# On Nonparametric Tests for Trend Detection in Seasonal Time Series

Oliver Morell and Roland Fried

**Abstract** We investigate nonparametric tests for identifying monotone trends in time series as they need weaker assumptions than parametric tests and are more flexible concerning the structure of the trend function. As seasonal effects can falsify the test results, modifications have been suggested which can handle also seasonal data. Diersen and Trenkler [5] propose a test procedure based on records and Hirsch et. al [8] develop a test based on Kendall's test for correlation. The same ideas can be applied to other nonparametric procedures for trend detection. All these procedures assume the observations to be independent. This assumption is often not fulfilled in time series analysis. We use the mentioned test procedures to analyse the time series of the temperature and the rainfall observed in Potsdam (Germany) from 1893 to 2008. As opposed to the rainfall time series, the temperature data show positive autocorrelation. Thus it is also of interest, how the several test procedures behave in case of autocorrelated data.

## 1 Introduction

One interest in time series analysis is to detect monotonic trends in the data. Several parametric and nonparametric procedures for trend detection based on significance tests have been suggested. Parametric methods rely on strong assumptions for the distribution of the data, which are difficult to check in practice and possibly not fulfilled. Furthermore a parametric form of the trend has to be specified, where only some unknown parameters need to be estimated. Nonparametric test procedures are more flexible as they afford only rather general assumptions about the distribution. Also the trend often only needs to be monotonic without further specifications.

---

Oliver Morell (✉)
Fakultät Statistik, Technische Universität Dortmund, D-44221 Dortmund, Germany
morell@statistik.tu-dortmund.de

First ideas for nonparametric test procedures based on signs (see e.g. [3] or [13]), ranks (see e.g. [4] or [12]) and records [7] have been developed early. However, all these approaches need the assumption of i.i.d. random variables under the null hypothesis. For time series with seasonal behavior this assumption is not valid. One way to handle this problem is to estimate and subtract the seasonality. Another approach is to use tests which are robust against seasonal effects. Hirsch et. al. [8] develop a test procedure based on Kendall's test of correlation [10]. Diersen and Trenkler [5] propose several tests based on records. They show that splitting the time series increases the power of the record tests, especially when seasonal effects occur. The procedures of Hirsch et. al. and Diersen and Trenkler use the independence of all observations to calculate a statistic separately for each period and sum them to get a test statistic for a test against randomness. The same ideas can be used for the above mentioned tests based on signs or ranks.

We apply the procedures to two climate time series from a gauging station in Potsdam, Germany: mean temperature and total rainfall. Such climate time series often show seasonality with a period of one year. Section 2 introduces the test problem of the hypothesis of randomness against a monotonic trend as well as test procedures which can also be used for seasonal data, namely some tests based on records for the splitted time series [5] and the seasonal Kendall–Test [8]. We also modify other nonparametric test statistics to consider seasonality. The mentioned sign– and rank–tests are transformed to new seasonal nonparametric tests. In Sect. 3 we compare the power of the several test procedures against different types of monotone trends and in the case of autocorrelation. In Sect. 4 the two climate time series are analysed. In particular, the test procedures are used to check the hypothesis of randomness. Section 5 summarizes the results.

## 2 Nonparametric Tests of the Hypothesis of Randomness

A common assumption of statistical analysis is the hypothesis of randomness. It means that some observed values $x_1, \dots, x_n$ are a realisation of independent and identically distributed (i.i.d.) continuous random variables (rv) $X_1, \dots, X_n$, all with the same cumulative distribution function (cdf) $F$. There are several test procedures which can be used to test the hypothesis of randomness $\mathrm{H}_0$ against the alternative $\mathrm{H}_1$ of a monotonic trend. However, in time series analysis the observed values $x_1, \dots, x_n$ are a realisation of a stochastic process and can be autocorrelated, implying a lack of independence of $X_1, \dots, X_n$. Additionally, many time series show seasonal effects and so $X_1, \dots, X_n$ are not identically distributed, even if there is no monotonic trend. We modify the hypothesis of randomness for seasonal data to handle at least the second problem:

Firstly, if there is a cycle of $k$ periods, the random sample $\mathbf{X} = (X_1, \dots, X_n)$ is splitted into $k$ parts

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k) \text{ with } \mathbf{X}_j = (X_{1,j}, X_{2,j}, \dots, X_{n_j,j}) \text{ and } X_{i,j} = X_{k(i-1)+j} \quad (1)$$

for $j = 1,\dots,k$ and $i = 1,\dots,n_j$. $\mathbf{X}_j$ thus includes all $n_j$ observations of season $j$. Under the null hypothesis $H_0$ of no trend the continuous rv $X_1,\dots,X_n$ are still considered to be independent but only for each $j$ the rv's $X_{1,j},\dots,X_{n_j,j}$ are identically distributed with common cdf $F_j$. Under the alternative $H_1$ of a monotonic trend there are values $0 = a_{1,j} \le a_{2,j} \le \dots \le a_{n_j,j}$ with $a_{i,j} < a_{i+1,j}$ for at least one $i \in \{1,\dots,n_j-1\}$ and $j \in \{1,\dots,k\}$ such that $F_{i,j}(x) = F_j(x-a_{i,j})$ in case of an increasing and $F_{i,j}(x) = F_j(x+a_{i,j})$ in case of a decreasing trend. Under $H_0$ the hypothesis of randomness within each period is fulfilled. In the following we denote the test problem of the hypothesis of randomness for seasonal data against a monotone trend alternative with $\mathscr{H}_R$ and introduce test procedures for $\mathscr{H}_R$.

## 2.1 Tests Based on Record Statistics

Foster and Stuart [7] introduce a nonparametric test procedure for $\mathscr{H}_R$ based on the number of upper and lower records in the sequence $X_1,\dots,X_n$ and the reversed sequence $X_n,\dots,X_1$. A test procedure for $\mathscr{H}_R$ based on this approach which is robust against seasonality is introduced by Diersen and Trenkler [5]. A first application of their procedure is given in [6].

Using (1) we define upper and lower record statistics $U^o_{i,j}$, $L^o_{i,j}$, $U^r_{i,j}$ and $L^r_{i,j}$ of the original and the reversed sequence for all periods $j = 1,\dots,k$ at $i = 2,\dots,n_j$ as

$$U^o_{i,j} = \begin{cases} 1 \text{ , if} & X_{i,j} > \max\{X_{1,j}, X_{2,j},\dots,X_{i-1,j}\} \\ 0 \text{ otherwise} & \end{cases} \tag{2}$$

$$L^o_{i,j} = \begin{cases} 1 \text{ , if} & X_{i,j} < \min\{X_{1,j}, X_{2,j},\dots,X_{i-1,j}\} \\ 0 \text{ otherwise} & \end{cases} \tag{3}$$

$$U^r_{n_j-i+1,j} = \begin{cases} 1 \text{ , if} & X_{n_j-i+1,j} > \max\{X_{n_j-i+2,j}, X_{n_j-i+3,j},\dots,X_{n_j,j}\} \\ 0 \text{ otherwise} & \end{cases} \tag{4}$$

$$L^r_{n_j-i+1,j} = \begin{cases} 1 \text{ , if} & X_{n_j-i+1,j} < \min\{X_{n_j-i+2,j}, X_{n_j-i+3,j},\dots,X_{n_j,j}\} \\ 0 \text{ otherwise} & \end{cases} \tag{5}$$

with

$$U^o_{1,j} = L^o_{1,j} = U^r_{n_j,j} = L^r_{n_j,j} = 1 \tag{6}$$

as the first value of a sequence is always an upper and a lower record.

Under $H_0$ for a larger $i$ the probability of a record will get smaller. Therefore Diersen and Trenkler [5] recommend to use linear weights $w_i = i-1$ for a record at the $i-$th position of the original or reversed sequence. The sum of the weighted records of the original sequence

$$U^o = \sum_{j=1}^{k}\sum_{i=1}^{n_j} w_i U^o_{i,j} \text{ and } L^o = \sum_{j=1}^{k}\sum_{i=1}^{n_j} w_i L^o_{i,j} \ , \tag{7}$$

and the sum of the records of the reversed series

$$U^r = \sum_{j=1}^{k} \sum_{i=1}^{n_j} w_i U^r_{n_j-i+1,j} \text{ and } L^r = \sum_{j=1}^{k} \sum_{i=1}^{n_j} w_i L^r_{n_j-i+1,j} \tag{8}$$

can be used as test statistics for $\mathscr{H}_R$. They are sums of independent rv and all have the same distribution under $\mathrm{H}_0$. The expectations and variances are given by

$$\mathrm{E}(U^o) \;=\; \sum_{j=1}^{k} \sum_{i=1}^{n_j} \frac{w_i}{i} \text{ and } \mathrm{Var}(U^o) \;=\; \sum_{j=1}^{k} \sum_{i=1}^{n_j} w_i^2 \frac{i-1}{i^2} \tag{9}$$

and especially

$$\mathrm{E}(U^o) \;=\; k \sum_{i=1}^{n_1} \frac{i-1}{i} \text{ and } \mathrm{Var}(U^o) \;=\; k \sum_{i=1}^{n_1} \frac{(i-1)^3}{i^2} \tag{10}$$

if linear weights $w_i = i-1$ are used and all periods $j$ have the same number of observations $n_1$.

