Trends and extreme values in climatological time series : what do they tell about climate change? : part 2 : recurrence times in series of Bernoulli trials

Technische
Universiteit
Eindhoven

Instituut
Wiskundige Dienstverlening
Eindhoven

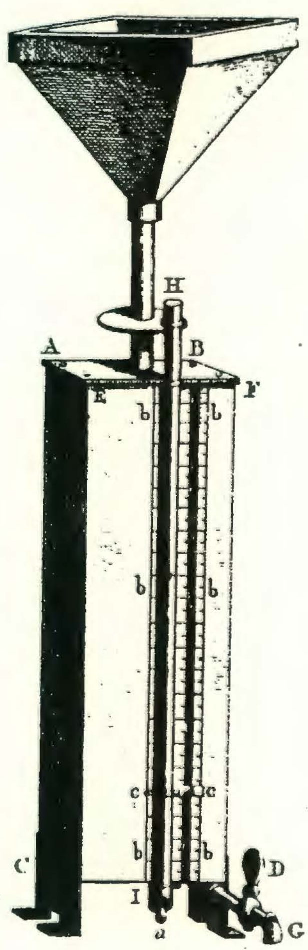
report IWDE 93 - 05

TRENDS AND EXTREME VALUES IN CLIMATOLOGICAL TIME SERIES : WHAT DO THEY TELL ABOUT CLIMATE CHANGE ?

Part 2 : RECURRENCE TIMES IN SERIES OF BERNOULLI TRIALS

J. Molenaar and H. Visser

May 1993



Den Dolech 2
Postbus 513
5600 MB Eindhoven

KEMA Research and Development

report IWDE 93 - 05

TRENDS AND EXTREME VALUES IN CLIMATOLOGICAL TIME SERIES : WHAT DO THEY TELL ABOUT CLIMATE CHANGE ?

Part 2 : RECURRENCE TIMES IN SERIES OF BERNOULLI TRIALS

Authors : dr. J. Molenaar and ir. H. Visser

May 1993

Abstract

In this report recurrence times in series of Bernoulli trials are studied. We show that the results depend on the event definition one is interested in. Analytical expressions and numerical results are given for six different event definitions. In addition, we consider an event, which is quite general and close to intuition, but does not admit an analytical approach yet. From the numerical results, however, we conclude that the values of the corresponding recurrence times are very close to the values of a slightly less general event definition, for which analytical expressions are available.
The theory is applied to series of annual mean temperatures averaged over the Northern Hemisphere. This series has remarkably many hot years during the last decade. The recurrence times of this pattern appear to be quite long irrespective the event definition used. This is a strong indication that the warming of the earth due to the so-called Greenhouse Effect is already detectable. The theory of recurrence times presented in this report provides a useful tool to quantify this kind of phenomena.

This report has been written under a contract with Kema Environmental Services, Arnhem. The first part of this report has been issued by the Kema under number 92252-MOF 92-3683.

One of the authors (JM) wishes to thank F.W. Steutel from Eindhoven University of Technology for useful discussions.

Contents

1. Introduction and motivation

In this report we study recurrence times in series of Bernoulli trials. Such trials have two possible outcomes with constant chances p and 1-p; the generic example is the tossing of a coin.

Before going into mathematical details we wish to give our motivation for these investigations. Our interest in this subject stems from the analysis of climatological time series. If the trend is removed, a series of residuals remains, which represents the short term fluctuations in the signal. For trend estimation techniques we refer to Visser and Molenaar (1990). In KEMA (1992) and Visser and Molenaar (1992) we investigated annual mean temperature data and observed that the residuals of these series are uncorrelated. They are identically and independently distributed and follow normal distributions. One of the investigated series contains temperature data averaged over the Northern Hemisphere during the period 1881 - 1990. In the literature this kind of information attracts a lot of attention, because of the supposition that the earth is warming on the average due to the so-called Greenhouse Effect. This phenemenon is often taken for granted, although careful studies point out that the detection of these effects is not yet satisfactorily established. See, e.g., Wigley and Barnett (1990).

