Vol. 11: 97–107, 1999 | CLIMATE RESEARCH Clim Res | Published March 22

Estimation of the beginning and end of recurrent events within a climate regime

Friedrich-Wilhelm Gerstengarbe*, Peter C. Werner

Potsdam Institute for Climate Impact Research, Telegrafenberg, PO Box 60 12 03, D-14412 Potsdam, Germany

ABSTRACT: Exact knowledge of the temporal and spatial variations of the beginning, end and length of a quasi-periodic event can be of importance for estimating climate changes. An extended sequential version of the Mann-Kendall test is presented which makes it possible to describe such events in a statistically significant way and to represent these events in their temporal sequence. The usefulness of this procedure is demonstrated using 2 examples: rainy and dry periods in Northeast Brazil and the annual cycle of relative humidity in Central Europe.

KEY WORDS: Quasi-periodic events · Extended sequential version of the Mann-Kendall test · Rainy and dry periods · Semi-arid regions · Annual cycle of relative humidity

1. BASIC IDEA

The investigation of time series is of particular importance for both meteorology and climatology. For example, in studying trends, variations, etc., the occurrence of recurrent events is often the subject of investigation. In such studies, it is of interest to determine as accurately as possible the beginnings, ends and variations of such events. These results can be used for the detection of climate changes and as basic information for the climate impact research. Examples of this are the beginning and end of rainy or dry periods in the tropics (Rao & Hada 1990), of monsoons (Wagner & Ruprecht 1975), or simply of the seasons in the Central European region.

For strongly periodic events, this problem can be solved using spectral analysis or well-known filter techniques. However, these methods are not applicable if the processes to be investigated are quasi-periodic, i.e. those which consist of a number of periodically recurrent events that vary in length of period (Olberg & Rakoczi 1984). In this case it is also impossible to objectively estimate the beginnings or ends of such events since there is a fuzzy transition from one event to the other. Thus, another method has to be found.

If we examine only 1 of these events, the describing parameter (e.g. daily sum of precipitation) can be characterized by a smoothed curve, and this curve can be interpreted as an oscillation with a length of 1 period (e.g. year). If we follow the curve from the bottom to the top of the oscillation, a point must be reached where the beginning of a trend can be established. Defining this point as the beginning/end of the event it is then possible to estimate this point in a statistically clear way. One well-known application often used in solving such problems is the estimation of the beginning of a trend using the sequential version of the Mann-Kendall test (Sneyers 1975). The beginning of a trend is defined here as the (statistically sound) transition to an increase or decrease in the course of the event. To describe the temporal development of such beginnings of trends within a time series, the sequential version of the Mann-Kendall test has to be applied separately for each event. This procedure is not only ineffective, but it has a further disadvantage as illustrated in Fig. 1, which shows on the basis of decadal sums of precipitation the rainy and dry periods of the meteorological station Tamboril (4° 50' S, 40° 20' W; NE Brazil) from 1921 to 1930. The beginning of the dry period for each year was calculated on the basis of daily values using the sequential version of the Mann-Kendall test, and these dates are also shown in Fig. 1. One can see that the beginning varies from 11 May to 31 August within this

*E-mail: gerstengarbe@pik-potsdam.de

short period of 10 yr, i.e. that this is a quasi-periodic and not a periodic process. Due to this fact, it is necessary to define the length and the position of each investigation period individually in order to solve the sequential version of the Mann-Kendall test. Each variation in the length or position of the investigation period leads to a variation in the result of the sequential version of the Mann-Kendall test. So this partially subjective method of solving the problem reduces the quality of the expected solution. Therefore, an attempt is made to replace the single sequential version of the Mann-Kendall test by a shifted sequential version of this test. The basic idea of this test can be described as follows:

(1) Determination of a common starting point from which to apply the sequential version of the Mann-Kendall test to all recurrent events within a time series;
(2) Carrying out the test with changing interval lengths for the calculation of the beginning and end of the event;
(3) Selection of the best statistical estimation of the point where the trend which defines the beginning or end of the event starts;
(4) Re-run of the procedure by shifting the sequential version of the Mann- Kendall test to the next recurrent event;
(5) Continuation of the Steps (2) to (4) up to the end of the time series.

The complete algorithm is explained using an example in Section 3.

2. MATHEMATICAL-STATISTICAL BASIS

2.1. Sequential version of the Mann-Kendall test

The Mann-Kendall test is parameter-free and comes under the class of rank tests. By this test it is possible to estimate trends within time series. Although the sequential version of the Mann-Kendall test is frequently used, a short overview is given here in order to make the shifted sequential version easier to understand.

The sequential version of the Mann-Kendall test (Sneyers 1975, Taubenheim 1989) is used to test an assumption about the beginning of the development of a trend within a sample $(x_1, \ldots, x_m)$ of the random variable X, based on the rank series R of the progressive and retrograde rows of this sample. The assumption (null hypothesis) is formulated as follows: the sample under investigation shows no beginning of a develop-