If an upward trend exists, $U^o$ and $L^r$ become large while $L^o$ and $U^r$ become small. The opposite is true, if a downward trend exists. These informations can be used to combine the sums in (8) and (9) and to use the statistics

$$T_1 = U^o - L^o,\; T_2 = U^o - U^r,\; T_3 = U^o + L^r,\; T_4 = U^o - U^r + L^r - L^o \tag{11}$$

for $\mathscr{H}_R$. Under $\mathrm{H}_0$ the distributions of $T_1$, $T_2$ and $T_3$ will not change, if $\widetilde{T}_1 = L^r - U^r$, $\widetilde{T}_2 = L^r - L^o$ and $\widetilde{T}_3 = U^r + L^o$, respectively, are taken instead of the sums given in (11). From these statistics, only

$$T_1 = U^o - L^o \;=\; \sum_{j=1}^{k} \sum_{i=1}^{n_j} w_i \left(U^o_{i,j} - L^o_{i,j}\right) \tag{12}$$

can be expressed as a sum of independent rv, because here records from the same sequence are combined. We have under $\mathrm{H}_0$

$$\mathrm{E}(T_1) \;=\; 0 \text{ and } \mathrm{Var}(T_1) \;=\; 2 \sum_{j=1}^{k} \sum_{i=1}^{n_j} \frac{w_i^2}{i}\;. \tag{13}$$

In contrast to $T_1$, in $T_2$, $T_3$ and $T_4$ we use records from the original sequence as well as from the reversed sequence. So the summands here are not independent. We get the expectations

$$\mathrm{E}(T_2) \;=\; \mathrm{E}(T_4) \;=\; 0 \text{ and } \mathrm{E}(T_3) \;=\; 2 \sum_{j=1}^{k} \sum_{i=1}^{n_j} \frac{w_i}{i}\;. \tag{14}$$

while the variances of $T_2$, $T_3$ and $T_4$ become unwieldly expressions and are given in [6] for the case $n_1 = \ldots = n_k$.

Diersen and Trenkler [6] recommend a splitting with large $k$ and small $n_j$, $j = 1,\dots,k$. The first reason for this are the asymptotic properties of the statistics in (11). With $X_1,\dots,X_n$ assumed to be independent and $n_1 = \dots = n_k$, the statistics $T_1$, $T_2$, $T_3$ and $T_4$ are the sum of $k$ i.i.d. rv. So for $k \to \infty$ all four test statistics are asymptotically normal distributed. These asymptotics are not fulfilled, if the statistics in (11) are only weighted but not splitted. Diersen and Trenkler [5] showed for this case that the asymptotic distribution is not a normal one. The second reason is that compared to the best parametric test in the normal linear regression model and the (non seasonal) Kendall–Test the asymptotic relative efficiency vanishes for fixed $k$ and increasing $n_j$. So it is also an interesting question if the efficiency of other nonparametric tests can be increased, if the time series is splitted with a large $k$ and a small number $n_j$ of observations in each period $j$.

## 2.2 The Seasonal Kendall-test

Mann [12] introduced a test for $\mathscr{H}_R$ based on Kendall's test for independence of two random variables in a bivariate distribution [10]. It was modified by Hirsch et al. [8] to robustify the test statistic against seasonal effects. Taking the splitted series in (1), they use the test statistic

$$S = \sum_{j=1}^{k} S_j \quad \text{with} \quad S_j = \sum_{i=1}^{n_j-1} \sum_{i'=i+1}^{n_j} \operatorname{sgn}(X_{i',j} - X_{i,j}) \tag{15}$$

for $\mathscr{H}_R$. So in $S_j$ the number of pairs $(X_{i,j}, X_{i',j})$ with $X_{i,j} < X_{i',j}$ is subtracted from the number of pairs $(X_{i,j}, X_{i',j})$ with $X_{i,j} > X_{i',j}$, $i < i'$, for period $j$. If there is a positive (negative) monotonic trend in period $j$, the statistic $S_j$ is expected to be large (small) while it will probably realise a value near 0 if there is no monotonic trend. If the same positive (negative) monotonic behavior can be observed for all periods, the statistic $S$ will also become large (small). $S$ will also take a value close to 0, if no monotonic trend exists.

The exact distribution of $S$ under $\mathrm{H}_0$ is symmetric with

$$\mathrm{E}(S) = \sum_{j=1}^{k} \mathrm{E}(S_j) = 0 \tag{16}$$

and if there are no identical values (ties) in the observations of any period $j$, the variance is given by

$$\mathrm{Var}(S) = \sum_{j=1}^{k} \mathrm{Var}(S_j) = \sum_{j=1}^{k} \frac{n_j(n_j-1)(2n_j+5)}{18} \tag{17}$$

as $S_1,\dots,S_k$ are independent. A pair of observations is called a tie of extend $\delta$, if $\delta$ observations of $x_1,\dots,x_n$ have the same value. If $X_1,\dots,X_n$ are continuous rv, the probability of a tie is zero, but for rounded values, ties can be observed. Let $n_{\delta,j}$ be the number of ties within $\mathbf{X}_j$ with extend $\delta$. Then the variance of $S$ becomes smaller:

$$\mathrm{Var}(S) \;=\; \sum_{j=1}^{k} \frac{\left(n_j(n_j-1)(2n_j+5) - \sum\limits_{\delta=1}^{n_j} n_{\delta,j}\delta(\delta-1)(2\delta+5)\right)}{18} \tag{18}$$

As every $S_j$ is asymptotically normally distributed for $n_j \to \infty$, the statistic $S$ as a finite sum of independent asymptotically normally distributed rv is asymptotically normal, too, if $n_j$ converges to infinity for each $j$. The exact distribution of $S$ under $\mathrm{H}_0$ (neglecting ties) can be determined by enumerating all permutations of $X_{1,j},\ldots,X_{n_j,j}$ for each $j$ and calculating the values of $S_j$ for every permutation of each $j$. The individual values and their frequencies can be easily calculated with Chap. 5 of [11]. According to the frequencies of the single values for each $S_j$, the distribution of $S$ can be obtained by reconsidering every possible combination of the values and multiplying the corresponding frequencies. However, for large $n$ calculating the exact distribution of $S$ is time consuming, so the normal approximation should be used whenever possible. Hirsch et al. [8] state that already for $k = 12$ and $n_j = 3$ the normal approximation of $S_j$ works well. They also claim that their test is robust against seasonality and departures from normality, but not robust against dependence. Hirsch and Slack [9] develop a test for $\mathscr{H}_R$, which performs better than $S$ if the data are autocorrelated. This test uses estimates of the covariances between two seasons based on Spearman's rank correlation coefficient. The estimated covariances are used to correct the variance of $S$ in the normal approximation.