One of the techniques to detect possible warming effects is to estimate the trend using the data from the period 1880 - 1980 and extrapolating this trend estimation over the decade 1980 - 1990. With respect to this extrapolated trend we found that the residuals in the last decade are all positive and attain quite high values. This indicates that indeed climate is changing rather abruptly. Such a supposition has to be quantified and to that end we studied the recurrence times of the observed patterns in the series of exceedances of a given prescribed temperature level (KEMA, 1992; Visser and Molenaar, 1992). Wishing to use the notion of recurrence time we came to the conclusions that no much literature is available at this point and that this property has to be introduced with care. Wehad therefore to extend the existing theory. In the following we present the mathematical details of our approach.

2. Events and recurrence times

Let $\{n_t, t = 1,...,N\}$ be the series under consideration. This series is assumed to be independently and identically distributed (i.d.d). It is convenient to associate with this series a second series $\{x_t, t = 1,...,N\}$ defined by

$$x_t = 1 \ \mathit{if}\ n_t > L \qquad \qquad (1)$$
$$0 \ \mathit{otherwise}$$

with L a prescribed level. Because the x_t are i.d.d., we have for all t

$$P(x_t = 1) = 1 - P(x_t = 0) = 0 \qquad (2)$$

with p a constant chance which follows directly from the distribution and the level L. In the literature such a series is also referred to as a series of *Bernoulli trials* and instead of the symbols 1 and 0 one sometimes uses s(ucces) and f(ailure).

To facilitate the formulations we use the concept of window. For r, s positive integers with $s \leq r$ we define :

- a **window** consists of r consecutive elements of a given series;
- an **r-window** is a window with r elements;
- a **one-window** is a window containing only ones;
- a **zero-window** is a window containing only zeros.

We call the occurence of a specific pattern of ones and zeros in the x-series an *event*. One may be interested in several types of events. In practice it is not always clear in advance which event definition is most appropriate. In this report we shall therefore deal with a number of event definitions which may commonly come into consideration. These definitions are summarized in
We note that E_1 is a special case of E_5, and E_4 a special case of E_7, but it is convenient to deal with E_1 and E_4 separately. E_2 is introduced because Feller (1957) derived explicit formulae for the corresponding recurrence time.

The definition of recurrence time is most conveniently given in terms of a third series $\{z_t, t = 1,...,N\}$ associated with both the x-series and the event E under consideration:

$$z_t = 1 \ \mathit{if\ event\ E\ occurs\ at\ position\ t} \qquad (3)$$
$$0 \ \mathit{otherwise}$$

To illustrate this point we give in Table 2 an arbitrarily chosen x-series together with the corresponding z-series.

Table 1
Let r, r', s, t be positive integers.
We define the following **events** to occur at position t in the x-series :

E_1: position t is the last element of an r-one-window.

E_2: position t the last element of an r-one-window, and this does not hold for the other positions of this window.

E_3: position t is element of an r-one-window, and position t is no element of an r'-one-window with r' > r.

E_4: position t is element of an r-one-window.

E_5: position t is the last element of an r-window with exactly s ones.

E_6: position t is the last element of an r-window with at least s ones.

E_7: position t is element of an r-window with at least s ones.

Table 2.
In this example r = 3, s = 2.

x-series	: 0 0 1 1 1 1 0 0 0 1 0 1 1 0 1 0 0 0 1 1 1 0 0
z-series according to E_1	: 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
z-series according to E_2	: 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
z-series according to E_3	: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
z-series according to E_4	: 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
z-series according to E_5	: 0 0 0 1 0 0 1 0 0 0 0 1 1 1 1 0 0 0 0 1 0 1 0
z-series according to E_6	: 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 0
z-series according to E_7	: 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0 1 1 1 1 1 0

In the following we call a zero-window **maximal** if it is preceded and succeeded by a one.

Definition : According to Kottegoda (1980) we define the **recurrence time** R as the average length of maximal zero-windows in the z-series.