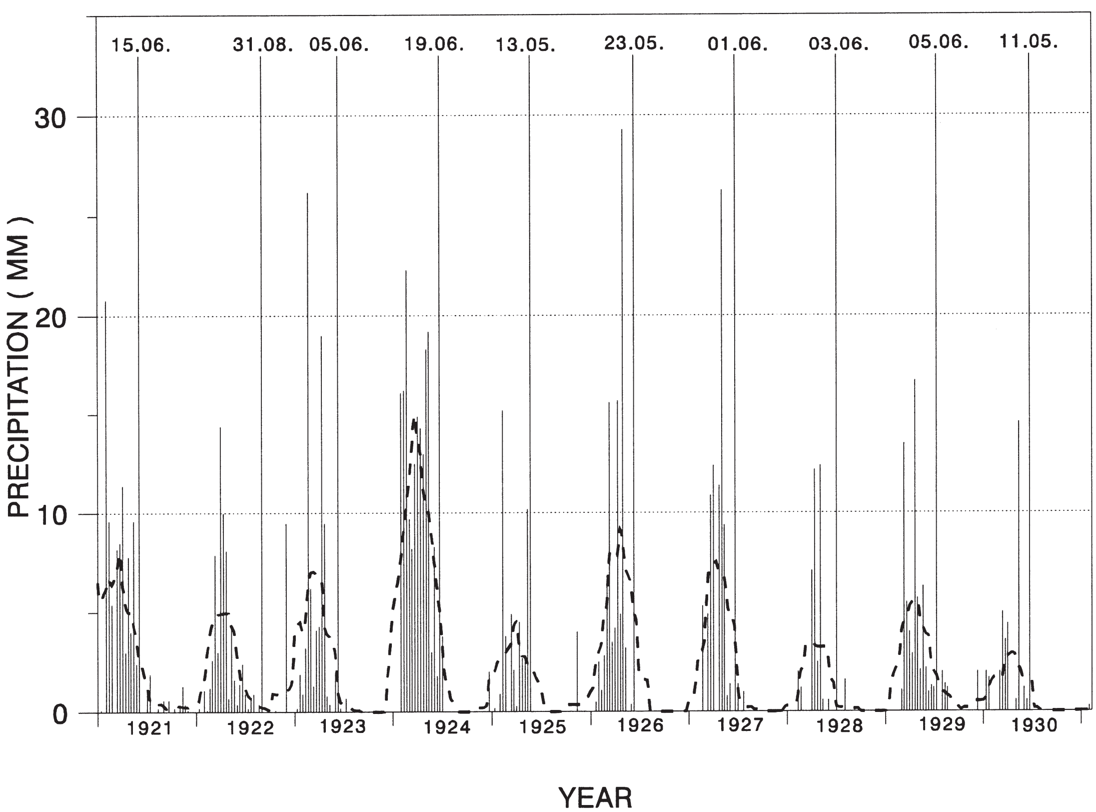

Fig. 1. Precipitation amounts for meteorological station Tamboril (NE Brazil), 1921–1930. (---) Smoothed rainy and dry periods. Dates (dd.mm): beginning of the dry period for each year (for explanation see text)

ing trend. In order to prove or to disprove the assumption, we use the following procedure:

(1) Definition of test statistic. As shown in the Mann-Kendall test (Mann 1945), the test statistic variable is calculated as

$$a = \sum_{i=2}^{m} R_i \qquad (1)$$

where R_i is, for each element x_i, the number of precedent elements x_j $(i > j)$ that are smaller than x_i $(x_i > x_j)$, with $i = 2, \ldots, m$ (size of sample) and $j = 1, \ldots, i - 1$.

In contrast to the Mann-Kendall test, which calculates this test statistic variable only once for the whole sample, the sample is separated into $m - 1$ subseries in the sequential version of the Mann-Kendall test (thus, the first subseries is composed of the sample values x_1, x_2, the second of the values x_1, x_2, x_3, etc.), and the test statistic variable a_k is determined for each of these subseries, k. Thus, $m - 1$ test statistic variables are given as:

$$a_k = \sum_{i=1}^{k} R_i \qquad (2)$$

where R_i is, for each element x_i of the subseries k, the number of elements x_j $(i > j)$ such that $x_i > x_j$, with $i = 2, \ldots, k$ and $j = 1, \ldots, i - 1$.

(2) Calculation of the reduced variables. The introduction of reduced variables is necessary to use the Gaussian statistic in a simple way. For each of the $m - 1$ test statistic variables of the rank series, a reduced variable $u(a_k)$ may be calculated using the following equation:

$$u(a_k) = \frac{a_k - Ea_k}{\sqrt{\sigma^2_{a_k}}} \qquad (3)$$

where Ea_k is the expected value of the respective subseries with

$$Ea_k = \frac{l_k(l_k - 1)}{4} \qquad (4)$$

and $\sigma^2_{a_k}$ is the respective variance given by

$$\sigma^2_{a_k} = \frac{l_k(l_k - 1)(2l_k - 5)}{72} \qquad (5)$$

For $l_k \rightarrow \infty$ (l_k = length of the subseries k) the a_k are approximately of Gaussian type with $u(a_k)$ as its test statistic.

(3) Determination of the beginning of a development of a trend. If the test statistic variables for all subseries are calculated according to Eq. (3), one obtains $m - 1$ reduced variables for the so-called progressive rows, u_p. The corresponding rank series for the so-called retrograde rows, u_r, are similarly obtained for the resorted sample $(x_m, \ldots, x_1)$. The intersection point of the progressive and retrograde rows of the reduced variables gives the point in time of the beginning of a developing trend within the time series.

Thus, the intersection point $s(t_i)$ at time t_i of progressive and retrograde series can be calculated

$$s(t_i) = t_i \quad \text{if} \quad u_p(t_{i-1}) > u_r(t_{i-1}); \quad u_p(t_i) \le u_r(t_i) \qquad (6)$$

$$\text{or} \quad u_p(t_{i-1}) < u_r(t_{i-1}); \quad u_p(t_i) \ge u_r(t_i) \qquad (7)$$

$$= 0 \quad \text{otherwise}$$

The null hypothesis (the sample is not affected by a trend) must be rejected if at least 1 of the reduced variables is greater than a chosen level of significance of the Gaussian distribution.

3. SHIFTED SEQUENTIAL VERSION OF THE MANN-KENDALL TEST

3.1. Basic principle

To describe the temporal sequence of beginnings and ends of recurrent events, the sequential version of the Mann-Kendall test should to be applied as often as necessary within the time series. This can be done by shifting the sequential version of the Mann-Kendall test from event to event over the whole time series. For this purpose, the length and the position of the investigation interval have to be defined. This must be done as objectively as possible. (One way to solve this problem is shown in Section 3.2). The beginning and end of each shifting interval have to be selected such that they are situated in an area between 2 recurrent events that shows no trend. The length of the shifting interval is calculated as the mean length of all recurrent events.