## 2.3 Some Rank Statistics for $\mathscr{H}_R$

Aiyar et al. [1] compare the asymptotic relative efficiencies of many nonparametric tests for the hypothesis of randomness against trend alternatives. They consider mostly linear and nonlinear rank statistics, which we will use in the following for $\mathscr{H}_R$:

Taken the splitted series from (1) let $R(X_{1,j}),\ldots,R(X_{n_j,j})$ be the ranks of the continuous rv $X_{1,j},\ldots,X_{n_j,j}$, for $j \in \{1,\ldots,k\}$. Then two linear rank test statistics based on Spearman's rank correlation coefficient are given by

$$R_1 = \sum_{j=1}^{k} \widetilde{R}_{1,j} \text{ with } \widetilde{R}_{1,j} = \sum_{i=1}^{n_j} \left(i - \frac{n_j+1}{2}\right)\left(R(X_{i,j}) - \frac{n_j+1}{2}\right) \text{ and} \tag{19}$$

$$R_2 = \sum_{j=1}^{k} \widetilde{R}_{2,j} \text{ with } \widetilde{R}_{2,j} = \sum_{i=1}^{n_j} \left(i - \frac{n_j+1}{2}\right) \operatorname{sign}\left(R(X_{i,j}) - \frac{n_j+1}{2}\right) .$$

Both statistics are symmetric and have an expected value of 0. Their variances are given by

$$\mathrm{Var}(R_1) = \sum_{j=1}^{k} \mathrm{Var}(\widetilde{R}_{1,j}) = \sum_{j=1}^{k} \frac{n_j^2(n_j+1)^2(n_j-1)}{144} \quad \text{and}$$

$$\mathrm{Var}(R_2) = \sum_{j=1}^{k} \mathrm{Var}(\widetilde{R}_{2,j}) \text{ with } \mathrm{Var}(\widetilde{R}_{2,j}) = \begin{cases} \sum_{j=1}^{k} \frac{n_j^2(n_j+1)}{12} & , n_j \text{ even} \\ \sum_{j=1}^{k} \frac{n_j(n_j-1)(n_j+1)}{12} & , n_j \text{ odd} . \end{cases} \tag{20}$$

Instead of considering all rv like in (19), the $(1-2\gamma)$ truncated sample can be taken for all periods, with $\gamma \in (0, 0.5)$. Like [1] we define

$$c_{i,j} = \begin{cases} -1 , & 0 < i \leq \lfloor \gamma n_j \rfloor \\ 0 , & \lfloor \gamma n_j \rfloor < i \leq n_j - \lfloor \gamma n_j \rfloor \\ +1 , & n_j - \lfloor \gamma n_j \rfloor < i \leq n_j \end{cases} \tag{21}$$

so that the two statistics

$$\begin{aligned} R_3 &= \sum_{j=1}^{k} \widetilde{R}_{3,j} \text{ with} \\ \widetilde{R}_{3,j} &= \sum_{i=1}^{n_j} c_{i,j} \left( R(X_{i,j}) - \frac{n_j+1}{2} \right) = \sum_{j=1}^{k} \left( \sum_{i=n_j-\lfloor \gamma n_j \rfloor+1}^{n_j} R(X_{i,j}) - \sum_{i=1}^{\lfloor \gamma n_j \rfloor} R(X_{i,j}) \right) \text{ and} \\ R_4 &= \sum_{j=1}^{k} \widetilde{R}_{4,j} \text{ with} \\ \widetilde{R}_{4,j} &= \sum_{i=1}^{n_j} c_{i,j} \operatorname{sign} \left( R(X_{i,j}) - \frac{n_j+1}{2} \right) \\ &= \sum_{i=n_j-\lfloor \gamma n_j \rfloor+1}^{n_j} \operatorname{sign} \left( R(X_{i,j}) - \frac{n_j+1}{2} \right) - \sum_{i=1}^{\lfloor \gamma n_j \rfloor} \operatorname{sign} \left( R(X_{i,j}) - \frac{n_j+1}{2} \right) \end{aligned} \tag{22}$$

compare the sum of the most recent $\lfloor \gamma n_j \rfloor$ ranks (signs) with the sum of the first $\lfloor \gamma n_j \rfloor$ ranks (signs). Again the expectation of $R_3$ and $R_4$ is 0. Under the null hypothesis, the variances are given by

$$\begin{aligned} \mathrm{Var}(R_3) &= \sum_{j=1}^{k} \frac{n_j(n_j+1)\lfloor \gamma n_j \rfloor}{6} \quad \text{and} \\ \mathrm{Var}(R_4) &= \sum_{j=1}^{k} \mathrm{Var}(\widetilde{R}_{4,j}) \quad \text{with } \mathrm{Var}(\widetilde{R}_{4,j}) = \begin{cases} 2\frac{n_j}{n_j-1} \lfloor \gamma n_j \rfloor & , n_j \text{ even} \\ 2\lfloor \gamma n_j \rfloor & , n_j \text{ odd} . \end{cases} \end{aligned} \tag{23}$$

Again the above variances are only valid if all observations have different values. If ties occur, one possibility, which leads to a loss of power but keeps the variances from (20) and (23) under the null hypothesis is to give random ranks to tied observations. Alternatives like average ranks, which reduce the loss of power compared to random ranks, are not considered here.

In addition to this, [1] also consider nonlinear rank statistics. In analogy to them we define for each period $j$

$$I_{i,i',j} = \begin{cases} 1 \text{ , if } X_{i,j} < X_{i',j} \\ 0 \text{ , otherwise} \end{cases} , \tag{24}$$

$i, i' \in \{1, \ldots, n\}, i \neq i'$. Under the null hypothesis of randomness, we have

$$\mathrm{E}(I_{i,i',j}) = \frac{1}{2} \text{ and } \mathrm{Var}(I_{i,i',j}) = \frac{1}{4} . \tag{25}$$

Based on the sign difference test [13] we define for $\mathscr{H}_R$

$$N_1 = \sum_{j=1}^{k} \widetilde{N}_{1,j} \text{ with } \widetilde{N}_{1,j} = \sum_{i=2}^{n_j} I_{i-1,i,j} \tag{26}$$

which counts the number of pairs for each period $j$, where the consecutive observation has a larger value and then sums these pairs over all periods. For each $j$ we have $n_j - 1$ differences. Under $\mathrm{H}_0$ and from (25) we get

$$\mathrm{E}(N_1) = \sum_{j=1}^{k} \frac{1}{2}(n_j - 1) \text{ and } \mathrm{Var}(N_1) = \sum_{j=1}^{k} \frac{1}{12}(n_j + 1) . \tag{27}$$

For each $j$ the distribution of $\sum_{i=2}^{n_j} I_{i-1,i,j}$ converges to a normal distribution [13]. Therefore the distribution of $N_1$ converges to a normal distribution, too.

Another test for $\mathscr{H}_R$ based on Cox and Stuart [3] is given by

$$N_2 = \sum_{j=1}^{k} \widetilde{N}_{2,j} \text{ with } \widetilde{N}_{2,j} = \sum_{i=1}^{\lfloor n_j/2 \rfloor} (n_j - 2i + 1) I_{i,n_j-i+1,j} . \tag{28}$$

Cox and Stuart [3] show that $N_2$ leads to the best weighted sign test with respect to the efficiency of a sign test of $\mathscr{H}_R$. The linear rank test statistics $R_1$ and $R_2$ and the procedure $S$ of Kendall compare all pairs of observations, while in (28) each observation is taken only for one comparison. Using (25) we get under $\mathrm{H}_0$

$$\mathrm{E}(N_2) = \sum_{j=1}^{k} \mathrm{E}(\widetilde{N}_{2,j}) \quad \text{with} \quad \mathrm{E}(\widetilde{N}_{2,j}) = \begin{cases} \frac{n_j^2}{8} & , n_j \text{ even} \\ \frac{n_j^2 - 1}{8} & , n_j \text{ odd} \end{cases}$$

$$\text{and } \mathrm{Var}(N_2) = \sum_{j=1}^{k} \tfrac{1}{24} n_j (n_j^2 - 1) . \tag{29}$$

Cox and Stuart [3] also introduce a best unweighted sign test, which can be formulated for $\mathscr{H}_R$ as follows

$$N_3 = \sum_{j=1}^{k} \widetilde{N}_{3,j} \text{ with } \widetilde{N}_{3,j} = \sum_{i=1}^{\nu_j} I_{i,n_j-\nu_j+i,j} \; . \tag{30}$$

The value $\nu_j \leq \frac{1}{2}n_j$ is taken to compare observations further apart. We get

$$\mathrm{E}(N_3) = \sum_{j=1}^{k} \frac{\nu_j}{2} \text{ and } \mathrm{Var}(N_3) = \sum_{j=1}^{k} \frac{\nu_j}{4} \tag{31}$$

under $\mathrm{H}_0$. Cox and Stuart [3] recommend $\nu_j = \frac{1}{3}n_j$.

Again a splitting with small $n_1 = \ldots = n_k$ and large $k$ leeds asymptotically to a normal distribution for all introduced test statistics, as $k$ i.i.d. rv are added.