The average number of zeros in the z-series is, for $N \rightarrow \infty$, given by $N * P(z_t = 0)$. Note that $P(z_t = 0)$, and thus $P(z_t = 1)$, becomes independent of t in the limit $N \rightarrow \infty$. Because maximal zero-windows are uniquely connected with the occurrence of the pattern (1,0) (or the pattern (0,1)), the average number of such windows is $N * P(z_t = 1, z_{t+1} = 0)$ for $N \rightarrow \infty$. Because of symmetry we have $P(z_t = 1, z_{t+1} = 0) = P(z_t = 0, z_{t+1} = 1)$.
The recurrence time R is, in the limit $N \rightarrow \infty$, given by (Lawrence and Kottegoda, 1977)

$$R = \frac{P(z_t = 0)}{P(z_t = 1,\ z_{t+1} = 0)} \tag{4a}$$

Although we shall apply in this report only expression (4a), it might be of general importance in this kind of investigations to have at hand other formulae for R. They involve conditional probabilities. Because $P(z_t = 1 , z_{t+1} = 0) = P(z_t = 1 \mid z_{t+1} = 0) * P(z_{t+1} = 0) = P(z_{t+1} = 0 \mid z_t = 1) * P(z_t = 1)$ two alternative formulae are

$$R = \frac{1}{P(z_t=1 \mid z_{t+1}=0)} \tag{4b}$$

and

$$R = \frac{1 - P(z_t=1)}{P(z_{t+1}=0 \mid z_t=1)P(z_t=1)} \tag{4c}$$

Here, we have used that $P(z_{t+1} = 0) = P(z_t = 0)$ in the limit $N \rightarrow \infty$.

3. Derivations

The different event definitions E_1 - E_7 give rise to different recurrence times R_1 - R_7. We shall derive expressions for R_1 - R_6, and point out how R_7 can be conveniently calculated.

3.1 Recurrence time R_1 for event E_1

Event E_1 occurs if position t is the last element of an r-one-window.
We apply formula (4a). The numerator follows directly from

$$P(z_t=1) = p^r \tag{5}$$

and thus

$$P(z_t=0) = 1-P(z_t-1) = 1-p^r \tag{6}$$

The denominator in (4a) is in this case found from

$$P(z_t=1,\ z_{t+1}=0) = P(x_{i-r+1}=1,\ \ldots\ldots,\ x_i=1,\ x_{i+1}=0) = (1-p)p^r \tag{7}$$

Thus

$$R_1(p,r) = \frac{1-p^r}{(1-p)p^r} = \frac{1}{(1-p)p^r} - \frac{1}{1-p} \tag{8}$$

For p ↓ 0 R_1 converges to $1 / p^r$. Figure 1 shows R_1 for several values of r :