3.2. Procedure

This new technique is further described using the sequence of dry and wet periods in semi-arid regions as an example. This was selected because the wet and dry periods are clearly separated from each other. The following question is to be answered: Is there a temporal variation in the beginning and end of the dry period for the region of the meteorological station Tamboril? The investigation carried out on the basis of daily sums of precipitation for the period 1921–1980.

Steps of the procedure:

(1) For each recurrent process, the number, n, of periods (events) within a selected time interval and the mean length, l, of the interval have to be selected based on the data.

With a time series of daily values for 60 yr with annually recurring events, n would be 60 and l would be 365 in the present case.

(2) The mean point of onset at the minimum or maximum of a recurrent phenomenon within the interval I is determined. For all i values within the interval (i = 1, …, 365), the mean value is formed over the n periods (i.e. displacement steps). To avoid the inclusion of secondary minima and maxima in the calculation, the mean values are smoothed (for example, by means of a triangular filter adjusted to the length of the window). Subsequently, the minimum and/or maximum of the smoothed curve are used to calculate the mean point of onset.

Since the beginning and the end of the dry period are to be determined, the mean point of onset at the precipitation minimum is to be found. Therefore, it is necessary to calculate the 60 yr mean of the daily sums of precipitation.

After averaging (using a triangular filter with a length of 20% of the displacement interval), one finds the minimum on the 247th day, as can be seen in Fig. 2. This figure also shows that smoothing was necessary in order to avoid a selection of secondary minimum values, as for example between the 211th and 214th day.

(3) The sequential version of the Mann-Kendall test is applied with a 2-directional displacement step, to determine both the beginning and the end of each quasi-periodic event. For the first shifting of the sequential version of the Mann-Kendall test, the interval l_b (b: beginning) runs from 1 to i_{mi} (i_{mi}: day of minimum value), for the second, the interval l_e (e: end) runs from i_{mi} to l (see Fig. 3). In general, these intervals have to be successively extended. This is necessary in order to record the most statistically sound variant of a development of a trend, due to the fact that existing natural variations within the observed data could mean that secondary trends are caught. That is why the sequential version of the Mann-Kendall test has to be calculated for each event several times. For the beginning of an event, one can, for example, extend the interval by a maximum of $l_{be} = l_e/2 = (l - i_{mi})/2$ (except for the first recurrent event of the time series). For the end, one can extend the interval to the nth displacement step by a maximum of $l_{ee} = l_b/2 = i_{mi}/2$ (except for the last event of the time series) (e: expanded). The direction of extension, however, depends on whether the beginning and/or the end of a development of a trend is based on a minimum or a maximum. Using even longer intervals involves the risk that the development of trends which do not belong to the investigated displacement step will be recorded.

The sequential version of the Mann-Kendall test to determine the beginning of the dry period was calculated in the first step for the interval from the 1st to the 247th day. The interval was successively extended in 10 d steps.

Because $l_{be} = (l - i_{mi})/2 = (365 - 247)/2 = 59$, there are 6 extension steps n_{be}. This means that the sequential version of the Mann-Kendall test is further calculated for the intervals –9th to +247th, –19th to +247th, …, –59th to +247th day. The –9th day is equivalent to the 357th day of the previous year.

The sequential version of the Mann-Kendall test to determine the end of the dry period was calculated in the same way. The interval was also successively extended in 10 d steps. Because $l_{ee} = i_{mi}/2 = 247/2 = 123.5$, there are 12 extension steps n_{ee}. This means that the sequential version of the Mann-Kendall test is calculated for the interval 247th to 365th day as well as for the 247th to 375th day, etc. The 375th day is equivalent to the 10th day of the following year.

(4) The sequential version of the Mann-Kendall test produces an intersection point between the progressive and retrograde series for each interval which represents the beginning and/or the end of the investigated event. This is done for both Mann-Kendall tests (beginning and end of the event) and for each shift j (with $j = 1, \ldots, n_{be} + 1$, or $n_{ee} + 1$ respectively). Assuming a minimum, Condition (6) is used to determine the beginning of a recurrent phenomenon; assuming a maximum, Condition (6) is used to determine the end of a recurrent phenomenon. Condition (7) is analogous (assuming a minimum or maximum, it is used to determine the end or beginning, respectively, of a recurrent phenomenon).

In the present example, the calculation leads to the result that, for the beginning of the dry period, there are 7 possible intersection points (6 extension steps + starting step) and, for the end of the dry period, there are 13 steps (12 extension steps + starting step).

(5) The test values are determined by the maximum value of the reduced variables u_p and u_r (according to Eq. 3) in the sequential version of the Mann-Kendall test performed for each extension step. Then the greatest calculated test value is selected. The intersection point of this extension step marks the most probable beginning of a development of a trend (beginning and/or end of a recurrent phenomenon).

For the recurrent event during 1923-1924 (3rd displacement step within the time series), the values for the mean point of the end of the dry period were calculated and are shown in Table 1.

As can be seen in Table 1, with the fourth investigation interval (Days 247 to 395) the level of significance steps over the 99% confidence interval of 2.326 of the Gaussian distribution. That means that a significant trend starts on the intersection point Day 386.