## 3 Comparison of the Nonparametric Tests for $\mathscr{H}_R$

Now we compare the different tests presented in Sect. 2 for different sample sizes and splitting factors and for various alternatives. We consider the time series model

$$X_{i,j} = a_{i,j} + E_{i,j} \;\; j = 1,\ldots,k, \; i = 1,\ldots,n_j, \tag{32}$$

where $E_{1,1},\ldots,E_{n_k,k}$ are Gaussian white noise with expected value 0 and constant variance $\sigma_E^2 = 1$. $X_{i,j}$ is the $i-$th observation for season $j$. For simplicity we fix the number of seasons to $k = 4$ and assume that each season has the same sample size $n_1$. Furthermore, the slopes are given by $a_{1,j} \leq \ldots \leq a_{n_1,j}$. We are interested in particular in three different kinds of monotone trends, with the same trend structure in each season. This means that for each $j$ we have the same slopes. With $a_{i,j} = i\theta$ we achieve a linear trend, where the parameter $\theta$ controls the slope of the straight line. We also consider a concave case with $a_{i,j} = \theta\sqrt{n_1 i}$, and a convex case with $a_{i,j} = \theta i^2/n_1$, so that all trends increase to $\theta n_1$. We consider sample sizes $n \in \{12,24,32,48,64,96,120\}$ and splittings into $\widetilde{k} \in \{1,4,8,12,16,24,32\}$ groups whenever $\widetilde{n}_1 = n/\widetilde{k}$ is an integer. We do not consider splittings with $\widetilde{n}_1 = 2$ as here $R_3$ and $R_4$ for $\gamma = \frac{1}{3}$ as well as $N_3$ with $\nu_1 = \ldots = \nu_{\widetilde{k}} = \frac{1}{3}$ are not defined. The other test statistics are equivalent in this case, as they all consider an unweighted ranking of two observations in each splitting. With $\widetilde{k} = 1$ the unsplitted case is also taken into account. In case of seasonal effects the power of all tests will probably be reduced if $\widetilde{k} = 1$ is chosen. We compare the power of the tests of Sect. 2 for all reasonable combinations of $\widetilde{k}$ and $n$ from above and take 1000 random samples from (32) for each combination. We use the asymptotic versions of the tests at a significance level of $\alpha = 0.05$. The percentage cases of rejections of $\mathrm{H}_0$ estimate the power of the several test procedures. Here we only consider the case of an upward trend, i.e. $\theta > 0$. We consider the linear, the convex and the concave case from above and calculate the power of all tests for $\theta \in \{0.01, 0.02, \ldots, 0.49, 0.50\}$. To achieve monotone power functions, we use the R–function `isotone` from the R–package `EbayesThresh` for monotone least squares regression to smooth the simulated power curves ([14, 15]).

Firstly we compare the weighted record statistics. For $n \geq 64$ all power functions take values close to 1, independently of the splitting factor $\widetilde{k}$, if a linear trend with $\theta > 0.1$ exists. In the concave case only $U^o$ and $T_2$ with $\widetilde{k} = 1$ perform worse for $n = 64$. An explanation for this is the strength of the slope. A positive concave trend increases less towards the end of the time series. Hence there will be fewer records at the end of the time series and $U^o$ will perform worse than $L^r$. As our version of $T_2$ also uses $U^o$ we receive similar results for this test statistic. In the convex case similar results can be obtained for $L^r$ as a convex upward trend of the original sequence means a concave downward trend of the negative reversed series. The power functions of the record tests for $\widetilde{k} = 1$ and $\widetilde{k} = 4$ can be seen in Fig. 1 for the linear, the concave and the convex case. Looking also at other sample sizes $n$ in the linear case (see Fig. 2), we find that $T_3$ performs best among the record tests in most of the cases. Generally, the power of the record tests gets larger in the above situations, if a larger $\widetilde{k}$ is chosen. Only $T_3$ performs better for a medium value of $\widetilde{k}$, e.g. $\widetilde{k} = 4$ for $n = 32$ or $\widetilde{k} = 12$ for $n = 96$. The previous findings are confirmed in the case of a convex or concave trend.

In Fig. 3 the power functions of the rank tests are shown, when different $\widetilde{k}$ for a fixed $n = 64$ are used. We show the concave case here, because the differences are qualitatively the same, but slightly bigger than for the linear or the convex trend. The seasonal Kendall–Test $S$ and Spearman–Test $R_1$ perform best, when a small $\widetilde{k}$ is used. Conclusions about an optimal splitting for the other rank tests are hard to state. If $\widetilde{k}$ is large compared to $n$, the power of the tests is reduced for most of the situations. However, generally we observe for all these tests (except $N_1$) good results, if $\widetilde{k} = 4$ is chosen. $N_1$ performs worse than the other tests in most situations even though it is the only test statistic with an increasing power in case of a larger splitting factor $\widetilde{k}$. From the rank tests $S$ and $R_1$ achieve the largest power in most situations. Comparing the best rank tests $S$ and $R_1$ with $\widetilde{k} = 4$ and the best record tests $T_3$ and $T_4$ with a large splitting factor $\widetilde{k} = 4$, $S$ and $R_1$ have a larger power in every situation.

Next we consider a situation with autocorrelated data. Here the hypothesis of randomness is not fulfilled, but no monotone trend exists. It is interesting which test procedures are sensitive to autocorrelation in the sense that they reject $H_0$ even though there is no monotone trend. We consider an autoregressive process of first order (AR(1))

$$E_t = \rho E_{t-1} + \varepsilon_t \ , \ t = 1, \ldots, n \ , \tag{33}$$

with autocorrelation coefficient $\rho$, i.e. we assume the sequence $E_1, \ldots, E_n$ to be autocorrelated with correlation $\rho$ and hence the autocorrelation within $E_{1,j}, \ldots, E_{n_1,j}$ with $E_{i,j} = E_{k(i-1)+j}$ is smaller than $\rho$. The innovations $\varepsilon_{1,j}, \ldots, \varepsilon_{n_1,j}$ are i.i.d. normally distributed random variables with expectation 0 and variance $\sigma_\varepsilon^2$, where

$$\sigma_\varepsilon^2 = (1 - \rho^2)\sigma_E^2 = (1 - \rho^2) \tag{34}$$

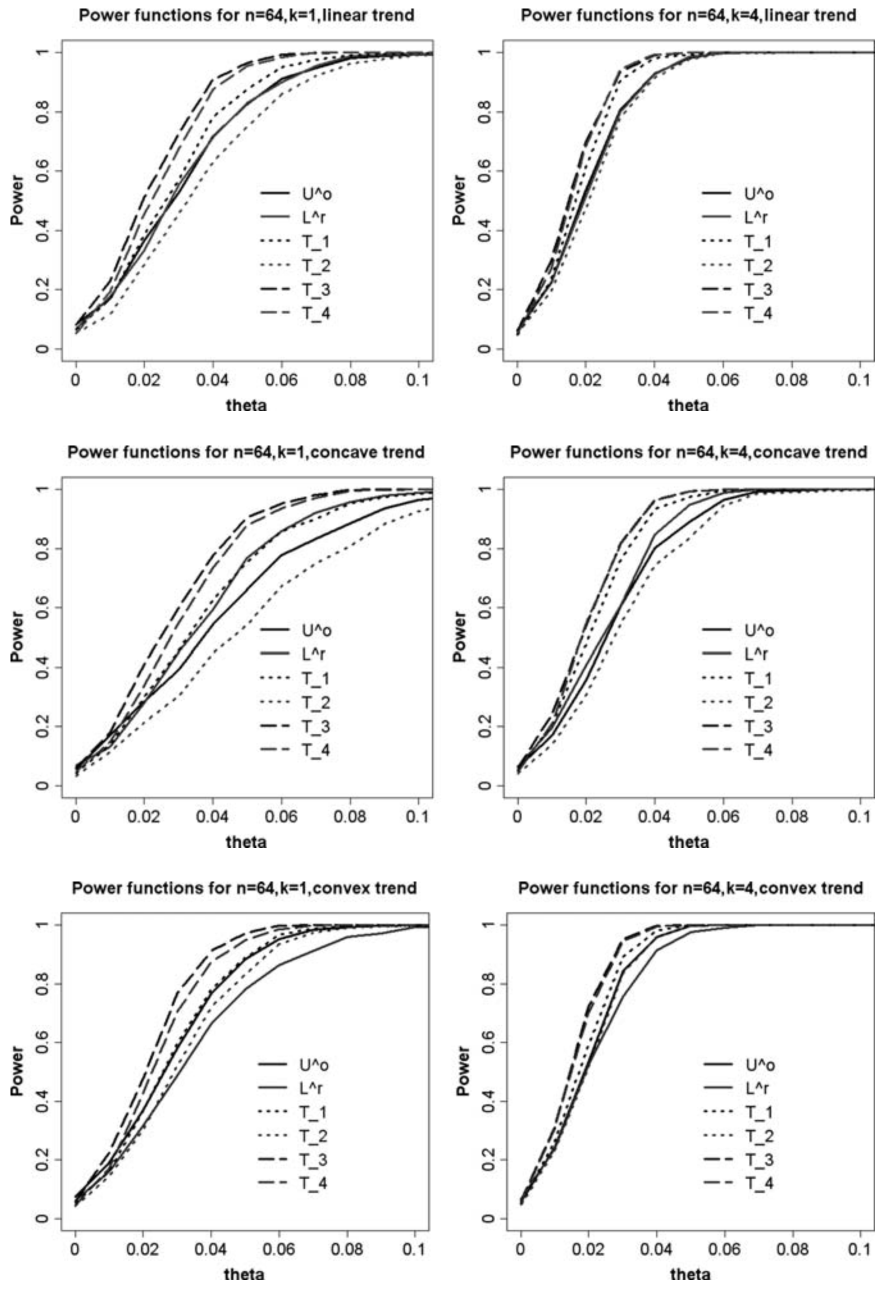


**Fig. 1** Power functions of the record tests for $n = 64$, small $\theta$ and $\widetilde{k} = 1$ (*left*) and $\widetilde{k} = 4$ (*right*)

