. If the value obtained from Eq. (16) is smaller than S_{min} (larger than S_{max} $=[N/N_{min}]$), then we impose $S^{est}=S_{min}(S^{est}=S_{max})$. Since, as a rule, N is not exactly divisible by S^{est}, the distribution of the series values into S^{est} homogeneous intervals is performed so that the number of values N_s for different s may differ at most with a unit.

The distribution of the signal values into S^{est} homogeneous intervals must be corrected in order to take into account some numerical characteristics of the ACD algorithm. In Appendix B we describe an algorithm of splitting and merging the homogeneous intervals, in the end resulting S^* nonhomogeneous intervals. The ACD method with nonhomogeneous intervals cannot be evaluated as easy as that for homogeneous ones. The values of the series $\{x_n\}$ can be distributed into a given number of intervals in many ways. In an exhaustive analysis, the algorithm described in Appendix B should be compared with all the possible distributions of the signal values for all the allowed values of S. However, we are satisfied with an automatic algorithm having an accuracy comparable with that of the simplest nonautomatic form of the ACD method. Therefore we use as a reference in our comparisons the ACD method with homogeneous intervals of signal values. Figure 4(d) presents the average of the quantity $\ln(\theta/\theta^*)$, where θ is the evaluation index in Eq. (15) obtained for different numbers S of homogeneous intervals and θ^* is the evaluation index for S^* nonhomogeneous intervals obtained according to Appendix B. Since the maximum value $S_{max}=[N/N_{min}]$ of the parameter S depends on the signal length N, in Fig. 4(d) the ratio $r_S=(S-S_{min})/(S_{max}-S_{min})$ is used as an independent variable. The result of a trend removal is better when the evaluation index θ is smaller and then the positive values of the quantity $\ln(\theta/\theta^*)$ indicate that the ACD algorithm with S^* nonhomogeneous intervals gives a more correct estimated trend than with homogeneous intervals.

In Fig. 4(d) one notices that in most of the presented cases the accuracy of the two forms of the ACD method is approximately equal and in the rest of the cases the accuracy obtained with nonhomogeneous intervals is better. Only in a single case (signals with $1/f$ noise of length $N=1000$ and trend parameter $a=1.1$) there are values of S for which the accuracy obtained with homogeneous intervals is significantly better. For all the analyzed signals with AR(1) noise (results not presented here) the ACD algorithm with nonhomogeneous intervals have comparable or better accuracy. From these results we can conclude that although the distribution of the signal values into S^* nonhomogeneous intervals is not the best solution, however, its accuracy is good enough to be used in the automatic form of the ACD method.

Since the verification of the optimum values of the parameters ρ and K_f has been performed using the distribution of the signal values into homogeneous intervals, it is necessary to repeat the tests using nonhomogeneous intervals. The re-
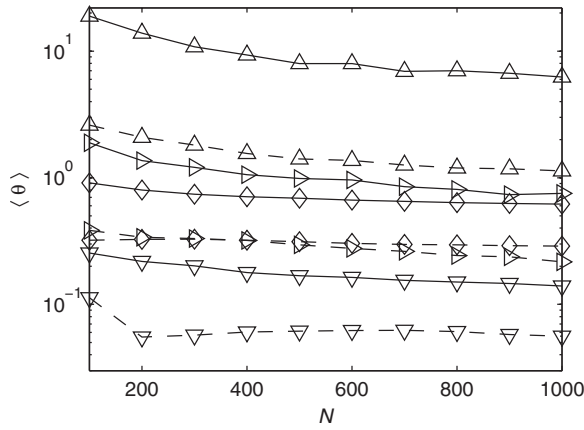
FIG. 5. Dependence of the average evaluation index of the ACD algorithm on the time series length. The time series differ from each other by the parameter a in trend formula $f(t) = t/(a-t)$, $t \in [0,1]$ (continuous lines correspond to $a=2.0$ and dashed lines to $a=1.1$) and by the standard deviation of the AR(1) noise with $\phi=0.9$ (marker $\triangle$ corresponds to $\sigma_Z=10$, $\triangleright$ to $\sigma_Z=1$, and $\triangledown$ to $\sigma_Z=0.1$). The signals with $1/f$ noise are marked with $\diamond$. A standard error smaller than 5% is obtained with statistical ensembles containing 300 time series.

sults are almost identical to those in Fig. 4 and we do not present them here.

Thus we have obtained an automatic form of the ACD method for monotone trend removal theoretically and numerically substantiated for time series with stationary noise which can be useful as well if the noise is nonstationary. Figure 5 shows the average evaluation index for the final automatic form of the ACD algorithm with nonhomogeneous intervals applied to the artificial time series discussed in Sec. III, with different lengths N. In all the cases, when N increases, the accuracy of the trend removal is improving. For signals with large noise and small slope ($\sigma_Z=10$ and $a=2.0$) for all the values of N, $\langle\theta\rangle \sim 10$ showing that the estimated trend follows more the noise in the signal than the real trend. Increasing the trend slope ($a=1.1$) or decreasing the noise ($\sigma_Z=1$), the evaluation index reaches the usefulness limit $\langle\theta\rangle \simeq 1$. For the rest of the time series, even for those with $1/f$ noise, we obtain $\langle\theta\rangle < 1$, indicating that they allow an efficient trend removal by means of the ACD algorithm.

On a standard 3 GHz Pentium 4 PC processing a time series with $N=1000$ required 0.4 CPU seconds for AR(1) noise and 0.8 CPU seconds for $1/f$ noise. The maximum number of partial estimated trend components was 4, and it was not significantly influenced by the series length or noise type. The number of the moving averagings increases directly proportional to the series length. For $N=1000$, on average, 80 smoothings were needed and the maximum number of smoothings was 140.

V. EVALUATION OF THE ACD ALGORITHM

The evaluation of the ACD method is performed by comparison with two trend removal methods used at present, one parametrical, the other nonparametrical. The trend is re-

moved from time series of the same type as those described in Sec. III. The comparison of the accuracy of different methods is realized by means of the quantity $\ln(\theta/\theta_{ACD})$, where θ is the evaluation index in Eq. (15) of the method to be compared with and θ_{ACD} is the evaluation index of the automatic ACD algorithm. We use the logarithm of the quantity θ/θ_{ACD} because this way the cases when $\theta \simeq \theta_{ACD}$ are better visualized. The result of a trend removal is better when the evaluation index is smaller and then the positive values of the quantity $\ln(\theta/\theta_{ACD})$ indicate that the ACD method gives a more accurate estimation of the trend than the reference method.