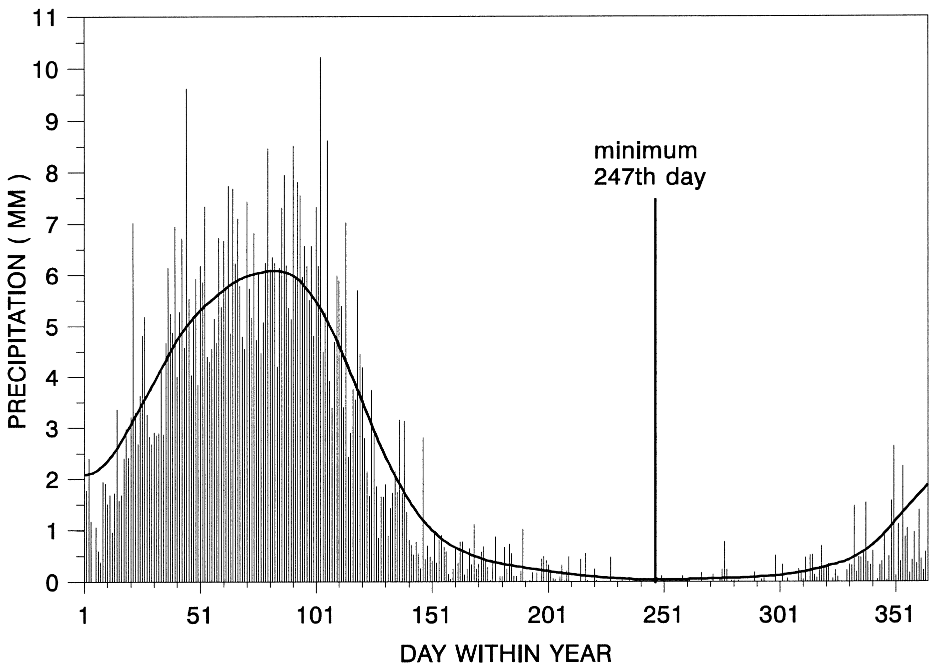

Fig. 2. Means of daily sum of precipitation at Tamboril over the 60 yr interval 1921–1980. Smoothed curve and minimum value obtained after averaging are also shown

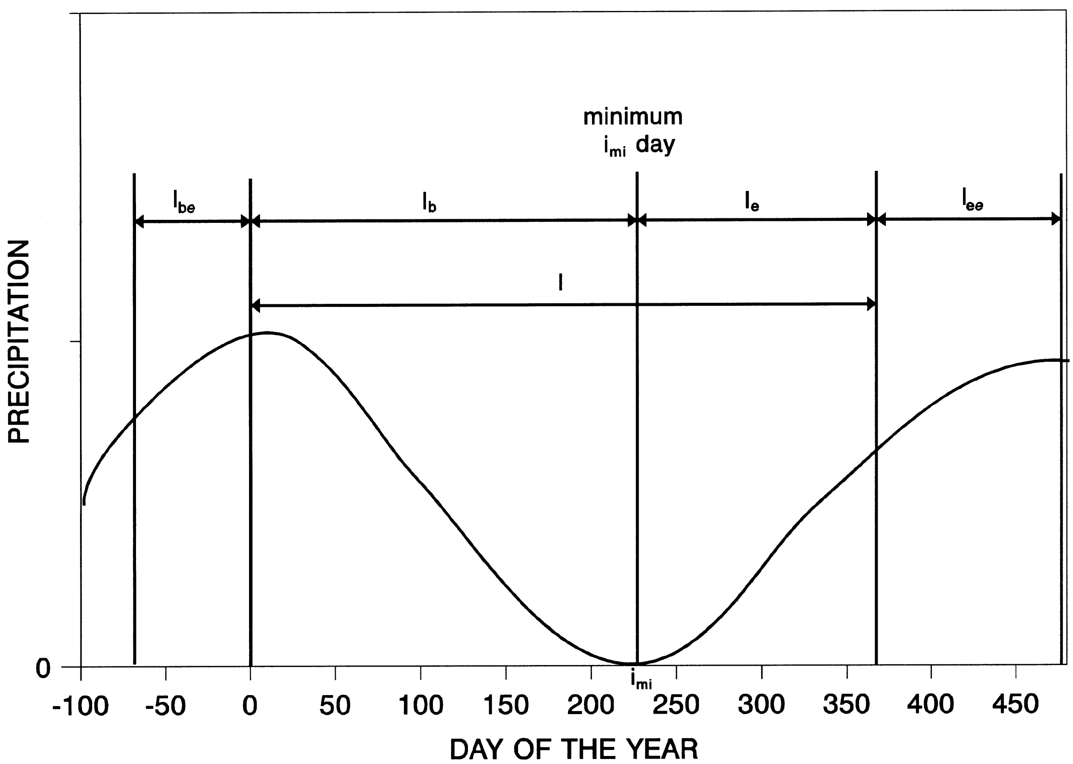

Fig. 3. Principle of definition of investigation intervals. i_{mi}: day of minimum value; l: mean length of interval; l_b: 1 to i_{mi}; l_e: i_{mi} to l; l_{be}: maximum extension at beginning of event; l_{ee}: maximum extension at end of event (for explanation see text)

Table 1. Results of the determination of the mean point of the end of the dry period for Tamboril in 1923-1924

No.	Investigation interval (days)	Intersection point (day)	Test value
1	247–365	–	–
2	247–375	–	–
3	247–385	379	1.02
4	247–395	386	2.71
5	247–405	386	4.04
6	247–415	386	4.96
7	247–425	379	4.80
8	247–435	382	5.33
9	247–445	382	5.77
10	247–455	382	6.16
11	247–465	381	6.26
12	247–475	382	6.77
13	247–485	384	7.20

The same result can be seen for the investigation intervals with the numbers 5 to 13. The highest test value (7.20) was obtained for the investigation interval Days 247 to 485. In this interval, the intersection point is at Day 384, which is the 19th day of the following year (coincidentally, the intersection point is here in accordance with the last possible investigation interval). The good quality of the sequential version of the Mann-Kendall test is also shown by the fact that there are only small changes between investigation intervals. In Fig. 4, the results for the above example are presented and compared with the actual course of precipitation. One can see that the intersection points for the beginning and the end of the dry period correspond well with the actual distribution of precipitation. Similar results are obtained for all recurrent events within the time series.