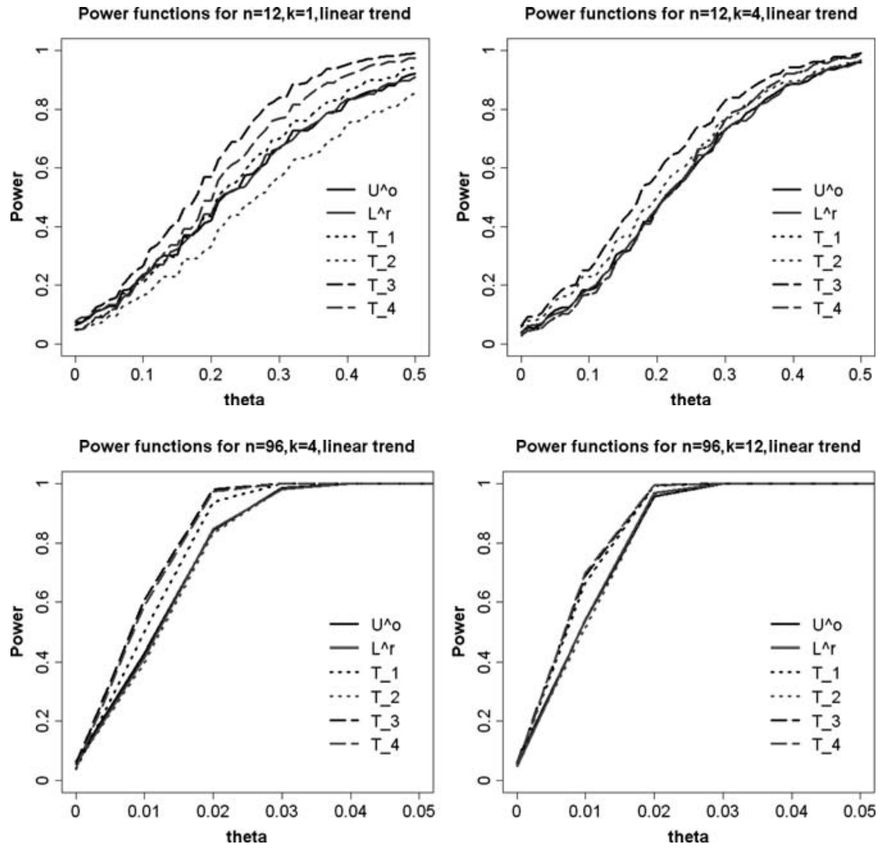


**Fig. 2** Power functions of the record tests for $n = 12$ (*top*) and $n = 96$ (*bottom*) for different $\widetilde{k}$

as we want to keep $\sigma_E^2$ equal to 1 again. We vary $\rho$ in $\{0.025, 0.05, \ldots, 0.875, 0.9\}$. The resulting detection rates of the record tests can be seen in Fig. 4 for $n = 96$ and different values of $\widetilde{k}$. $T_3$ is more sensitive to positive autocorrelation than $T_1$, $T_2$ and $T_4$ if a small $\widetilde{k}$ is used, but this difference vanishes for a large $\widetilde{k}$. The better performance of $T_1$, $T_2$ and $T_4$ for small $\widetilde{k}$ can be explained by the fact that they subtract statistics which become large in case of monotonically decreasing sequences from statistics which become large in case of monotonically increasing sequences. Positive autocorrelations cause both patterns to occur so that the effects cancel out.

For the rank tests we get the following findings: $N_2$ becomes robust against autocorrelations $\rho \leq 0.6$ for larger sample sizes $n \geq 48$, if we choose $\widetilde{k}$ so that we have three observations in each split. We observe for the pairs $n = 48$, $\widetilde{k} = 16$ and $n = 96$, $\widetilde{k} = 32$ for most of the values of $\rho$ a power of less than $\alpha = 0.05$. If we choose a splitting factor leading to $n_1 > 3$ this robustness is lost (see Fig. 5). $N_1$ behaves the most insensitive against autocorrelation for a large $\widetilde{k}$, but $N_1$ was also the test with the smallest power if a trend exists. For the other tests we have for