Figure 6 shows the results of the comparison with the polynomial fit of degrees from 1 to 10. First we notice that the polynomial fitting does not dispose of a parameter equivalent to ρ of the ACD algorithm, i.e., the polynomial trend is estimated by a single step and there is no possibility to control the magnitude of the final residual. In addition let us remind that if the signal is dominated by noise, then, as the degree of the polynomial increases, the number of coefficients to be determined by least-squares fit increases too and they are more strongly affected by the random component of the signal. Therefore the accuracy of the polynomial fit for noise dominated signals decreases with the polynomial degree. Reversely, if the signal is dominated by trend variation, a greater polynomial degree implies an improvement of the trend extraction results.

Since θ_{ACD} does not depend on the polynomial degree, the curves plotted in Fig. 6 describe the dependence of the quantity $\ln\theta_{poly}$ on the polynomial degree, but translated with a constant $-\ln\theta_{ACD}$, different for each type of time series. Hence comparing Figs. 6(b) and 6(c) one can see that, as expected, the polynomial fitting gives better results if the polynomial has a larger (smaller) degree when the signal is dominated by trend (noise). This contradictory behavior makes very difficult an automatic choice of the optimum value of the polynomial degree for all the possible types of time series.

From the first three panels of Fig. 6, it is difficult to draw a definite conclusion regarding the relation between the ACD accuracy and that of the polynomial fit. For $\sigma_Z=1$ [continuous lines in Fig. 6(a)], the ACD algorithm has better performance for short time series ($N=100$), when the fitted polynomial follows more the noise fluctuations than the trend, especially for higher polynomial degrees. If the resolution is increased ($N=1000$), then for degrees higher than 3, the two methods have comparable accuracy. For the signals dominated by trend in Fig. 6(b) ($\sigma_Z=0.1$) the accuracies are comparable for low resolution. At high resolution the results of the two methods are very sensitive to the polynomial degree. For degrees smaller than 5 the ACD accuracy is significantly better and for degrees higher than 5 significantly worse. For the signals dominated by noise in Fig. 6(c) ($\sigma_Z=10$), the ACD is better in almost all cases. Although the accuracy of the ACD algorithm is not significantly better than that of the polynomial fitting, this comparison indicates that the automatic form of the ACD algorithm has been obtained preserving performance comparable with that of existing algorithms.

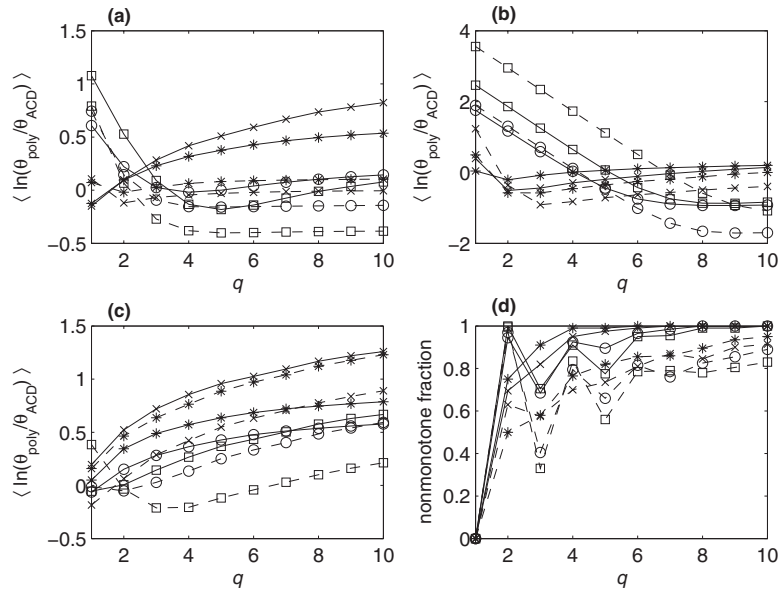For signals with $1/f$ noise [dashed lines in Fig. 6(a)], except the polynomials with degree 1 or 2, the ACD algo-

FIG. 6. Comparison of the ACD accuracy with that of the polynomial fit in terms of the polynomial degree q. The time series are characterized by different values of their length N and of the parameter a in the trend formula $f(t)=t/(a-t)$, $t \in [0,1]$. The couple (N,a) takes the values $(100,2.0)$ (asterisk marker $*$), $(100,1.1)$ (cross marker $\times$), $(1000,2.0)$ (circle marker $\bigcirc$), and $(1000,1.1)$ (square marker $\square$). (a) The random component is a stationary AR(1) noise with $\sigma_Z=1$ and $\phi=0.9$ (continuous lines) and a nonstationary fractional Brownian motion with $\gamma=1.4$ (dashed lines). (b) The random component is a stationary AR(1) noise with $\sigma_Z=0.1$ and $\phi=0.9$ (continuous lines) and $\phi=0$ (dashed lines). (c) The random component is a stationary AR(1) noise with $\sigma_Z=10$ and $\phi=0.9$ (continuous lines) and $\phi=0$ (dashed lines). (d). The random component is as in panel (a) but the represented quantity is the fraction of the time series for which the polynomial estimated trend is not monotone. A standard error for $\langle \theta_{\mathrm{poly}} \rangle$ smaller than 5% is obtained with statistical ensembles containing 200 time series.

rithm has worse results. An explanation of this behavior could be that many times it is difficult to differentiate a nonstationary noise containing a so-called "stochastic trend" from a variation due to a deterministic trend. For such signals the ability of the ACD algorithm to adjust itself to the signal shape is a drawback since the estimated trend describes not only the deterministic trend but also the nonstationary noise.

Finally, Fig. 6(d) shows the fraction of the time series for which the estimated polynomial trend is not monotone. Obviously, the polynomials of order 1 preserve the monotony, but for higher orders more than one-half of the estimated trends are not monotone. It is interesting to remark that the polynomials of odd degree allow a better identification of the trend monotony.

The second comparison method is the "jump process" in [5] and it consists in an iterative weighted averaging

$$x_n^{(i+1)} = x_n^{(i)} + R(x_{n-1}^{(i)} - 2x_n^{(i)} + x_{n+1}^{(i)}), \qquad (17)$$

where ratio R $(0<R<0.5)$ and iteration parameter M $(i \leq M)$ are user-specified constants. At boundaries a symmetric extension of data is employed. As in the case of the polynomial fitting, the averaging in Eq. (17) has opposite effects on the accuracy of the estimated trend. When the effect of the averaging is greater (larger values of the parameters R and M), the noise fluctuations are more strongly smoothed, but at the same time the trend shape is more distorted.