(6) If one carries out this procedure for the n displacement steps of the time series, the calculated results describe the temporal behaviour of the beginning and the end of recurrent phenomena as well as the statistical soundness of the results for each displacement step.

The entire procedure described so far is calculated for the n = 60 present displacement steps. One obtains the results presented in Fig. 5, which shows the beginning and the end of the dry period from 1921 to 1980. A comprehensible result is obtained if the beginnings and ends of the dry period are distinctly separated from each other, which is the case in Fig. 5.

From these results further information can be obtained, for example:

–lengths of recurrent phenomena (time between beginning and end)
–mean beginning, mean end and mean length
–range of variation of the before-mentioned values
–temporal trends or fluctuations of beginning, end, and length

By means of the presented results, the climatology given in Table 2 can be obtained for precipitation in the Tamboril region.

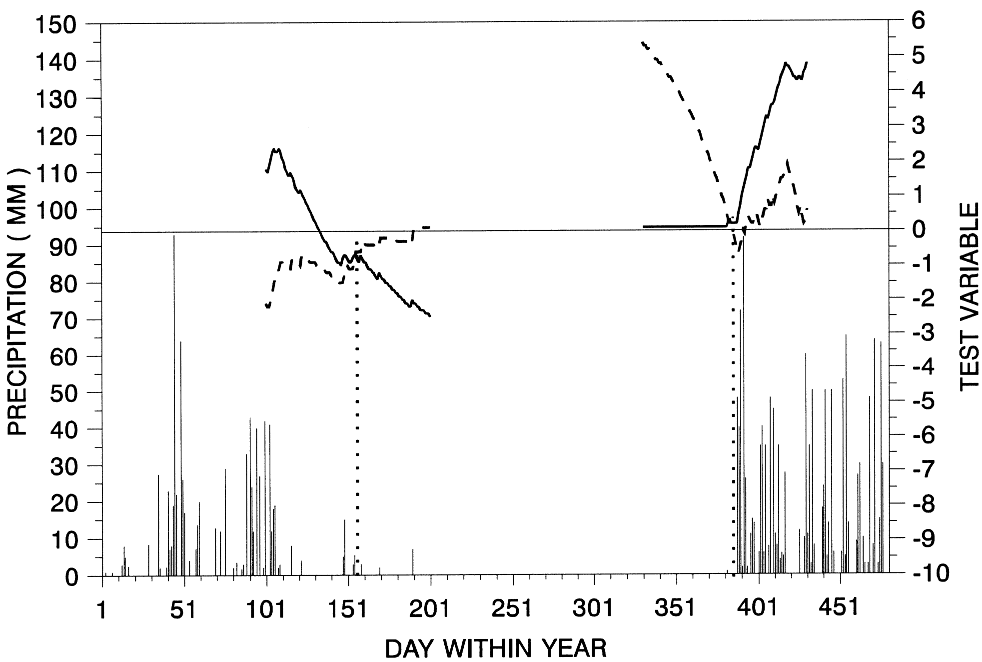

Fig. 4. Beginning and end of the dry period at Tamboril in 1923-1924. The points of intersection of the sequential version of the Mann-Kendall test shows the beginning (156th day) and the end (384th day). Columns represent daily sums of precipitation

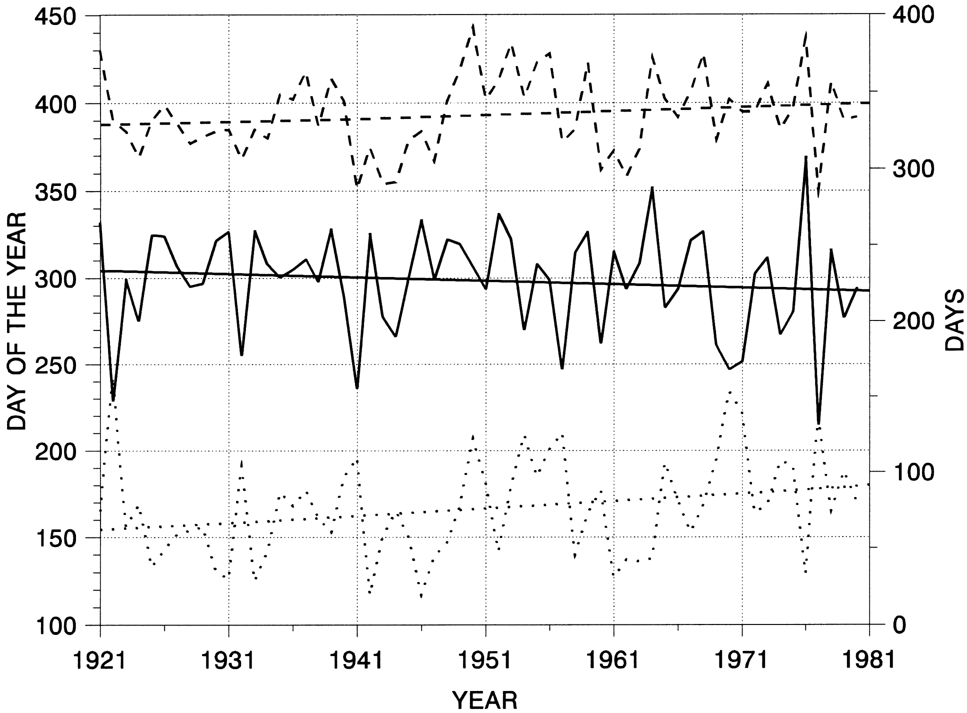

Fig. 5. Beginning (dotted), end (dashed) and duration (solid) of the dry period at Tamboril, 1921–1980

Table 2. Climatology of the dry period at Tamboril, 1921–1980

	Minimum	Maximum	Mean	Standard deviation (d)
Beginning	27 Apr	31 Aug	16 Jun	29.0
End	01 Nov	20 Mar	26 Jan	25.1
Length (d)	131	308	225	36.6

The linear trend indicated in Fig. 5 shows that the beginning of the dry period in 1980 starts 25 d later than in 1921. For the end of the dry period, a similar but weaker trend can be obtained (only 11 d difference between 1921 and 1980); therefore, the length of the dry period is necessarily decreasing. The climatological conclusion from this is that the duration of the dry period has shortened by about 2 wk in the course of the 60 yr between 1921 and 1980.