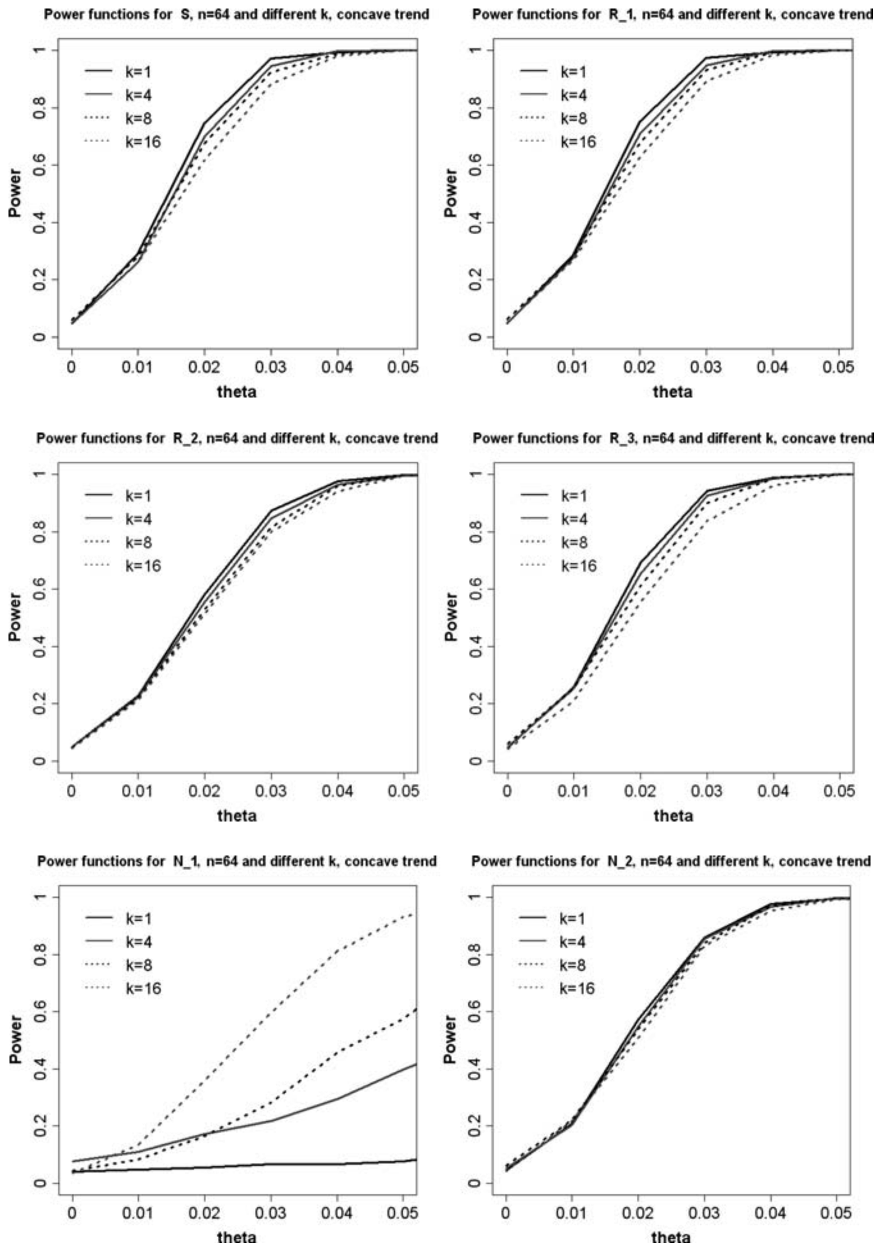


**Fig. 3** Power functions of the rank tests for different $\widetilde{k}$ with $n = 64$ and a concave trend

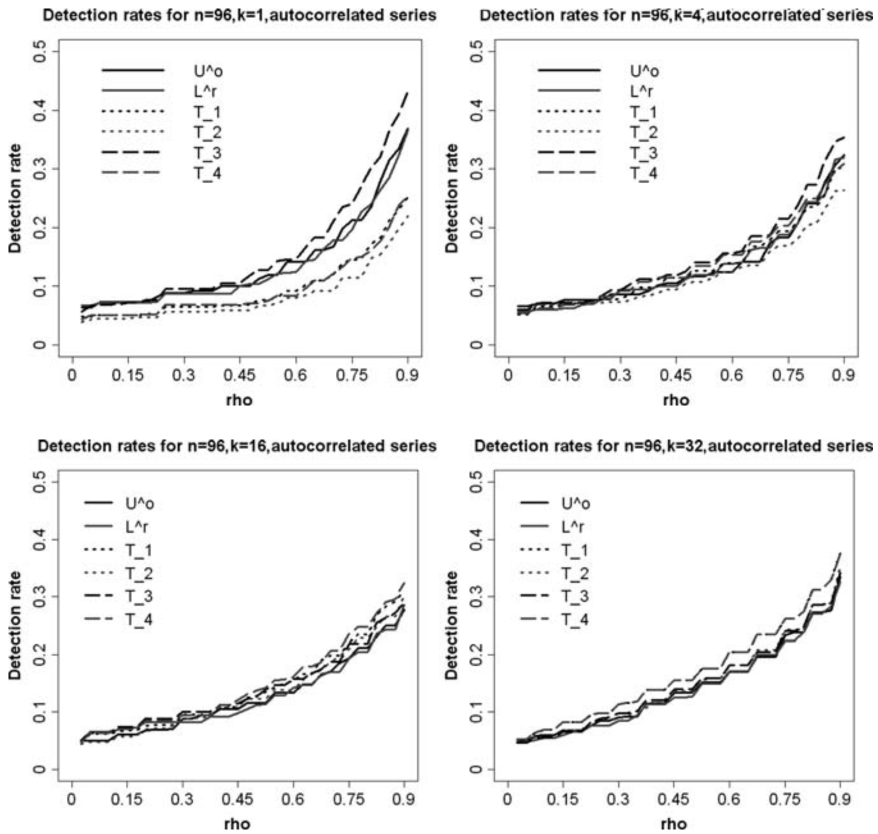


**Fig. 4** Detection rates of the record tests for $n = 96$ and different $\widetilde{k}$ with autocorrelation

a fixed $n$ a higher detection rate, when a smaller splitting factor $\widetilde{k}$ is used. If we compare the record tests with the rank tests, we find that $T_3$ reacts less sensitive to autocorrelation than the rank tests in most situations.

## 4 Analysis of the Climate Time Series from Potsdam

Now the methods from Sect. 2 are applied to some real time series data. The two series analysed here consist of the monthly observations of the mean air temperature and the total rainfall in Potsdam between January 1893 and April 2008. There are no missing values. The secular station in Potsdam is the only meteorological station in Germany for which daily data have been collected during a period of over 100 years without missings. The measures are homogeneous, what is due to the facts that the

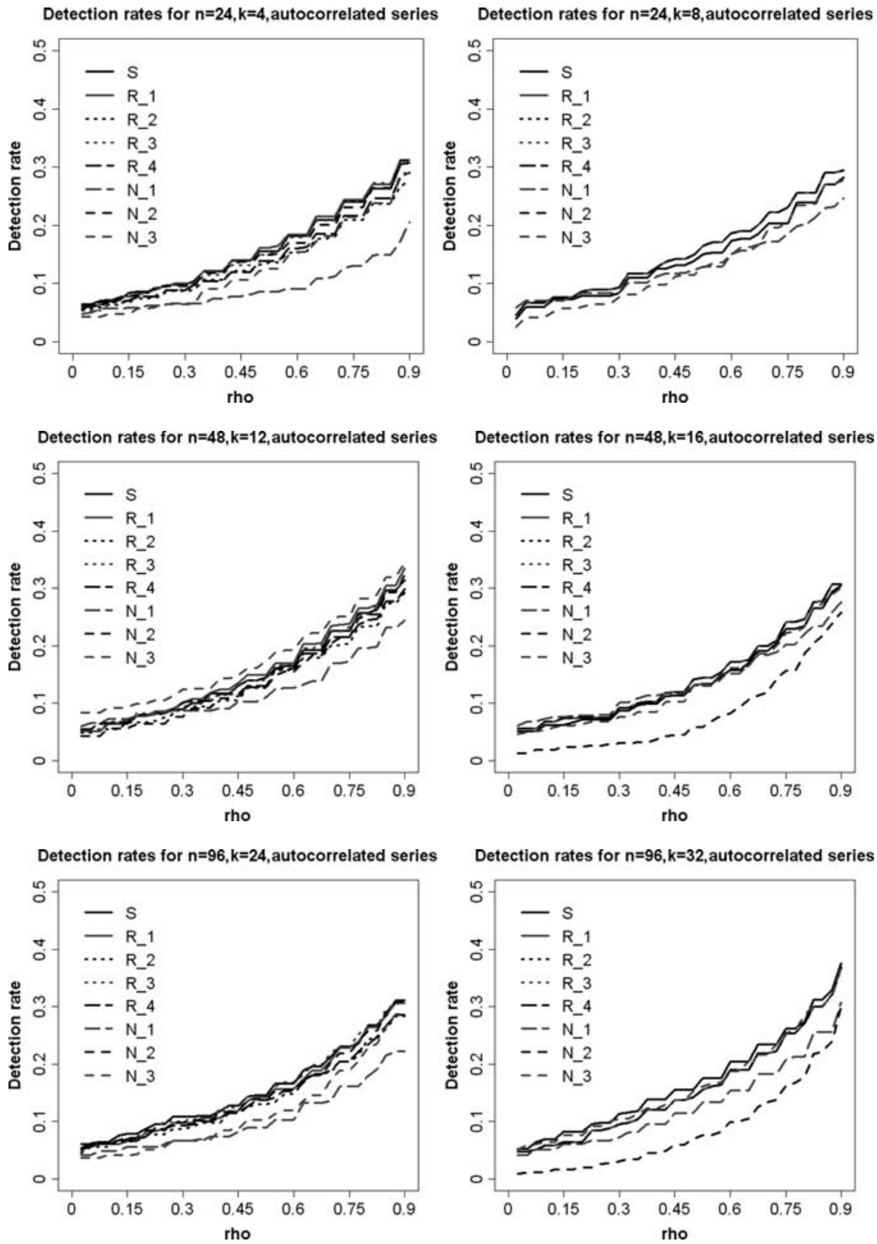


**Fig. 5** Detection rates of the rank tests with $n_1 = 4$ (*top*) and $n_1 = 3$ (*bottom*) observations in each splitting with autocorrelation

station has never changed its position, the measuring field stayed identical and the sort of methods, prescriptions and instruments, which are used for the measuring, have been kept.

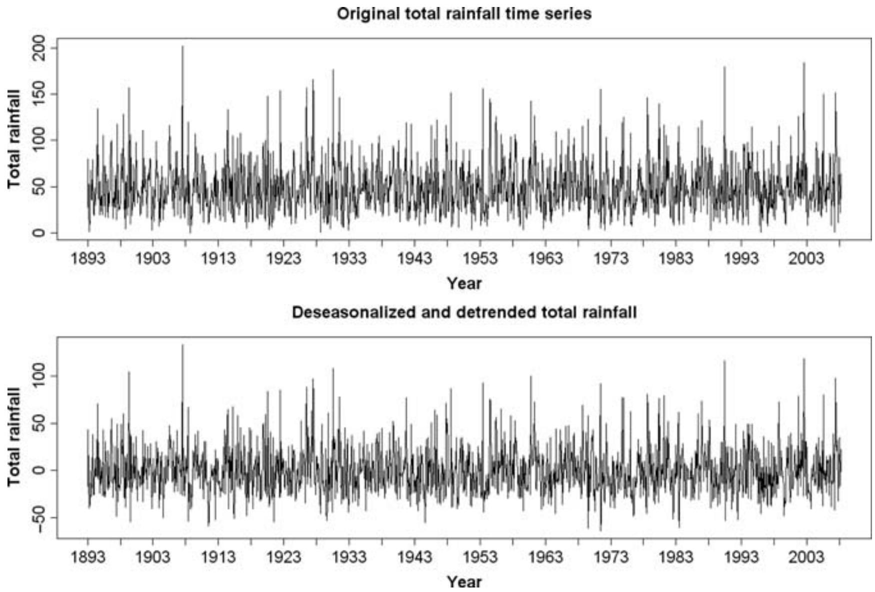


**Fig. 6** Original (*top*) and detrended and deseasonalized (*bottom*) total rainfall time series

Before the methods from Sect. 2 can be applied, we have to check if the assumptions are fulfilled. Independence of the observations can be checked with the autocorrelation function (ACF) and the partial autocorrelation function (PACF). Before this we detrend the time series by subtracting a linear trend. We also deseasonalize the time series by estimating and subtracting a seasonal effect for each month. The original and the detrended deseasonalized time series can be found in Fig. 6 for the total rainfall and in Fig. 7 for the mean temperature. The autocorrelation functions of the detrended and deseasonalized time series show positive autocorrelations at small time lags in case of the temperature and no correlation in case of the rainfall (see Fig. 8). In the former case, a first order autoregressive model with a moderately large AR(1) coefficient gives a possible description of the correlations. We use the test statistics from Sect. 2 to test the hypothesis of randomness against the alternative of an upward trend in both time series.