Figure 7 contains the results of the comparison of the ACD algorithm with the jump process for different values of the parameter R and for $M=100$ fixed. For signals with AR(1) noise, excepting a few cases with $\sigma_Z=0.1$ in Fig. 7(b), the ACD method has better accuracy. For signals with $1/f$ [dashed lines in Fig. 7(a)], although not so bad as for the polynomial fit, the results of the ACD algorithm continue to be worse than those of the comparison method. As discussed above, comparing Figs. 7(b) and 7(c) one notices that, as expected, the jump process has better results if the averaging parameter R is smaller (larger) when the signal is dominated by trend (noise). From Fig. 7(d), which shows the fraction of time series for which the estimated trend is not monotone, it follows that when R increases, more and more estimated trends preserve the monotony of the real trend. The fraction of the nonmonotone estimated trends depends also on the standard deviation of the noise and on the signal length and it becomes smaller if we increase the number of iterations M.

From the comparisons discussed above, we conclude that, on average, the ACD algorithm removes monotone trends with slightly better accuracy than the other two methods if the noise is stationary. In the case of nonstationary noise the accuracy is worse. Nevertheless, the ACD algorithm preserves the advantage of its automatic form. In addition, the ACD algorithm removes only monotone trend, not nonmonotone trend as the existing algorithms do. To illustrate these special characteristics we estimate by means of the three algorithms the monotone trend of a real time series from paleoclimatology. Figure 8(a) shows the global mean annual temperature anomalies during the period A.D. 200–
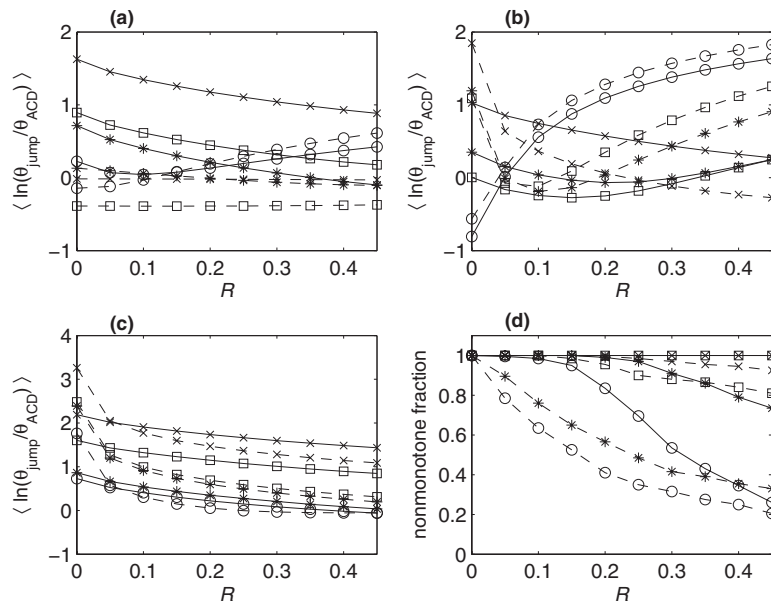
FIG. 7. Comparison of the ACD accuracy with that of the jump process in terms of the averaging parameter R for a fixed number of iterations $M=100$. The time series are characterized by different values of their length N and of the parameter a in the trend formula $f(t)$ $=t/(a-t)$, $t \in [0,1]$. The couple (N,a) takes the values $(100,2.0)$ (asterisk marker $*$), $(100,1.1)$ (cross marker $\times$), $(1000,2.0)$ (circle marker $\bigcirc$), and $(1000,1.1)$ (square marker $\square$). (a) The random component is a stationary AR(1) noise with $\sigma_Z=1$ and $\phi=0.9$ (continuous lines) and a nonstationary fractional Brownian motion with $\gamma=1.4$ (dashed lines). (b) The random component is a stationary AR(1) noise with σ_Z $=0.1$ and $\phi=0.9$ (continuous lines) and $\phi=0$ (dashed lines). (c) The random component is a stationary AR(1) noise with $\sigma_Z=10$ and ϕ $=0.9$ (continuous lines) and $\phi=0$ (dashed lines). (d). The random component is as in panel (a) but the represented quantity is the fraction of the time series for which the estimated trend is not monotone. A standard error for $\langle \theta_{jump} \rangle$ smaller than 5% is obtained with statistical ensembles containing 200 time series.

1995 with respect to the Northern Hemisphere mean annual temperature over 1856–1980 discussed in [9] and freely accessible from [10].

The temperature anomalies series contains $N=1796$ values, many of them repeating themselves. In fact there are only 48 distinct values and the distribution of the time series values into disjoint intervals used in the ACD algorithm demands that all values should be distinct. To satisfy this request without distorting the initial signal, we superpose on the original values numbers randomly generated with a homogeneous probability distribution on a range 1000 times

smaller than the minimum difference between two distinct signal values. The ACD algorithm has estimated the trend in a single removal performed after 129 averagings. The series values are distributed into $S^*=31$ nonhomogeneous intervals obtained by splitting $S^{est}=9$ homogeneous intervals. The estimated standard deviation of the noise is $\sigma_Z^{est}=0.056$ in comparison with the standard deviation of the initial series $\hat{\sigma}_x$ $=0.063$. In Fig. 8(b) the ACD estimated trend is compared with those estimated by the other two methods. As it follows from Fig. 6(d), most of the estimated trends by nonlinear polynomial fit are nonmonotone. The linear trend, the only
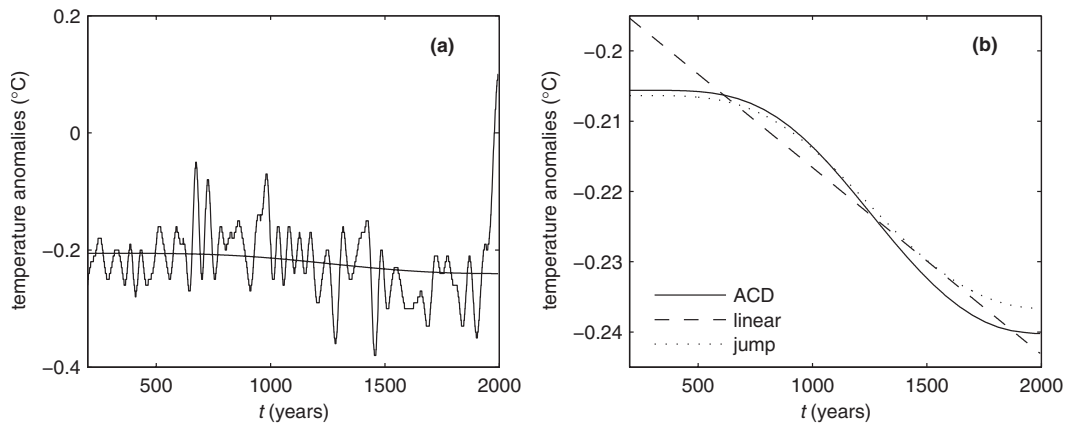