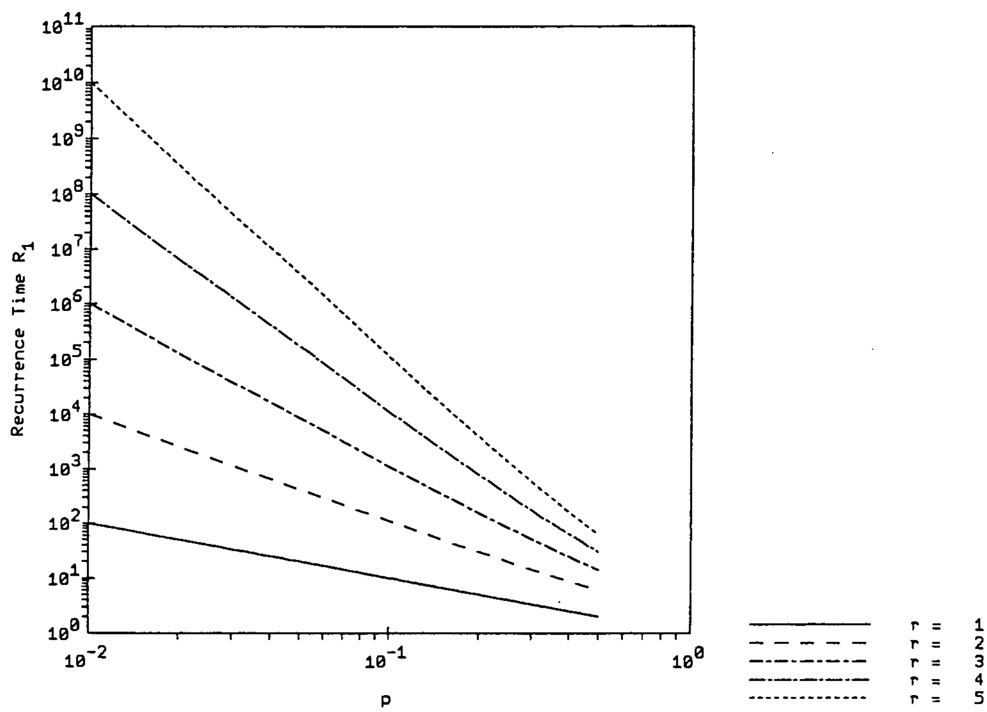


Fig.1. Recurrence time R_1 as a function of chance p for five values of parameter r.

3.2 Recurrence time R_2 for event E_2

Event E_2 occurs if position t is the last element of an r-one-window, while this does not hold for the other window elements. The definition of E_2 stems from Feller (1957). It implies that the ones in the z-series are always separated from each other by a zero-window of length at least r-1. We assume in this case r > 1. Because of the isolation of the ones, the denominator in (4a) follows from

$$P(z_t=1,\ z_{t+1}=0) = P(z_t=1) \tag{9}$$

To calculate $P(z_t = 1)$ and thus $P(z_t = 0) = 1 - P(z_t = 1)$, we enumerate those patterns in the x-series, which yield $z_t = 1$. For example, using r = 3 we find the patterns :

$$\begin{array}{rl}
 & \textit{position } t \\
 & \quad\downarrow \\
\ldots\ldots\ldots\ldots\ldots\ldots\ 0\ 1\ 1\ 1 & \textit{with chance } (1-p)\ p^r \\
\ldots\ldots\ldots\ 0\ 1\ 1\ 1\ 1\ 1\ 1 & \textit{with chance } (1-p)\ p^{2r} \\
\ldots\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1 & \textit{with chance } (1-p)\ p^{3r} \\
 & \textit{etcetera}
\end{array} \tag{10}$$

From this we see that

$$P(z_t = 1) = \sum_{j=1}^{\infty} (1-p)p^{jr} = \frac{(1-p)\ p^r}{1-p^r} \tag{11}$$

Application of (4a) yields

$$R_1(p,r) = \frac{1-p^r}{(1-p)p^r} = \frac{1}{(1-p)p^r} - \frac{1}{1-p} \tag{12}$$

This is in agreement with Feller's result (Feller, 1957, chapter 13). There is a slight difference, but that is simply because Feller's definition of recurrence time includes one position more than ours. In Feller's approach it is essential that the windows are not-overlapping. The present approach is more general and applicable without the latter restriction.

In the first instance it seems remarkable that R_1 and R_2 differ so little. However, it should be noticed that definition E_2 leads on the one hand to more zero-windows of length greater than r, but on the other hand to an enhanced number of zero-windows of length r-1.

For p $\downarrow$ 0 R_2 converges to $1 / p^r$. Figure 2 shows R_2 for several values of r :