We consider all test statistics except $L^o$ and $U^r$ as these tests are only useful to detect a downward trend. As we have in both time series monthly observations for more than 115 years, we choose the splitting factor $\widetilde{k}$ as multiples of 12, more precisely $\widetilde{k} \in \{12, 24, 60, 120, 240, 360\}$. This guarantees that even $R_3$, $R_4$ (with $\gamma = \frac{1}{3}$) and $N_3$ (with $\nu_j = \frac{1}{3}n_j$) can be computed for each split. For every test procedure we use the asymptotic critical values, which seems to be reasonable for the above $\widetilde{k}$. The resulting p–values can be seen in Table 1 for the total rainfall time series and in Table 2 for the mean temperature.

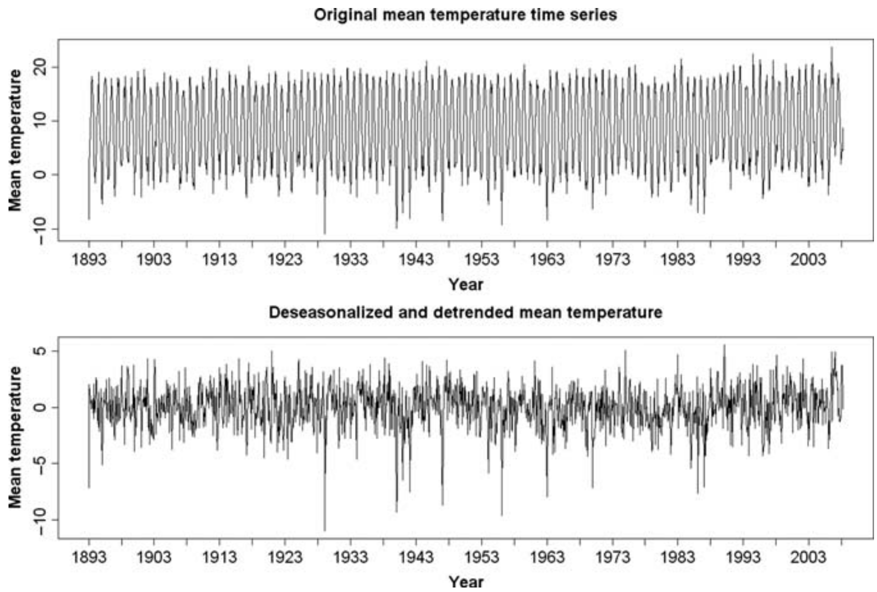


**Fig. 7** Original (*top*) and detrended and deseasonalized (*bottom*) mean temperature time series

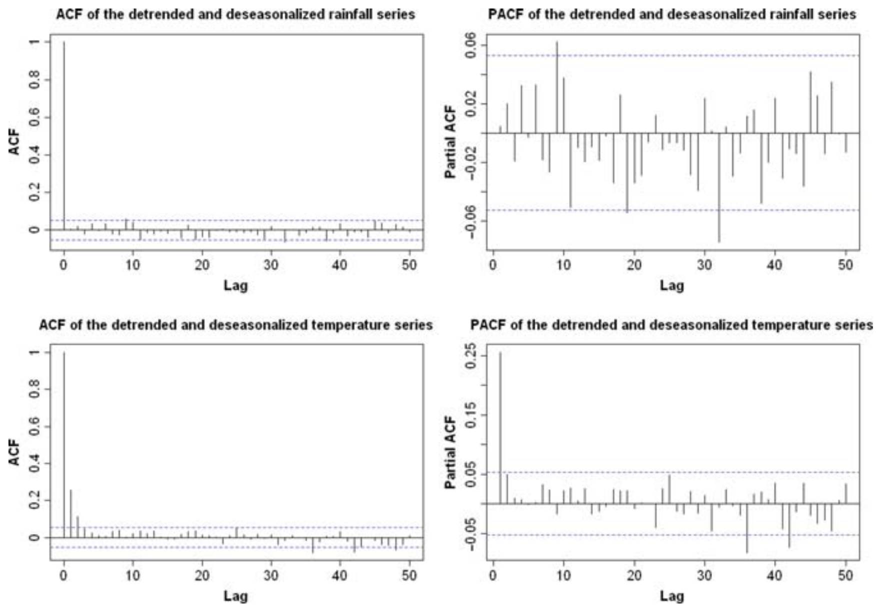


**Fig. 8** Autocorrelation (*left*) and partial autocorrelation function (*right*) of the detrended and deseasonalized rainfall (*top*) and temperature time series (*bottom*)

**Table 1** p–values for the total rainfall time series (in percent)

| $\widetilde{k}$ | 12 | 24 | 60 | 120 | 240 | 360 |
|---|---|---|---|---|---|---|
| $U^o$ | 6.4 | 40.9 | 11.7 | 18.3 | 11.1 | 6.1 |
| $L^r$ | 9.3 | 21.3 | 32.4 | 26.8 | 38.7 | 7.9 |
| $T_1$ | 4.2 | 34.9 | 14.2 | 7.9 | 14.8 | 9.8 |
| $T_2$ | 4.3 | 31.8 | 3.3 | 11.9 | 12.8 | 7.4 |
| $T_3$ | 2.3 | 23.7 | 12.9 | 15.7 | 17.8 | 4.6 |
| $T_4$ | 1.9 | 22.5 | 6.0 | 7.8 | 17.6 | 7.5 |
| $S$ | 17.2 | 12.8 | 28.1 | 25.6 | 24.1 | 9.1 |
| $R_1$ | 19.4 | 15.7 | 33.2 | 39.2 | 37.5 | 13.0 |
| $R_2$ | 26.7 | 19.2 | 36.3 | 42.2 | 33.1 | 26.5 |
| $R_3$ | 44.0 | 38.6 | 57.0 | 58.9 | 45.5 | 11.1 |
| $R_4$ | 48.7 | 44.8 | 63.4 | 61.8 | 41.2 | 20.5 |
| $N_1$ | 8.2 | 35.6 | 32.4 | 18.6 | 5.1 | 5.8 |
| $N_2$ | 4.6 | 5.1 | 58.4 | 61.7 | 49.1 | 20.0 |
| $N_3$ | 61.1 | 61.1 | 46.1 | 46.1 | 46.1 | 14.6 |

**Table 2** p–values for the mean temperature time series (in percent)

| $\widetilde{k}$ | 12 | 24 | 60 | 120 | 240 | 360 |
|---|---|---|---|---|---|---|
| $U^o$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $L^r$ | 0.00 | 0.03 | 0.01 | 0.00 | 0.00 | 0.00 |
| $T_1$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $T_2$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $T_3$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $T_4$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $S$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $R_1$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $R_2$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $R_3$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $R_4$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $N_1$ | 97.42 | 13.40 | 5.04 | 21.07 | 0.05 | 0.06 |
| $N_2$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $N_3$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

For the total rainfall time series the record tests $T_1$, $T_2$, $T_3$ and $T_4$ with $\widetilde{k} = 12$ detect a monotone trend at a significance level of $\alpha = 0.05$. From the rank tests only $N_2$ finds a monotone trend at this $\alpha$. Using a larger splitting factor we only find a monotone trend with $T_2$ for $\widetilde{k} = 60$. Of course we need to keep in mind that we perform multiple testing and thus expect about four significant test statistics among the more than 80 tests performed here even if there is no trend at all.