FIG. 8. Global mean annual temperature anomalies during the period A.D. 200–1995 with respect to the Northern Hemisphere mean annual temperature over 1856–1980 [9] and the trend estimated by means of ACD algorithm (continuous line). In panel (b) the trends estimated by means of the linear fit (dashed line) and jump process (dotted line) are also presented.

polynomial trend which is assuredly monotone, exaggerates the global variation of the time series and does not provide any information on the slope variation over the analyzed interval. In the case of the jump process, the time series must be averaged until all the nonmonotone variations of the estimated trend are eliminated. For $R = 0.4$, $M = 329\,637$ averagings were needed such that the computing time became prohibitive. Besides that, even if the estimated trend has the same shape as the ACD trend, due to the large number of averagings, the magnitude of the global temperature variation is underestimated.

As shown in [8], it is possible that, when the noise is large, the trend removal should not be recommended. The series in Fig. 8(a) is such a case. In order to verify that the estimated trend $\{\hat{F}_n\}$ is significant, we build surrogate series $\hat{x}_n = \hat{F}_n + \hat{\zeta}_n$. Since the serial dependence of the noise has a significant influence on the estimated trend quality, we choose the surrogate series $\{\hat{\zeta}_n\}$ as realizations of an AR(1) stochastic process. We note that the white noise is obtained for $\phi = 0$, so it is implicitly included among the possible surrogate series. From the standard deviation and the autocorrelation at one time step of the estimated noise in Eq. (10) we compute the value of the parameter ϕ used in surrogates generation. The fraction of the estimated trends for the surrogate series $\{\hat{x}_n\}$ having the same sign of the global variation as the estimated trend $\{\hat{F}_n\}$ for the original series is a measure of the probability that the estimated trend should correspond to a real one in the original series. For example, from the estimated trends for 100 surrogate series of the time series in Fig. 8(a), only 58 are decreasing showing that the association of a trend to the initial climatological series is hazardous.

Climatologists are interested in the time periods with a monotone temperature variation associated with geophysical processes of global scale, as, for example, the global warming in the last century. I have applied the surrogate series method described above on the global temperature anomalies over time intervals measured in centuries, i.e., intervals $[t_1, t_2]$ with

$$t_1 = 200, 300, \ldots, 1900,$$

$$t_2 = 300, 400, \ldots, 1900, 1995$$

and $t_2 > t_1$. The results are presented in Fig. 9. There are only three time periods to which we can associate a monotone trend with a confidence level of 99%. Two of them are related to the global warming over the last 2 centuries and the third one corresponds to the 10th century. For 19 time periods the ACD algorithm did not succeed in associating a monotone trend to the temperature variation. The rest of the estimated trends cannot be considered "significant," a conclusion which coincides with that in [9].

VI. NONMONOTONE TREND

So far we have analyzed only time series with a monotone trend. In this section we consider time series without trend or
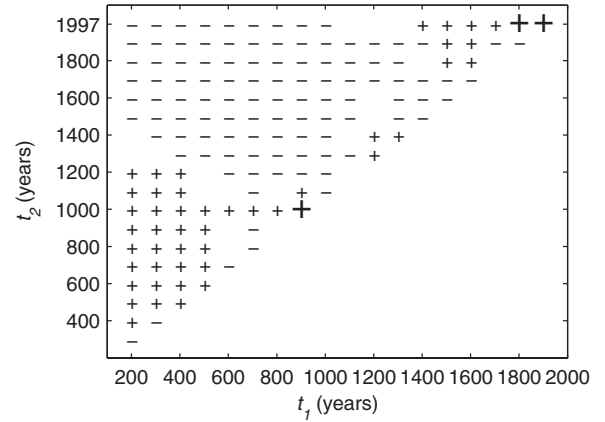

FIG. 9. Time periods $[t_1, t_2]$ over which the global average temperature has an increasing ($+$ sign) or decreasing ($-$ sign) trend. The three bigger signs correspond to significant trends with a confidence level of 99%. The ACD method did not succeed to estimate a monotone trend for the time periods to which no sign is attached.

with nonmonotone trend. In these cases, as well, the ACD evaluated trend is monotone and it may or may not correspond to a monotone component of the actual trend. Therefore an evaluation of the statistical significance of the estimated trend is needed, either using the surrogate series method presented in the previous section, or applying a statistical test for monotone trend detection. In the following we compare the results of the ACD algorithm with the Mann-Kendall test [11].

The Mann-Kendall test is based on the quantity $S = P - M$, where P is the number of the pairs $x_n > x_m$ with $n > m$ and M is the number of pairs $x_n < x_m$ with $n > m$. If $\{x_n\}$ are independent observations, then for $N > 10$ the random variable

$$Z_{MK} = \begin{cases} (S-1)/\sigma_S & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ (S+1)/\sigma_S & \text{if } S < 0 \end{cases}$$

with $\sigma_S = [N(N-1)(2N+5)/18]^{1/2}$ follows a standard normal distribution. The null hypothesis that there is no monotone trend is rejected when the computed Z_{MK} value is greater in absolute value than the critical value $z_{\alpha/2}$, where α is the chosen significance level. This test was extended to a serial dependent time series [12], but here we restrain ourselves to the simple original form of the test.