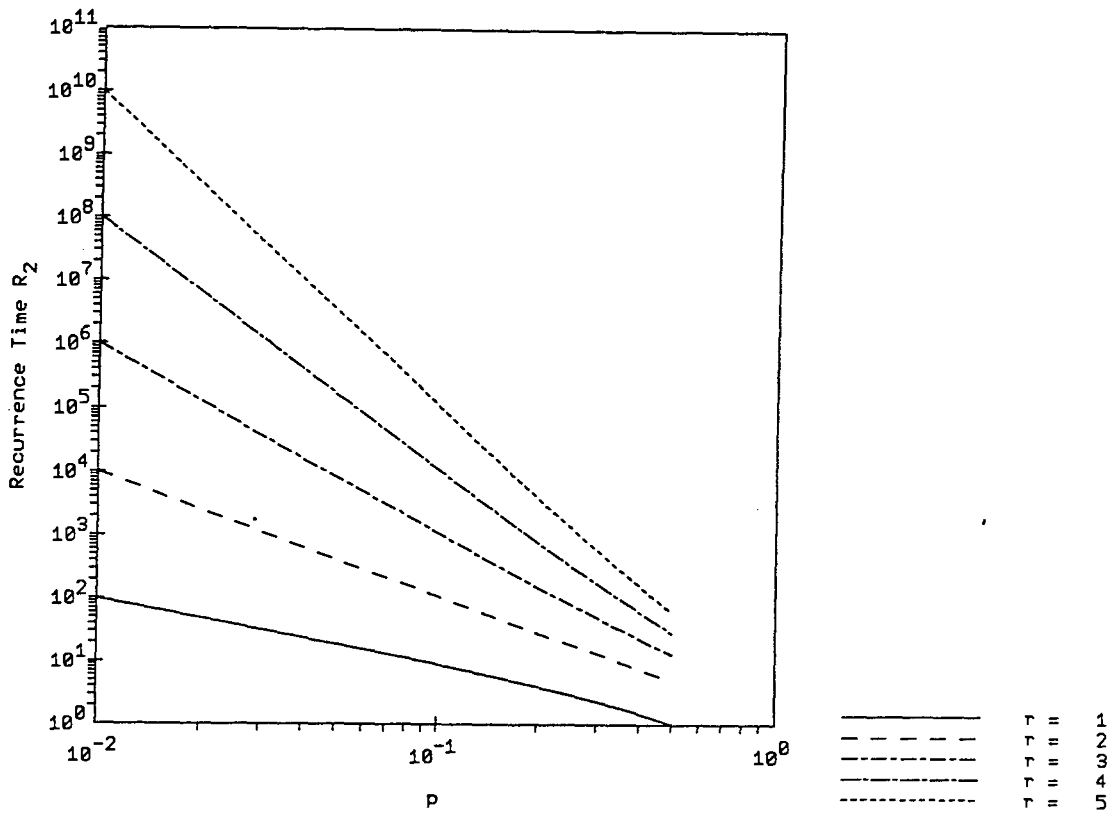


Fig.2. Recurrence time R_2 as a function of chance p for five values of parameter r.

3.3 Recurrence time R_3 for event E_3

Event E_3 occurs if position t is element of an r-one-wimdow, but no element of an r'-one-window with r' > r. In this case $P(z_t = 1)$ is the chance that position t is element of the window (0,1,...,1,0) in the x-series with exactly r ones on a row. Thus

$$P(z_t=1) = r(1-p)^2p^r \tag{13}$$

As for the denominator of (4a) we note that the combination ($z_t = 1$, $z_{t+1} = 0$) can occur only if the last one in the window (0,1,...,1,0) in the x-series occupies position t. Thus,

$$P(z_t=1,\ z_{t+1}=0) = (1-p)^2p^r \tag{14}$$

Application of (4a) yields

$$R_3(p,r) = \frac{1}{(1-p)^2p^r} - r \tag{15}$$

For p $\downarrow$ 0 R_3 converges to $1 / p^r$. Figure 3 shows R_3 for several values of r :