All tests except $N_1$ detect a monotone trend in the temperature time series for all splittings $\widetilde{k}$. The statistic $N_1$ only detects a monotone trend, if $\widetilde{k}$ is large. But as all tests need the assumption of independence, the results of Table 2 can not be interpreted as p–values of unbiased tests. This is why we deseasonalize the temperature time series and fit an AR(1)–Model to the deseasonalized series by maximum likelihood. If the data generating mechanism is an AR(1) process with uncorrelated innovations, then the residuals of the fitted AR(1) model are asymptotically uncorrelated. The residuals are even asymptotically independent, if the innovations are i.i.d. The residuals are asymptotically normal, if the innovations are normally distributed (see Section 5.3 of [2]). Looking at the plot of the scaled residual time series in Fig. 9 and its ACF in Fig. 10, we do not find significant autocorrelations between the residuals. However, the residuals do not seem to be identically normally distributed, as we can find some outliers in the residual plot. Table 3 shows the p–values of the record and rank tests for the residuals. We find mostly larger p–values than in Table 2, but again all tests except $N_1$ detect a positive monotone trend at $\alpha = 0.05$, what confirms the previous findings.

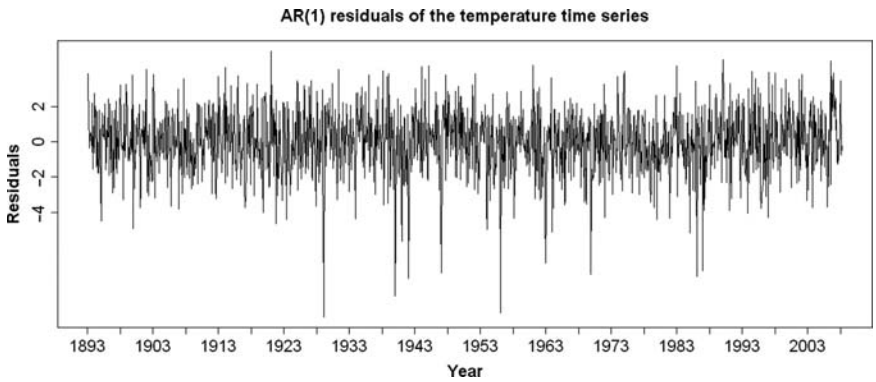


**Fig. 9** Residuals of the temperature time series obtained from fitting an AR(1) model to the deseasonalized temperature time series

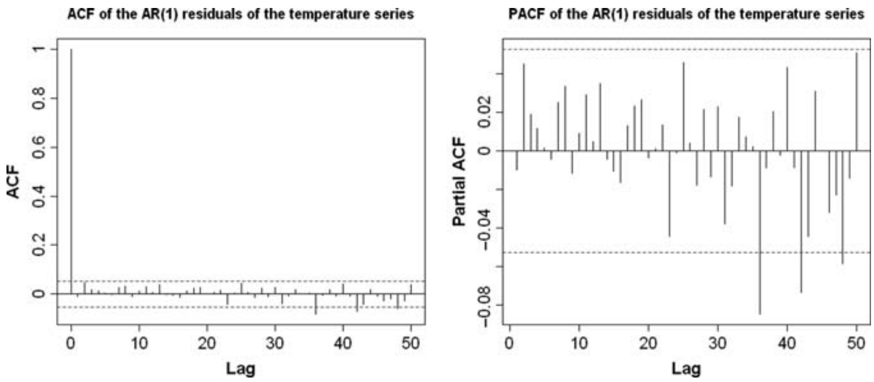


**Fig. 10** ACF (*left*) and PACF (*right*) of the AR(1) residuals of the deseasonalized temperature series

**Table 3** p–values for the residual temperature time series (in percent)

| $\widetilde{k}$ | 12 | 24 | 60 | 120 | 240 | 360 |
|---|---|---|---|---|---|---|
| $U^o$ | 0.30 | 0.19 | 0.07 | 0.07 | 0.00 | 0.15 |
| $L^r$ | 2.77 | 0.41 | 0.13 | 0.93 | 0.24 | 0.09 |
| $T_1$ | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 |
| $T_2$ | 0.44 | 0.07 | 0.05 | 0.08 | 0.00 | 0.12 |
| $T_3$ | 0.05 | 0.01 | 0.00 | 0.01 | 0.00 | 0.03 |
| $T_4$ | 0.02 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 |
| $S$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| $R_1$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $R_2$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 |
| $R_3$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $R_4$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| $N_1$ | 93.10 | 23.01 | 11.80 | 53.56 | 0.10 | 1.91 |
| $N_2$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 |
| $N_3$ | 0.01 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 |

## 5 Conclusions

We have considered nonparametric tests for trend detection in time series. We have not found large differences between the power of the different tests. All tests based on records or ranks react sensitive to autocorrelations. Our results confirm findings by Diersen and Trenkler that $T_3$ can be recommended among the record tests because of its good power and its simplicity. Robustness of $T_3$ against autocorrelation can be achieved for the price of a somewhat reduced power by choosing a large splitting factor $\widetilde{k}$. However, even higher power can be achieved by applying a nonparametric rank test like the seasonal Kendall–Test $S$ or the Spearman–Test $R_1$ with a small $\widetilde{k}$, even though for the price of a higher sensitivity against positive autocorrelation. The power of all rank tests except $N_1$ gets smaller, if a larger splitting factor is used. For $N_1$ a larger splitting factor enlarges the power, but $N_1$ is not recommended to use, as even with a large splitting factor it is less powerful than the other tests. From the rank tests the test $N_2$ seems robust against autocorrelations below 0.6, if only a few observations are taken in each block. Another possibility to reduce the sensitivity to autocorrelation is to fit a low order AR model and consider the AR residuals. We have found a significant trend in the time series of the monthly mean temperature in Potsdam both when using the original data and the AR(1) residuals. Since in the plot of the scaled residuals for this series we find some outliers, another interesting question for further research is the robustness of the several tests against atypical observations.

## References


[1] Aiyar, R.J., Guillier, C.L., Albers, W.: Asymptotic relative efficiencies of rank tests for trend alternatives. J. Am. Stat. Assoc. **74**, 227–231 (1979)

[2] Brockwell, P.J., Davis, R.A.: Introduction to time series and forecasting. Springer, New York (2002)

[3] Cox, D.R., Stuart, A.: Some quick sign tests for trend in location and dispersion. Biometrika **42**, 80–95 (1955)

[4] Daniels, H.E.: Rank correlation and population models. J. Roy. Stat. Soc. B **12**, 171–181 (1950)

[5] Diersen, J., Trenkler, G.: Records tests for trend in location. Statistics **28**, 1–12 (1996)

[6] Diersen, J., Trenkler, G.: Weighted record tests for splitted series of observations, In: Kunert, J., Trenkler, G. (eds.) Mathematical Statistics with Applications in Biometry, Festschrift in Honour of Prof. Dr. Siegfried Schach, pp. 163–178. Eul, Lohmar (2001)

[7] Foster, F.G., Stuart, A.: Distribution-free tests in time-series based on the breaking of records. J. Roy. Stat. Soc. B **16**, 1–22 (1954)

[8] Hirsch, R.M., Slack, J.R., Smith, R.A.: Techniques of trend analysis for monthly water quality data. Water Resour. Res. **18**, 107–121 (1982)

[9] Hirsch, R.M., Slack. J.R.: A nonparametric trend test for seasonal data with serial dependence. Water Resour. Res. **20**, 727–732 (1984)

[10] Kendall, M.G.: A new measure of rank correlation. Biometrika **30**, 81–93 (1938)

[11] Kendall, M.G., Gibbons, J.D.: Rank correlation methods. Arnold, London (1990)

[12] Mann, H.B.: Non-parametric tests against trend. Econometrica **13**, 245–259 (1945)

[13] Moore, G.H., Wallis, W.A.: Time series significance tests based on signs of differences. J. Am. Stat. Assoc. **38**, 153–164 (1943)

[14] R Development Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing. Vienna, Austria, ISBN 3-900051-07-0, URL: http://www.R-project.org (2008)

[15] Silverman, B.W.: EbayesThresh: Empirical Bayes thresholding and related methods. R package version 1.3.0., URL: http://www.bernardsilverman.com (2005)

Springer

http://www.springer.com/978-3-7908-2120-8

Statistical Inference, Econometric Analysis and Matrix Algebra
Festschrift in Honour of Götz Trenkler
Schipp, B.; Krämer, W. (Eds.)
2009, XVI, 434 p., Hardcover
ISBN: 978-3-7908-2120-8
A product of Physica-Verlag Heidelberg