4. AN INVESTIGATION OF RELATIVE HUMIDITY

4.1. Problem definition

As described above the method can be used for the characterization of the temporal behaviour of recurrent events. Often these events are not clearly divided into separate periods, as discussed in Section 3.2. The behaviour of the relative humidity in Central Europe is characterized by only a weakly structured annual course (as an example see Fig. 6, which shows the course for the Potsdam station). Two questions arise:

– Can the proposed method be used on weakly structured samples?
– Are there temporal developments within the annual cycles of the relative humidity for selected stations in Central Europe?

4.2. Data

Daily mean values of relative humidity were available for 6 stations for the period 1901–1990. These stations are

Hamburg, Germany (53.55° N, 09.97° E; 22 m above MSL [mean sea level]), Potsdam, Germany (52.38° N, 13.07° E; 81 m above MSL), Prague, Czech Republic (50.10° N, 14.30° E; 381 m above MSL), Karlsruhe, Germany (49.03° N, 08.37° E; 112 m above MSL), Hohenpeissenberg, Germany (47.80° N, 11.02° E; 977 m above MSL), and Zürich, Switzerland (47.40° N, 08.60° E; 569 m above MSL). All data were tested for homogeneity by various methods in a preliminary investigation (Gerstengarbe & Werner 1993a). One of the findings was that some of the data series show non-homogeneities within the period from 1940 to 1945 caused by a lack of measurements. These non-homo-

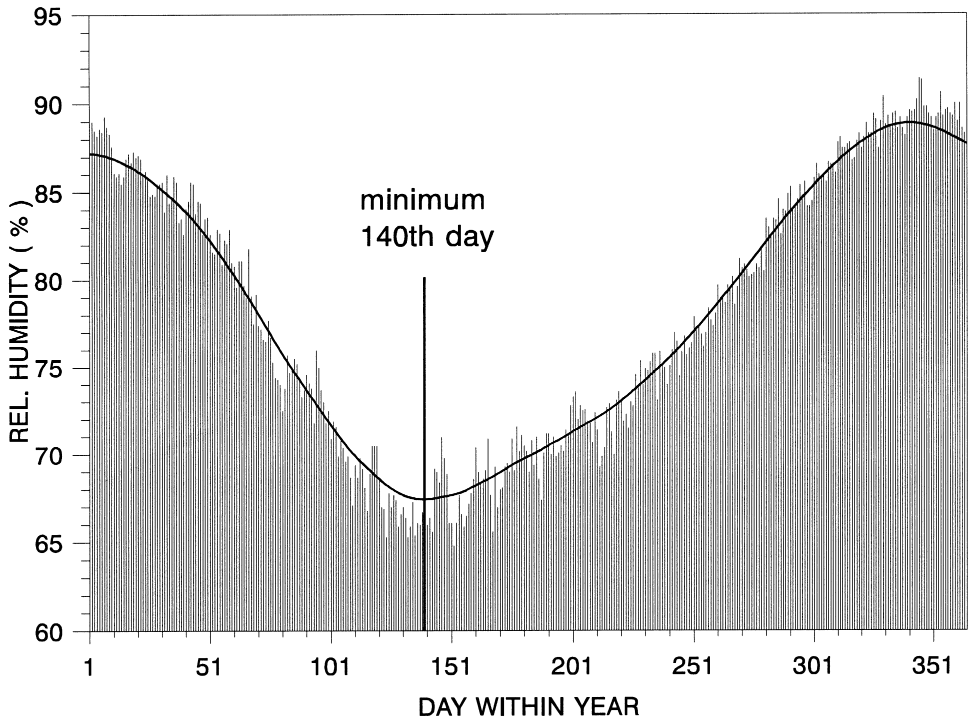

Fig. 6. Means of daily mean relative humidity at Potsdam over the 90 yr interval 1901–1990. Smoothed curve and minimum value obtained after averaging are also shown