The Mann-Kendall test is used especially in hydrology, other environmental sciences, and econometrics for short time series, even containing only several dozens of values. In order to obtain a complete analysis of the ACD algorithm performance, we process time series of lengths down to the inferior limit of validity of the Mann-Kendall test, i.e., $N = 10$. To make the ACD algorithm applicable to such time series, we modify it as follows: for $N < 2N_{\min} = 28$, instead of nonhomogeneous intervals we use $S = 2$ homogeneous intervals to distribute the signal values. According to Appendix A, in such cases the estimated noise standard deviation has a
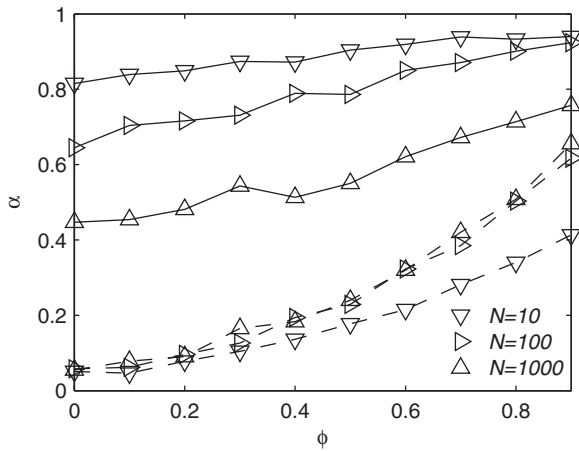
FIG. 10. The probability of the type I error for the Mann-Kendall test with 5% significance level applied to AR(1) noise with different serial dependence (dashed line). The continuous line represents the relative frequency of the monotone estimated trends by the ACD algorithm for the same time series. For the results of the ACD algorithm a standard error smaller than 4% is obtained with statistical ensembles containing 1000 time series.

larger error, however, the ACD algorithm can be applied without any further modifications.

The so-called type I error of a statistical test occurs when the null hypothesis is rejected although it is true. Figure 10 shows the probability α of type I error occurrence for the Mann-Kendall test with 5% significance level applied to AR(1) noise without trend, in terms of the correlation parameter ϕ, for different lengths of the time series. The figure also shows the relative frequency of the monotone trends estimated by means of the ACD algorithm for the same time series. One notices that the ACD algorithm estimates incorrectly a monotone trend much more frequently than it could occur as a random event. The error of the ACD algorithm diminishes when the time series length increases, however, it remains much greater than the probability α of type I error occurrence for the Mann-Kendall test.

All the trend estimation methods present errors due to the lack of discrimination between monotone and nonmonotone trends. As shown in Fig. 6(d) for nonlinear polynomial fit and Fig. 7(d) for the jump process, these methods estimate many times a nonmonotone trend when the actual one is monotone. Being designed to remove the monotone component of the trend, the ACD algorithm has an opposite behavior.

The type II error of a statistical test occurs when the null hypothesis is accepted although it is false. Our null hypothesis is false if the signal contains a monotone trend. To analyze this kind of situation we introduce a trend composed of a monotone linear trend and a sinusoid $f(t) = p(t + a \sin 2\pi t)$, $t \in [0,1]$, where p and a are positive real parameters. This formula allows us to make a continuous transition from a monotone to a nonmonotone trend. When $a = a_0 = 1/(2\pi) \simeq 0.159$, the function $f(t)$ has an inflection point at $t = 0.5$ with the tangent parallel to the Ox axis. Therefore if $a \leq a_0$ the trend is monotone, and if $a > a_0$ it is nonmonotone.

The power of a statistical test is defined as one minus the probability of type II error. Figure 11 shows the power of the
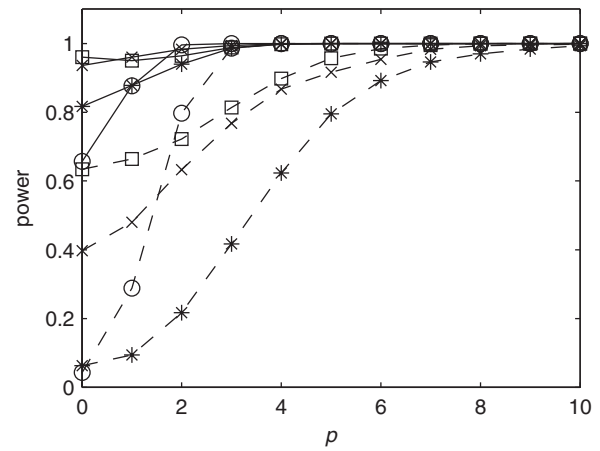

FIG. 11. The power of the Mann-Kendall test with 5% significance level (dashed lines) for time series with different lengths N obtained by superposing AR(1) noise with $\sigma_Z = 1$ on the monotone trend $f(t) = p(t + a_0 \sin 2\pi t)$, $t \in [0,1]$, in terms of the parameter p. The continuous line represents the relative frequency of the monotone estimated trends by the ACD algorithm for the same time series. The couple (N, ϕ) takes the values $(10,0)$ (asterisk marker $*$), $(10,0.9)$ (cross marker $\times$), $(50,0)$ (circle marker $\circ$), and $(50,0.9)$ (square marker $\square$). For the results of the ACD algorithm a standard error smaller than 2% is obtained with statistical ensembles containing 1000 time series.

Mann-Kendall test for time series with $\sigma_Z = 1$ and the monotone trend described in the previous paragraph with $a = a_0$ and different values of the parameter p. As expected, the closest shape to the ideal test power (power vanishing for $p = 0$ and rapidly increasing up to 1) is obtained for white noise ($\phi = 0$) and longer series ($N = 50$). For correlated noise the results are worse because the probability of type I error is much larger than the significance level. Like in Fig. 10, the relative frequency of the monotone trends estimated by the ACD algorithm for the same time series is much greater than the Mann-Kendall test power, reinforcing the previous conclusion that the ACD algorithm estimates the monotone component of the trend.

In Fig. 12 we give an example of a monotone component estimated by the ACD algorithm for a nonmonotone trend. We have used a signal with $N = 1000$ values given by the nonmonotone trend $f(t) = t + 2a_0 \sin 2\pi t$, $t \in [0,1]$, without superposed noise. One notices that the estimated trend is significantly different from the "real" monotone component of the trend, which is a straight line. This example shows that for a nonmonotone trend one cannot define unequivocally a monotone component and the ACD monotone estimated trend is only one of the infinitely existing possibilities.

To evaluate the ACD algorithm results for nonmonotone trend, the same method as that in Sec. V has been used. The time series have $N = 10$ and, respectively, $N = 50$ values, in order to provide information on the ACD performance for very short time series too. The trend is given by $f(t) = t + a \sin 2\pi t$ with variable a, such that the trend is continuously modified from a linear one ($a = 0$) to the nonmonotone one represented in Fig. 12 ($a = 2a_0$). Figure 13(a) shows the dependence of the average evaluation index of the ACD al-
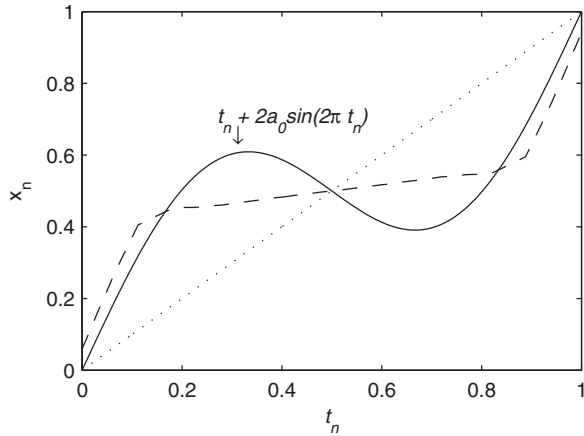
FIG. 12. The estimated ACD trend (dashed line) for a time series (continuous line) with $N=1000$ values, without noise, obtained by superposing a sinusoid on a linear component (dotted line), using dimensionless time $t_n=n/N$.

gorithm on the parameter a. From these results the same conclusions as those in Sec. IV in connection with Fig. 5 can be drawn: the accuracy is improving when N increases or ϕ decreases and the trend removal is efficient if the signal is not dominated by noise. Furthermore, the ACD accuracy is worsening when the contribution of the harmonic component increases.

