Chapter 10

The monotone trend estimation

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 [9] or wavelet analysis [54]. 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 chapter a trend removal algorithm, which needs no initial assumptions, is proposed. It is obtained by generalizing the coarse-grained averages analyzed in the previous chapters if the signal values are divided in disjoint intervals with different lengths and the temporal scale is greater than the length of the time series. Then the trend is approximated by a piecewise linear curve the slope of each line segment being proportional to the average one-step displacement of the signal values included into the corresponding interval. This method allows the development of an automatic numerical algorithm for the evaluation of a monotone trend that was tested on artificial time series [61].

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 method presented here is that it can describe a much richer set of monotone trends as piecewise linear functions.

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 chapter. But even when the trend monotony is not significant, the initial separation of a monotone component from the global non-monotonic 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.

10.1 The mean slope of the estimated trend

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

$$X_n = f_n + Z_n\,, \tag{10.1}$$

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 $\bar{f} = N^{-1} \sum_{n=1}^{N} f_n = 0$.

If the stochastic process $\{Z_n\}$ is stationary, then a mathematical justification of the proposed 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. (10.1), 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

$$\langle N_\xi \rangle = \sum_{n=1}^{N} p_z(\xi - f_n)\,. \tag{10.2}$$

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

$$g(\xi)\delta\xi = \frac{1}{\langle N_\xi \rangle} \sum_{n=1}^{N-1} p_z(\xi - f_n) \left\langle \delta X_n | X_n \in I_\xi \right\rangle\,, \tag{10.3}$$

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. (10.3) becomes

$$g(\xi) = \frac{1}{\langle N_\xi \rangle} \sum_{n=1}^{N-1} (f_{n+1} - f_n)\, p_z(\xi - f_n) + \langle \varepsilon_\xi \rangle\,. \tag{10.4}$$

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. (10.4)

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

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 above conditions, since $f_{n+1} - f_n = a\,\delta t$, from Eqs. (10.4) and (10.2) 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 estimated trend $F(t)$ 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

$$g(\xi) = F'(F^{-1}(\xi))\, \delta t\,. \tag{10.5}$$

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. (10.3) using the values of a time series $\{x_n\}$ obtained as a realization of the stochastic process $\{X_n\}$. The infinitesimal intervals I_ξ had to 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. (10.2). 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 trend estimation algorithm we use a distribution into nonhomogeneous intervals described later in this chapter.

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

$$\widehat{g}_s = \frac{1}{2N_s} \left(\sum_{x_n \in I_s} \delta x_n + \sum_{x_{n+1} \in I_s} \delta x_n \right) , \tag{10.6}$$

which corresponds to the definition in Eq. (10.3). If all the values $\widehat{g}_s$ have the same sign, then we can use them to determine a numerical approximation of the trend by a piecewise linear curve denoted $\widehat{F}(t)$ and corresponding to the theoretical one in Eq. (10.5). 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 $\widehat{F}(t)$ is different from that of the real trend $f(t)$. Therefore the numerical values of the estimated trend are obtained either by translation $\widehat{F}_n = \widehat{F}\left((\kappa + n)\delta t\right)$, or by scaling $\widehat{F}_n = \widehat{F}(nT/N)$. The optimal estimated trend is given by the requirement that the difference $\{x_n - \widehat{F}_n\}$ should have a minimum standard deviation.

If the fluctuations described by ε_ξ causes the quantities $\widehat{g}_s$ to have different signs, then the estimated trend $\widehat{F}_n$ cannot be determined and the fluctuations are smoothed by means of a moving average. Therefore the numerical algorithm of trend removal consists in 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 $\widehat{F}_n^{(i)}$

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

If $\widehat{g}_s$ have not the same sign ($\widehat{F}_n^{(i)} = 0$ 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

$$\widehat{F}_n = \sum_i \widehat{F}_n^{(i)}$$

and the estimated random component is

$$\widehat{z}_n = x_n - \widehat{F}_n . \tag{10.7}$$

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 $\{\widehat{F}_n^{(i+1)}\}$ the estimated trend $\{\widehat{F}_n\}$ is not monotone.

For the interior values two-sided moving averages are performed over an interval of length $2K + 1$. If $n \leqslant K$ ($n > N - K$), 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 $\widehat{g}_s$ 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 algorithm described above is the possibility that some values $\widehat{g}_s^{(i)}$ 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 $\widehat{g}_s^{(i)}$ 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 $\left|\widehat{g}_s^{(i)}\right| \geqslant (\xi_{s+1} - \xi_s)/\Delta n_s$.