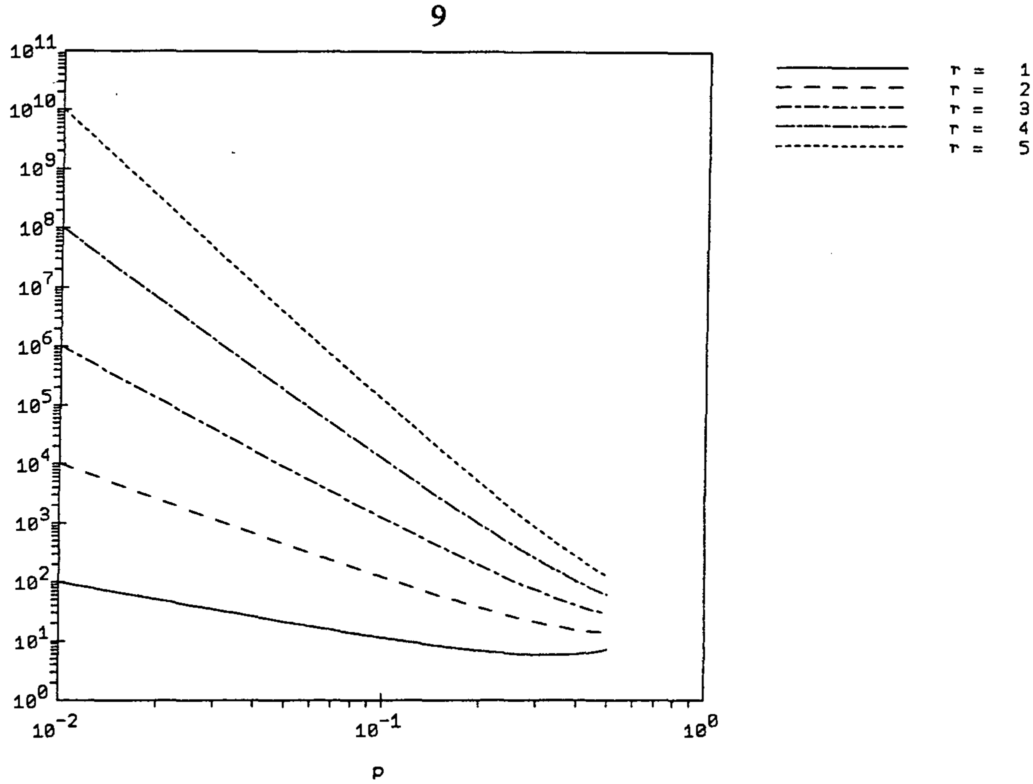


Fig.3. Recurrence time R_3 as a function of chance p for five values of parameter r.

3.4 Recurrence time R_4 for event E_4

Event E_4 occurs if position t is element of an r-one-window.
In this case $P(z_t = 1)$ is the chance that position t is element of the window (0,1,...,1,0) in the x-series with $j \geq r$ ones on a row. This chance is given by

$$\begin{aligned} P(z_t=1) &= \sum_{j=r}^{\infty} j\,(1-p)^2\,p^j \\ &= p\,(1-p)^2 \sum_{j=r}^{\infty} j\,p^{j-1} \\ &= p\,(1-p)^2 \frac{d}{dp}\left[\sum_{j=r}^{\infty} p^j\right] \\ &= p\,(1-p)^2 \frac{d}{dp}\left[\frac{p^r}{1-p}\right] \\ &= p\,(1-p)^2 \left[\frac{rp^{r-1}}{1-p} + \frac{p^r}{(1-p)^2}\right] \end{aligned} \tag{16}$$

The pair $(z_t = 1,\ z_{t+1} = 0)$ implies in the x-series the pattern (1,...,1,0) with r ones on a row and the last zero at position t. So,

$$P(z_t=1,\ z_{t+1}=0) = p^r(1-p) \tag{17}$$

Application of (4a) yields

$$R_4(p,r) = \frac{1}{(1-p)p^r} - r - \frac{p}{1-p} \tag{18}$$

For p $\downarrow$ 0 R_4 converges to 1 / p^r. Figure 4 shows R_4 for several values of r :