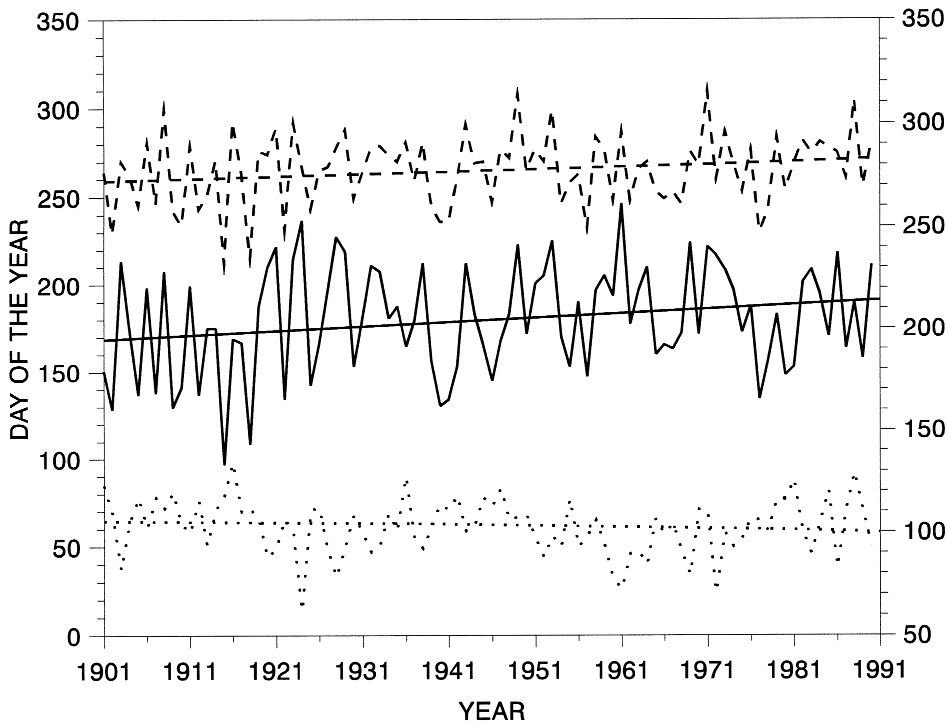

Fig. 7. Beginning (dotted), end (dashed) and duration (solid) of the 'dry' period at Potsdam, 1901–1990

Table 3. Means and trends of beginning, end and length of the 'dry' period for selected European stations over the period 1901–1990

Station	Beginning		End		Length	
	Mean	Trend (d)	Mean	Trend (d)	Mean	Trend (d)
Hamburg	8 Mar	−7.4	19 Sep	−3.8	195	3.6
Potsdam	2 Mar	−6.6	22 Sep	12.7	204	19.3
Prague	3 Mar	−1.8	12 Oct	4.6	223	6.4
Karlsruhe	21 Feb	5.1	13 Sep	27.0	204	21.9
Hohenpeissenberg	1 Mar	−2.3	30 Sep	21.1	213	23.4
Zürich	17 Feb	−0.5	13 Sep	16.4	208	16.9

geneities have been taken into consideration in the following investigations and interpretations.

4.3. Results

The investigations were carried out as described above. The following results were obtained:

– Two periods were selected: the 'dry' period is longer than the 'wet' period and starts in March and ends in September (for example, the results for Potsdam station are shown in Fig. 7). For all stations, with variations of a maximum of 18.7 d at the beginning and 28.8 d at the end of the 'dry' period were found (see Table 3).
– All 3 parameters (beginning, end and length) of the 'dry' period show a trend (see Fig. 7). The mean tendencies for the 3 parameters and all stations are given in Table 3. For all stations, the beginning of the 'dry' period becomes progressively earlier over the 90 yr period (except for Karlsruhe). For the end, the inverse tendency (except for Hamburg) is obtained. The result of this is an extension of the length of the 'dry' period for all stations.

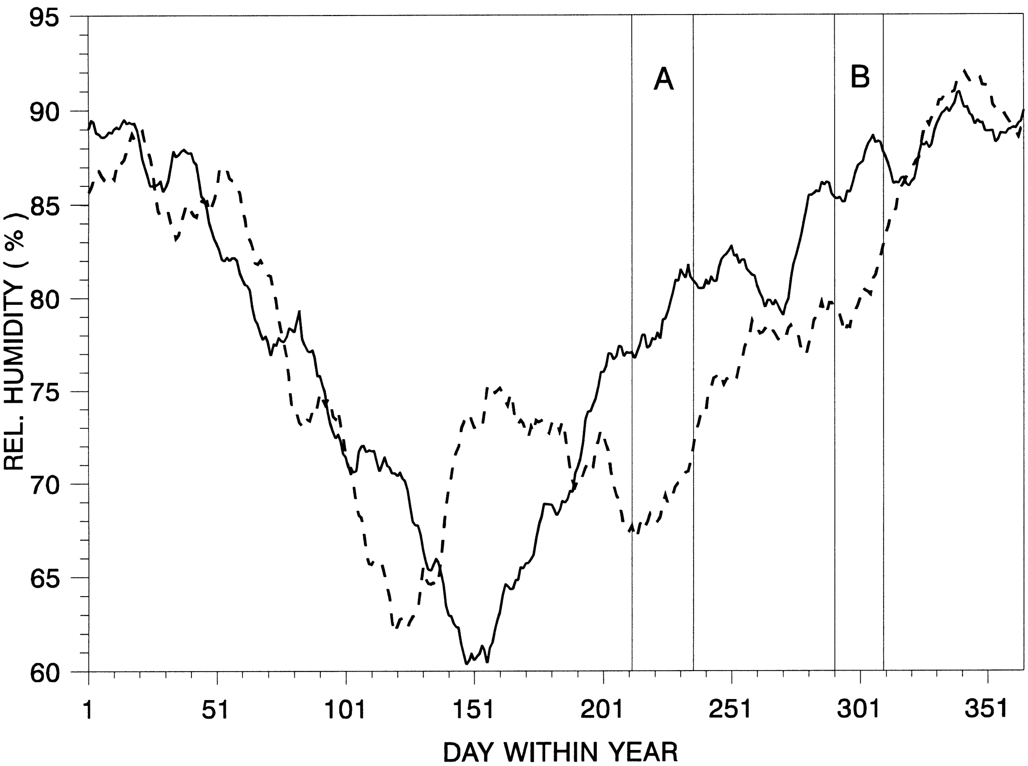

Fig. 8. Mean annual cycle of relative humidity for years with an early end (solid) and a late end (dashed) of the 'dry' period at Potsdam, 1901–1990. Interval for the end of the 'dry' period: A, early; B, late