10.2 The numerical algorithm for trend estimation

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 of [61], a workable estimation of the magnitude order of σ_Z is given by

$$\sigma_Z^{\text{est}} = \frac{1}{\sqrt{2}} \widehat{\sigma}(\delta_{m_0} x_n) , \tag{10.8}$$

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

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

where $\| \cdot \|$ is the usual quadratic norm.

In order to apply the method for trend estimation 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); 2. distribute the signal values into disjoint intervals I_s. These tasks can be automatically accomplished as is shown in [61]. First we have to establish the limits of the number S of intervals I_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.

The optimum value of the parameter ρ is found by means of the estimation (10.8) 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 $\sqrt{N_r}$ 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 new 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

$$\rho_{\text{opt}} = \frac{\sqrt{N}\widehat{\sigma}_x}{\sigma_Z^{\text{est}}} , \tag{10.9}$$

where $\widehat{\sigma}_x$ is the sample standard deviation of the original signal. There are situations when Eq. (10.8) 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 (10.8) is applied to the residual obtained. The trend removal is performed with $S = S_{\max}$ homogeneous intervals and $K_f = 0.1N$. If the trend removal is not possible, then we apply Eq. (10.8) 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. (10.9) can be calculated.

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. (10.6). Therefore we distribute the signals values into a number of disjoint intervals inversely proportional with the estimated standard deviation of the noise σ_Z^{est} given by Eq. (10.8)

$$S^{\text{est}} = (x_{\max} - x_{\min})/\sigma_Z^{\text{est}} , \tag{10.10}$$

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. (10.10) is smaller than $S_{\min}$ (larger than $S_{\max} = [N/N_{\min}]$), then we impose $S^{\text{est}} = S_{\min}$ ($S^{\text{est}} = S_{\max}$). 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 algorithm. The signals dominated by noise or trend impose different numerical restrictions and it is advisable that the numerical algorithm should adapt itself to the signal type. We introduce a new parameter

$$\eta^{\text{est}} = \left| \frac{\widehat{\sigma}_x}{\sigma_Z^{\text{est}}} - 1 \right| ,$$

where $\widehat{\sigma}_x$ is the sample standard deviation of the signal and σ_Z^{est} is given by estimation (10.8). If $\eta^{\text{est}} < 1$, then $|\widehat{\sigma}_x - \sigma_Z^{\text{est}}| < \sigma_Z^{\text{est}}$ and the noise standard deviation represents the largest part of the whole signal, i.e., the signal is dominated by noise. Conversely, if $\eta^{\text{est}} > 1$, then the signal is dominated by trend.

Equation (10.10) applied to a signal dominated by noise ($\eta^{\text{est}} < 1$) 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 $\eta^{\text{est}} < 1$, we split into two subintervals those intervals which by splitting generate larger number of values than the quantity $N_{\min}$. 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 $N'_s \geqslant N_{\min}$ and $N'_{s+1} \geqslant N_{\min}$, then the splitting is validated and the number of intervals I_s increases by one. If all S^{est} intervals can be split generating $2S^{\text{est}}$ new intervals, then the splitting process is repeated until at least one of the intervals does not satisfy any more the condition that the two subintervals should contain more than $N_{\min}$ values. In the end we obtain S^* nonhomogeneous intervals I_s containing arbitrary numbers of time series values.

Applied to a signal dominated by trend ($\eta^{\text{est}} > 1$), Eq. (10.10) 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. (10.6). Therefore we merge the intervals with smaller length than σ_Z^{est}, $\xi_{s+1} - \xi_s < \sigma_Z^{\text{est}}$. 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 $\xi_{s+m+1} - \xi_s < \sigma_Z^{\text{est}}$ and $\xi_{s+m+2} - \xi_s > \sigma_Z^{\text{est}}$. Thus we obtain S^* nonhomogeneous intervals.

10.3 The trend of the paleoclimatic temperature

The automatic algorithm presented above estimates only monotone trend. To illustrate how it works we apply it on a real time series from paleoclimatology. Figure 10.1(a) shows the global mean annual temperature anomalies during the period A.D. 200-1995 with respect to the Northern Hemisphere mean annual temperature over 1856-1980 discussed in [29].