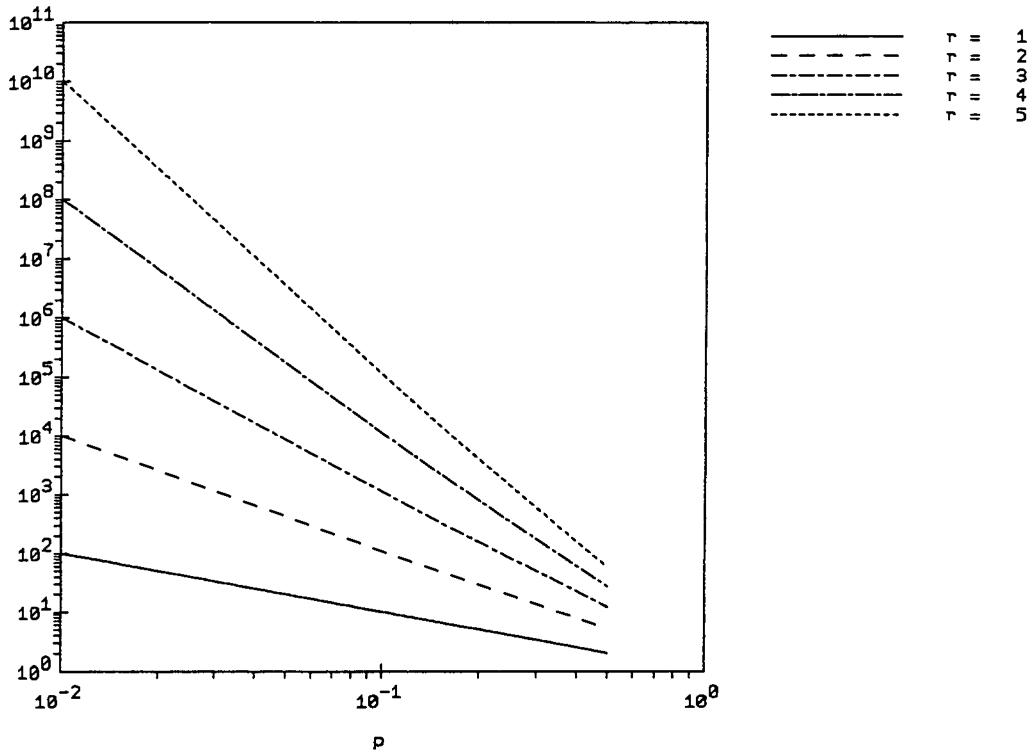


Fig.4. Recurrence time R_4 as a function of chance p for five values of parameter r.

3.5 Recurrence time R_5 for event E_5

Event E_5 occurs if position t is the last element of an r-window with exactly s ones.
In this case it is convenient to introduce the chance Q(r,s) that an r-window in the x-series contains exactly s ones with $0 \leq s \leq r$. This chance follows the binomial distribution :

$$Q(r,s) = \binom{r}{s} p^s (1-p)^{r-s} \tag{19}$$

Q has the properties

$$\begin{aligned} Q(r,r) &= p^r \\ \sum_{j=0}^{r} Q(r,j) &= 1 \end{aligned} \tag{20}$$

We consider the r-window with position t as its last element. $P(z_t = 0)$ is the chance to find less or more than s ones in this r-window :
The pair (z_t = 1, z_{t+1} = 0) implies that the r-window at positions (t-(r-1),..., t) contains exactly s

$$P(z_t=0) = \sum_{j=0,\, j\neq s}^{r} Q(r,j) = 1 - Q(r,s) \tag{21}$$

ones. If this window is shifted one position to the right, it contains (s-1) ones. From this we conclude that positions t-(r-1) and t+1 contain a one and a zero respectively. The (r-1)-window between these positions contains exactly s-1 ones. This means that

$$P(z_t=1, z_{t+1}=0) = pQ(r-1,s-1)(1-p) \tag{22}$$

Application of (4a) yields

$$R_5(p,r,s) = \frac{1 - Q(r,s)}{pQ(r-1,s-1)(1-p)} \tag{23}$$

If we set s = r, R_5 reduces to R_1, given in (8) :

$$R_5(p,r,r) = \frac{1 - Q(r,r)}{pQ(r-1,r-1)(1-p)} = \frac{1 - p^r}{p\, p^{r-1}\,(1-p)} \equiv R_1(p,r) \tag{24}$$

For p ↓ 0 we have

$$R_5 \approx \frac{1}{\binom{r-1}{s-1} p^s} \tag{25}$$

Figure 5 shows R_5 for r = 10 and several values of s.