Since, as seen in Sec. V, the performance of the polynomial fit increases for smaller degrees, we compare in Fig. 13(b) the ACD accuracy only with the linear fit. For small values of parameter a the trend is almost linear, such that the accuracy of the linear fit is better in all the cases. For greater values of the parameter a and for small noise, the ACD accuracy becomes better ($N=50$) or at least comparable ($N=10$). Only for a signal dominated by white noise ($\phi=0$ and $\sigma_Z=1$) the results of the linear fit are better in all the cases. These are the only time series for which the jump process has a better accuracy than the ACD algorithm [see

Fig. 13(c)]. So we can conclude that, also for very short time series or nonmonotone trend, the ACD accuracy remains comparable with that of the usual trend estimation methods. Besides, this result is obtained preserving the automatic feature of the ACD algorithm.

VII. CONCLUSIONS

The automatic ACD algorithm for monotone trend removal presented in this paper can be applied on a time series with stationary noise without any preliminary subjective inspection of the time series. The analysis of its performance on artificial signals with AR(1) noise indicates that the quality of the estimated trend is slightly better than that of the usual nonautomatic methods. This conclusion is confirmed by the results obtained for a real time series from paleoclimatology.

The preliminary tests for nonstationary $1/f$ noise show that the ACD accuracy is worse than that of the polynomial fit and likely of the parametric methods in general. The explanation is presumably related to the fact that the realizations of a nonstationary stochastic process are difficult to differentiate from a signal containing a deterministic trend component. Then the ability of the ACD method to adjust itself to the signal shape makes possible the confusion between the stochastic trend of the nonstationary noise and the deterministic trend, whereas the parametrical methods which have less degrees of freedom are less affected by such confusions. Therefore the application of the ACD method to nonstationary noises needs a special analysis. However, the possibility to apply automatically the ACD method to a large number of time series is an advantage which may justify its utilization in some applications like DFA.

The analysis of the automatic ACD algorithm presented in this paper should be extended. Further tests should be performed both on different real time series and artificial ones with greater lengths ($N>1000$), other types of stationary noises, and other functional forms of the monotone trend.
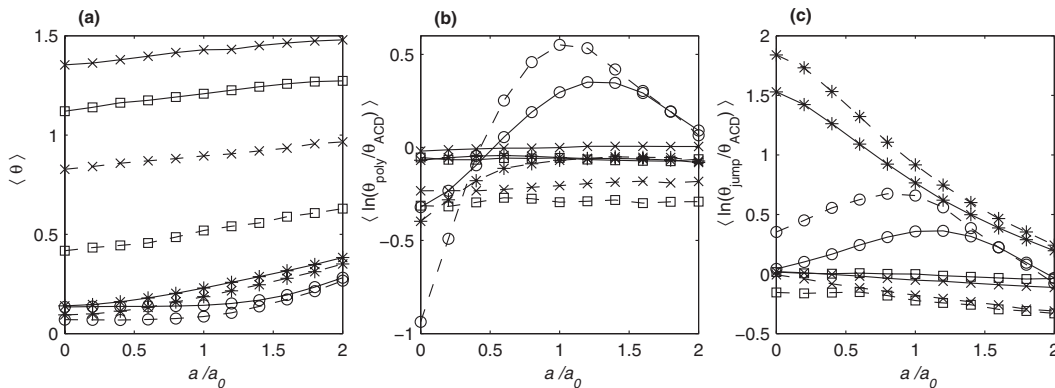

FIG. 13. (a) Dependence of the average evaluation index of the ACD algorithm on the parameter a in the trend formula $f(t)=t+a \sin 2\pi t$, $t \in [0,1]$. The time series are characterized by different values of their length N and the random component of the signals is a stationary AR(1) noise with standard deviation σ_Z and different serial dependence $\phi=0.9$ (continuous lines) and $\phi=0$ (dashed lines). The couple (N,σ) takes the values (10,0.1) (asterisk marker $*$), (10,1) (cross marker $\times$), (50,0.1) (circle marker $\bigcirc$), and (50,1) (square marker $\square$). A standard error smaller than 2% is obtained with statistical ensembles containing 1000 time series. (b) Comparison of the ACD accuracy with that of the linear fit for the same time series as for panel (a). (c) Comparison of the ACD accuracy with that of the jump process with $R=0.4$ and $M=100$ for the same time series as for panel (a).

Also it is necessary to compare the performance of the ACD algorithm with other trend removal methods besides the two algorithms presented here (polynomial fit and jump process), in comparison with which the ACD accuracy was proved to be slightly better.

The extension of the ACD method to nonmonotone trends can be achieved by applying the ACD method separately to the increasing, respectively, decreasing part of the signal. This is possible if the sample average in Eq. (7) is computed separately for the positive, respectively, negative variations of the time series values. In this case a unique trend cannot be determined because the final result depends on the manner in which the two monotone parts of the estimated trend are combined. However, the main advantage of the ACD method is preserved, i.e., the functional form of the estimated trend need not be imposed and the design of an automatic algorithm for nonmonotone trend removal is possible.

In essence the ACD method tries to use in a more complete way the information available in the transition probability of a stochastic process. The most elaborate method of this type is that in nonequilibrium statistical physics, where a hydrodynamical description of a fluid is obtained by averaging the two-particle distribution function. In [13] it was shown that a description of hydrodynamical type can be obtained for any corpuscular system using space-time coarse-grained averages of a kinematic description of the movement of the component particles. This approach was applied to a one-dimensional system [14] and then to a time series with the space coordinate replaced by the values of a share price [15]. From this point of view, the ACD method is a discretized form of the space-time coarse-grained averages with disjoint space averaging intervals and the time averaging interval identical with the signal duration. It is possible to develop the ACD method for a time series with nonmonotone trends as a more complete hydrodynamical description if multiple overlapping subintervals are used for both space and time.