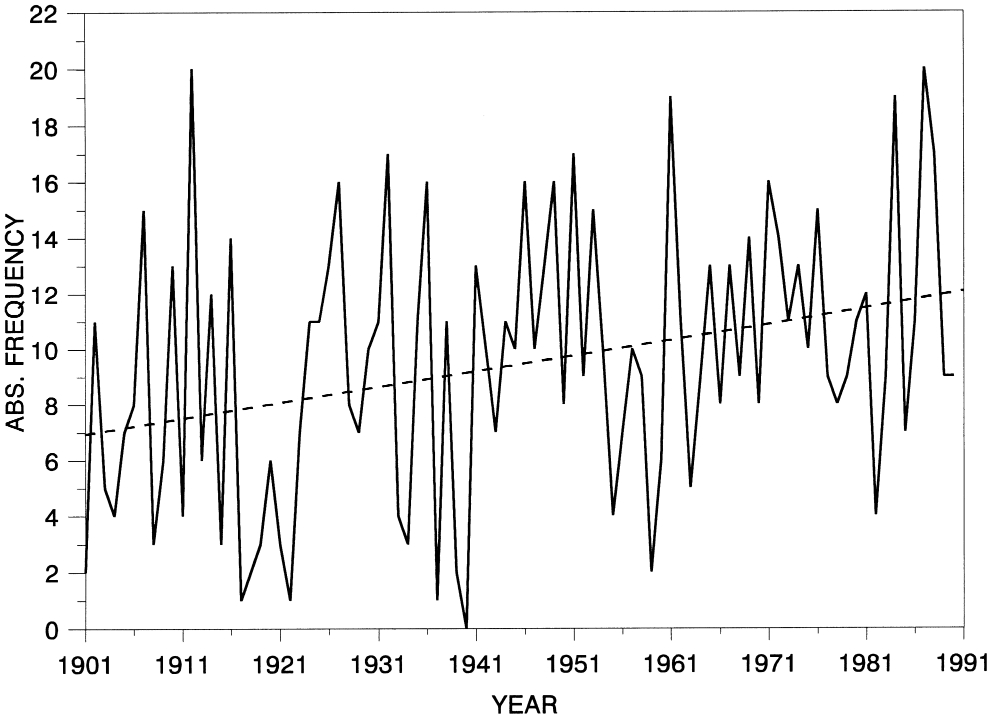

Fig. 9. Absolute frequency of cyclonal circulation patterns between 141st and 160th day of the year (derived from the Grosswetterlagen of Hess/Brezowsky for Central Europe, 1901–1990)

- The observed trends lead to a change in the characteristic annual course of the relative humidity. An example is given in Fig. 8. The figure shows the mean annual course of years with a very early end of the ‘dry’ period and those with a very late end. The expected differences in the structures are clearly visible.
- In addition, one can see an inverse behaviour of the 2 mean annual courses between the 141st (21st May) and 160th (9th June) day of the year; within this interval the years with a late end of the ‘dry’ period are markedly wetter than those with an early end.

In general, these interpretations have to be followed by a discussion of the climatic reasons. For example, different circulation patterns could be a source for the estimated trends. This is confirmed by the fact that the cyclonal circulation patterns for this interval occur much more frequently in combination with the late end of the ‘dry’ period than for the early end (55.7% to 32.8%). If one examines the temporal behaviour of the cyclonal circulation patterns (Gerstengarbe & Werner 1993b) for the interval from the 141st up to the 160th day of the year, an increase of about 5 d (i.e. 25%) is seen (see Fig. 9). Simultaneously, a weakening of the pre-monsoonal weather period can be observed (Baur 1948, Gerstengarbe & Werner 1987). These results are in good agreement with the trends described above.

This short interpretation is an example of how the results of the method can be used to detect and to discribe climate changes.

5. CONCLUSIONS

Using examples it has been shown that the suggested methodology is suitable for describing the temporal development of recurrent events. The advantage of the shifted sequential version of the Mann-Kendall test is, on the one hand, its effectiveness in calculating the temporal development of beginnings and ends of events within recurrent processes, and, on the other hand, the better precision of the results due to the optimal determination of the beginnings and ends by using expanded calculation intervals. It is concluded that the method can be applied in all areas where similar questions occur. The most important thing is the ‘right’ treatment of the investigation series in regard to the selection of the starting conditions and/or the shifting algorithm. The benefit of this method is that the shifted sequential version of the Mann-Kendall test enables a statistically significant statement about the entire temporal development of a series.

The examples show that the suggested methodology represents an improvement in describing climate processes because the tendencies detected are often a result of changes in atmospheric circulation.

LITERATURE CITED

Baur F (1948) Zur Frage der Echtheit der sogenannten Singularitäten im Jahresgang der Witterung. Ann Meteorol 12

Gerstengarbe FW, Werner PC (1987) Ist der Baur'sche Kalender der Witterungsregelfälle heute noch gültig? Z Meteorol 37(5):263–272

Gerstengarbe FW, Werner PC (1993a) Extreme klimatologische Ereignisse an der Station Potsdam und an ausgewählten Stationen Europas. Ber Dtsch Wetterdienst 186

Gerstengarbe FW, Werner PC (1993b) Katalog der Großwetterlagen Europas nach Paul Hess und Helmut Brezowsky 1881 – 1992. Ber Dtsch Wetterdienst 113(4)

Mann HB (1945) Non parametric test against trend. Econometrica 13:245–259

Olberg M, Rakoczi F (1984) Informationstheorie in Meteorologie und Geophysik. Akademie-Verlag, Berlin

Rao VB, Hada K (1990) Characteristics of rainfall over Brazil: annual variations and connections with the Southern Oscillation. Theor Appl Climatol 42:81–91

Sneyers R (1975) Sur l'analyse statistique des séries d'observations. WMO Tech Note

Taubenheim J (1989) An easy procedure for detecting a discontinuity in a digital time series. Z Meteorol 39(6): 344–347

Wagner M, Ruprecht E (1975) Materialien zur Entwicklung des Indischen Sommermonsuns. Bonn Meteorol Abh 23

Editorial responsibility: Hans von Storch, Geesthacht, Germany

Submitted: July 23, 1997; Accepted: January 8, 1998
Proofs received from author(s): February 16, 1999