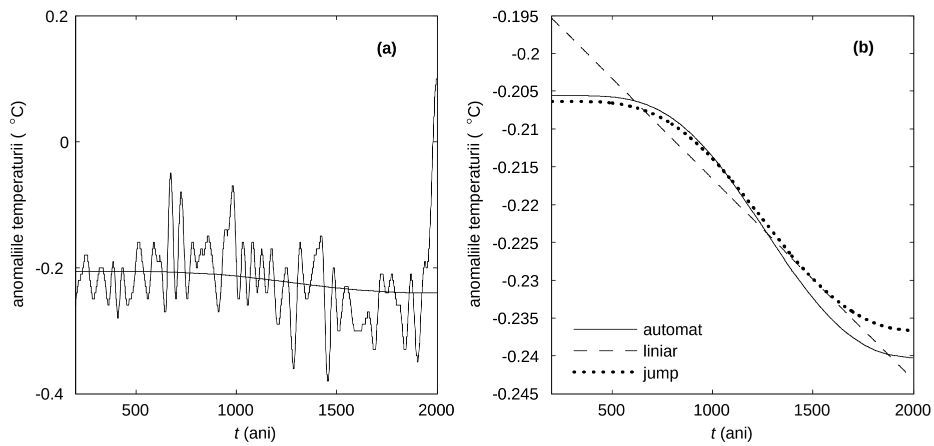

Figure 10.1: Global mean annual temperature anomalies during the period A.D. 200-1995 and the trend estimated by means of three algorithms.

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 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 automatic 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 $S^{\text{est}} = 9$ homogeneous intervals. The estimated standard deviation of the noise is $\sigma_Z^{\text{est}} = 0.056$ in comparison with the standard deviation of the initial series $\widehat{\sigma}_x = 0.063$.

In Fig. 10.1(b) the estimated trend is compared with those estimated by the other two methods. The linear trend, the only 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. The second comparison method is the "jump process" in [73] and it consists in an iterative weighted averaging

$$x_n^{(i+1)} = x_n^{(i)} + R(x_{n-1}^{(i)} - 2x_n^{(i)} + x_{n+1}^{(i)})\,, \tag{10.11}$$

where ratio R $(0 < R < 0.5)$ and iteration parameter M $(i \leqslant M)$ 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. (10.11) 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. The time series must be averaged until all the nonmonotone variations of the estimated trend are eliminated. For $R = 0.4$, $M = 329637$ averagings were needed such that the computing time became prohibitive. Besides that, even if the estimated trend has the same shape as the trend, due to the large number of averagings, the magnitude of the global temperature variation is underestimated.

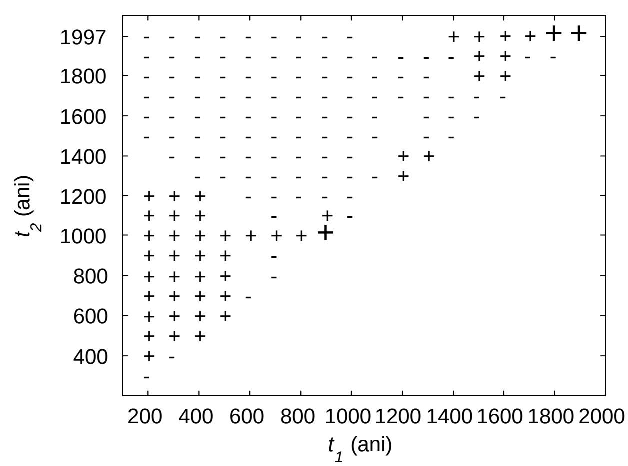

Figure 10.2: Time periods $[t_1, t_2]$ over which the global average temperature has an increasing (+ sign) or decreasing (- sign) trend. The three bigger signs correspond to significant trends with a confidence level of 99%.

As shown in [2], it is possible that, when the noise is large, the trend removal should not be recommended. The series in Fig. 10.1(a) is such a case. In order to verify that the estimated trend $\{\widehat{F}_n\}$ is significant, we build surrogate series $\widehat{x}_n = \widehat{F}_n + \widehat{\zeta}_n$. Since the serial dependence of the noise has a significant influence on the estimated trend quality, we choose the surrogate series $\{\widehat{\zeta}_n\}$ as realizations of an AR(1) stochastic process defined by the recurrence relation $Z_n = \phi Z_{n-1} + G_n$, where $0 \leqslant \phi < 1$ and $\{G_n\}$ is a Gaussian white noise with zero mean and standard deviation σ_G. The properties of the AR(1) process are well-known [10]. The probability distribution is Gaussian with zero mean and standard deviation $\sigma_Z^2 = \sigma_G^2(1-\phi^2)^{-1}$. If the parameter ϕ increases, the serial correlation of the noise increases too. 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. (10.7) we compute the value of the parameter ϕ used in surrogates generation. The fraction of the estimated trends for the surrogate series $\{\widehat{x}_n\}$ having the same sign of the global variation as the estimated trend $\{\widehat{F}_n\}$ 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. 10.1(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. 10.2. 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 two centuries and the third one corresponds to the 10th century. For 19 time periods the algorithm did not succeed to associate a monotone trend to the temperature variation. The rest of the estimated trends can not be considered “significant”, conclusion which coincides with that in [29].