APPENDIX A: ESTIMATION OF NOISE STANDARD DEVIATION

A justification (not a rigorous proof) of estimation (12) of the noise standard deviation for a realization $\{x_n\}$ of the stochastic process in Eq. (1) is based on the fact that Z_n does not depend on the trend values f_n. Then for N large enough we have the approximate equality

$$\|\delta_m x_n\|^2 \simeq \|\delta_m f_n\|^2 + \|\delta_m z_n\|^2, \tag{A1}$$

where $\|\cdot\|$ is the usual quadratic norm and, as in Sec. III, we use the notation $\delta_m x = x_{m+n} - x_n$. Using the property $\bar{f} = 0$, one can prove that if $\{f_n\}$ is monotone, then for $m < N/2$ the sequence $\|\delta_m f_n\|$ is increasing

$$\|\delta_{m+1} f_n\| > \|\delta_m f_n\|. \tag{A2}$$

It follows that the left-hand term in Eq. (A1) decreases when m increases only due to the term $\|\delta_m z_n\|^2$. According to Eq. (13), m_0 is the number of time steps when the first decrease of $\|\delta_m x_n\|^2$ takes place, i.e., the value of m for which the
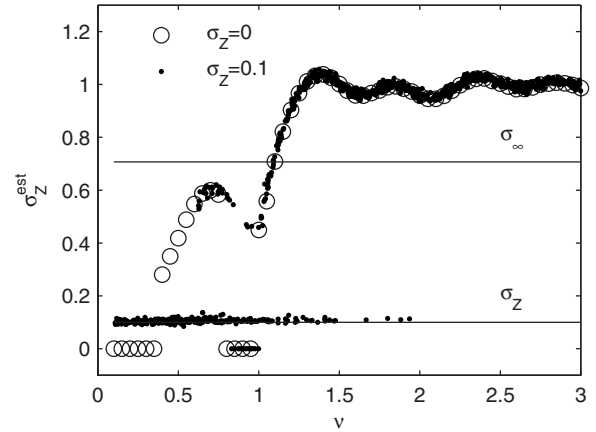

FIG. 14. Estimated standard deviation σ_Z^{est} of a Gaussian white noise superposed on a sinusoidal signal with frequency ν, $f(t) = \sin 2\pi\nu t$, $t \in [0,1]$, discretized on $N = 200$ time steps. The standard deviation of the noise is $\sigma_Z = 0$ (circle markers) and $\sigma_Z = 0.1$ (points markers). The horizontal lines represent the nonvanishing value of the noise standard deviation and the standard deviation of an infinite sinusoid $\sigma_\infty = 1/\sqrt{2}$.

variations of $\{x_n\}$ are dominated by noise. Since we intend to estimate only the magnitude order of σ_Z, we neglect in Eq. (A1) the term due to the trend. If m_0 is large enough to render Z_{m_0+n} and Z_n independent, then Eq. (A1) becomes

$$\|\delta_{m_0} x_n\|^2 \approx 2\|z_n\|^2. \tag{A3}$$

Since $\langle Z_n \rangle = 0$, the right-hand term is proportional with the sample standard deviation of the noise $\hat{\sigma}_Z$. For the left-hand term it is not rigorously true that it is proportional to the corresponding standard deviation, but taking into account the approximations made until now we write Eq. (A3) as estimation (12). The numerical tests presented in Fig. 3 show that the obtained estimation is remarkable for an approximation only of the magnitude order of the noise standard deviation. The estimation quality remains the same for other functional forms of the monotone trend (polynomial, exponential, logarithmic, etc.).

If we want to apply estimation (12) on real signals, we have to establish what information on the noise standard deviation is supplied by this formula if the trend is not monotone. Then Eq. (A2) is not true and the negative variations of $\|\delta_m x_n\|^2$ with respect to m can result from both right-hand terms in Eq. (A1). So, even if the signal does not contain noise, estimation (12) can take for noise the trend oscillations. Figure 14 shows the results obtained applying Eq. (12) to a sinusoidal signal $f(t) = \sin 2\pi\nu t$, $t \in [0,1]$. For $\nu < 0.25$, the signal is monotone and coincides with the trend, so that the noise term in Eq. (A1) vanishes and according to Eq. (A2) $\|\delta_m x_n\|$ is increasing. Then there is no m_0 satisfying Eq. (13) and it is normal to consider that $\sigma_Z^{\text{est}} = 0$. In Fig. 14 $\sigma_Z^{\text{est}} = 0$ not only for $\nu < 0.25$, but also for $\nu < \nu_1 = 0.4$. Hence according to evaluation (12), a sinusoid containing up to ν_1 periods is taken for a trend without noise, even if it is not monotone anymore. Consequently, the estimation (12) is able to recognize nonmonotone trends as well. For $\nu > \nu_2 = 1$, the estimated standard deviation oscillates about an asymptotic

value $\sigma_\infty^{est}=1$, a larger value than the standard deviation of an infinite sinusoid $\sigma_\infty=1/\sqrt{2}$. In other words, a sinusoid containing more than ν_2 periods is erroneously identified to a noise with standard deviation σ_∞^{est}. The interval $\nu \in (\nu_1,\nu_2)$ represents a transition domain between the two dominant behaviors of evaluation (12).

Actually we are interested first of all in the ability of estimation (12) to identify a noise superposed on a nonmonotone trend. Therefore, on the sinusoid analyzed above we superpose a white Gaussian noise with $\sigma_Z=0.1$. In this case the results strongly depend on the signal resolution. Figure 14 shows the estimations for $N=200$ and one notices that for different realizations of the noise, different values of σ_Z^{est} for the same ν can be obtained. However, if $\nu<\nu_1$, all values of σ_Z^{est} are clustered about the real value σ_Z, and if $\nu>\nu_2$, most values cluster about the value σ_∞^{est} obtained for sinusoid without noise. Hence if the sinusoid without noise is interpreted as a trend, then the noise standard deviation is correctly estimated and when it is interpreted as noise then the whole signal is considered as noise. If the signal resolution increases ($N>200$) then the correct value of σ_Z is obtained for $\nu>\nu_2$ too. For example, if $N=1000$, then $\sigma_Z^{est}\simeq\sigma_Z$ for $\nu<2$. This result proves that Eq. (12) gives a useful estimation of the order of magnitude of a noise superposed on a nonmonotone trend. A detailed analysis of the extension for nonmonotone trends of estimation (12) will be the subject of a future paper.

As a first application of Eq. (12) we estimate the minimum number of values N_{min} needed lest a Gaussian white noise should be interpreted as a trend. If the number of values is very small ($N\leq10$), then there is a significant probability that no m_0 satisfying Eq. (13) should exist and the noise is interpreted as a trend. For example, for $N=4$ [the minimum value of N for which Eq. (13) is meaningful], out of 1000 realizations of a Gaussian white noise with unit standard deviation 35% are not recognized as noise. The largest value of N for which we still obtain $\sigma_Z^{est}=0$ in 0.1% cases is $N=12$. However, we choose $N_{min}=14$ in order to be sure that in no case $\sigma_Z^{est}=0$, and then $\langle\sigma_Z^{est}\rangle=1.105$ with $\sigma_Z^{est}\in[0.38,2.26]$. These results are consistent with the Mann-Kendall test for monotone trends [11], which is applicable only if $N>10$. The estimation of the noise standard deviation is strongly improved when N increases. For $N=1000$ we obtain $\langle\sigma_Z^{est}\rangle=1.015$ with $\sigma_Z^{est}\in[0.935,1.110]$. These results are used in Sec. IV and in Appendix B to distribute the time series values into disjoint intervals.

In the case of correlated AR(1) noise, the value of m_0 increases when ϕ approaches 1. For example, for 1000 time series with $\phi=0.9$, $\sigma_Z=1$ and $N=1000$, the average value of m_0 is 26.9 and $\langle\sigma_Z^{est}\rangle=0.984$ with $\sigma_Z^{est}\in[0.724,1.365]$. The average value $\langle\sigma_Z^{est}\rangle$ is as accurate as that for the Gaussian

white noise, but the fluctuations of the individual values are greater. This behavior is due to the fact that for large values of ϕ the stochastic trend becomes important enough to be interpreted in Eq. (12) as a deterministic trend.

APPENDIX B: NONHOMOGENEOUS INTERVALS OF SIGNAL VALUES

The signals dominated by noise or trend impose different numerical restrictions on the ACD algorithm and it is advisable that the numerical algorithm should adapt itself to the signal type. Since the parameter η_0 defined by Eq. (11) can be calculated only if we know the signal components, we have to introduce a new parameter expressed by means of the values of the initial series

$$\eta^{est} = \left| \frac{\hat{\sigma}_x}{\sigma_Z^{est}} - 1 \right|, \tag{B1}$$

where $\hat{\sigma}_x$ is the sample standard deviation of the signal and σ_Z^{est} is given by estimation (12). If $\eta^{est}<1$, then $|\hat{\sigma}_x-\sigma_Z^{est}|<\sigma_Z^{est}$ and the noise standard deviation represents the largest part of the whole signal, i.e., the signal is dominated by noise. Conversely, if $\eta^{est}>1$, then the signal is dominated by trend.

Equation (16) applied to a signal dominated by noise ($\eta^{est}<1$) generates a small number of homogeneous intervals I_s, each of them containing a large number of values. For longer signals, the number of values in I_s increases and the computing time can become prohibitive. For this reason, when $\eta^{est}<1$, we split into two subintervals those intervals which by splitting generate a larger number of values than the quantity N_{min} defined in Appendix A. For example, consider the interval $I_s=(\xi_s,\xi_{s+1})$. By splitting it, we obtain two subintervals $I'_s=(\xi_s,\xi'_{s+1})$ and $I'_{s+1}=(\xi'_{s+1},\xi_{s+1})$, where $\xi'_{s+1}=(\xi_s+\xi_{s+1})/2$. The two subintervals contain in general different numbers of values $N'_s \neq N'_{s+1}$. If $N'_s \geq N_{min}$ and $N'_{s+1} \geq N_{min}$, then the splitting is validated and the number of intervals I_s increases by one. If all S^{est} intervals can be split generating $2S^{est}$ new intervals, then the splitting process is repeated until at least one of the intervals does not satisfy anymore the condition that the two subintervals should contain more than N_{min} values. In the end we obtain S^* nonhomogeneous intervales I_s containing arbitrary numbers of time series values.

Applied to a signal dominated by trend ($\eta^{est}>1$), Eq. (16) generates a large number of homogeneous intervals I_s, each of them with a small number of values inducing large fluctuations of the mean slope in Eq. (7). Therefore we merge the intervals with smaller length than σ_Z^{est}, $\xi_{s+1}-\xi_s<\sigma_Z^{est}$. Concretely, the successive intervals $I_s,I_{s+1},\ldots,I_{s+m}$ are joined in a new interval $I'_s=(\xi_s,\xi_{s+m+1})$, containing $N'_s=N_s+N_{s+1}+\cdots+N_{s+m}$ values if $\xi_{s+m+1}-\xi_s<\sigma_Z^{est}$ and $\xi_{s+m+2}-\xi_s>\sigma_Z^{est}$. Thus we obtain S^* nonhomogeneous intervals.

[1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forcasting and Control*, 2nd ed. (Holden-Day, San Francisco, 1976).

[2] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis* (Cambridge University Press, Cambridge, England, 2000).

[3] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994).

[4] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley, Phys. Rev. E **64**, 011114 (2001).

[5] S. Zhao and G. W. Wei, Comput. Stat. Data Anal. **42**, 219 (2003).

[6] P. J. Brockwell and R. A. Davies, *Time Series: Theory and Methods*, 2nd ed. (Springer-Verlag, New York, 1991).

[7] N.-N. Pang, Y.-K. Yu, and T. Halpin-Healy, Phys. Rev. E **52**, 3224 (1995).

[8] E. L. Andreas and G. Trevino, J. Atmos. Ocean. Technol. **14**, 554 (1997).

[9] P. D. Jones and M. E. Mann, Rev. Geophys. **42**, 1 (2004).

[10] P. D. Jones and M. E. Mann, *Climate over Past Millenia*, Data Contribution Series # 2004-085 (2004), in NOAA/NGDC Paleoclimatology Program, URL http://www.ncdc.noaa.gov/paleo/pubs/jones2004/

[11] M. G. Kendall, *Rank Correlation Methods*, 4th ed. (Griffin, London, 1975).

[12] R. M. Hirsch and J. R. Slack, Water Resour. Res. **20**, 727 (1984).

[13] C. Vamoş, A. Georgescu, N. Suciu, and I. Turcu, Physica A **227**, 81 (1996).

[14] C. Vamoş, N. Suciu, and A. Georgescu, Phys. Rev. E **55**, 6277 (1997).

[15] C. Vamoş, N. Suciu, and W. Blaj, Physica A **287**, 461 (2000).