3.6 Recurrence time R_6 for event E_6

Event E_6 occurs if position t is the last element of an r-window with at least s ones. This case is only a slight modification of the preceding one. Instead of (21) we now have

$$P(z_t=0) = \sum_{j=0}^{s-1} Q(r,j) \tag{26}$$

The reasoning leading to (22) applies here too. R_6 is given by

$$R_6(p,r,s) = \frac{\sum_{j=0}^{s-1} Q(r,\, j)}{pQ(r-1,s-1)(1-p)} \tag{27}$$

For p ↓ 0 R_6 has just the same limiting behaviour as R_5. This behaviour is given by (25).

$$R_6 \approx \frac{1}{\binom{r-1}{s-1} p^s} \tag{28}$$

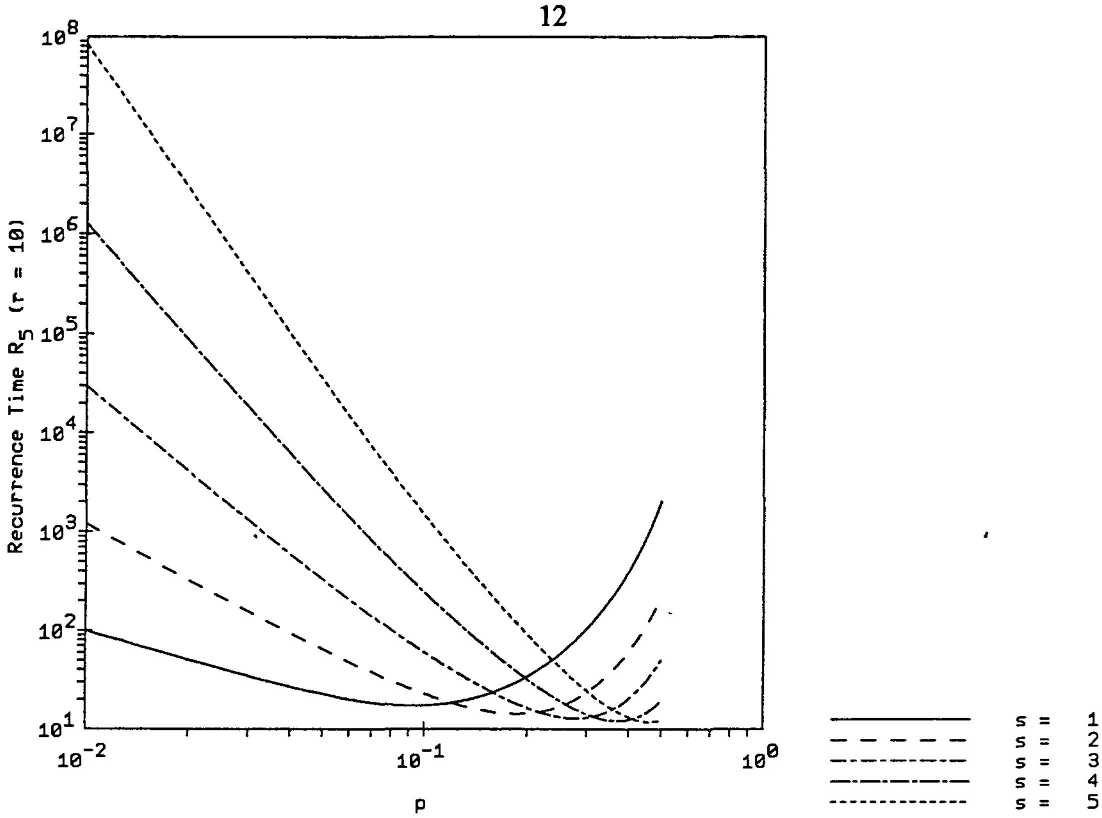


Fig.5. Recurrence time R_5 as a function of chance p for r = 10 and five values of parameter s.

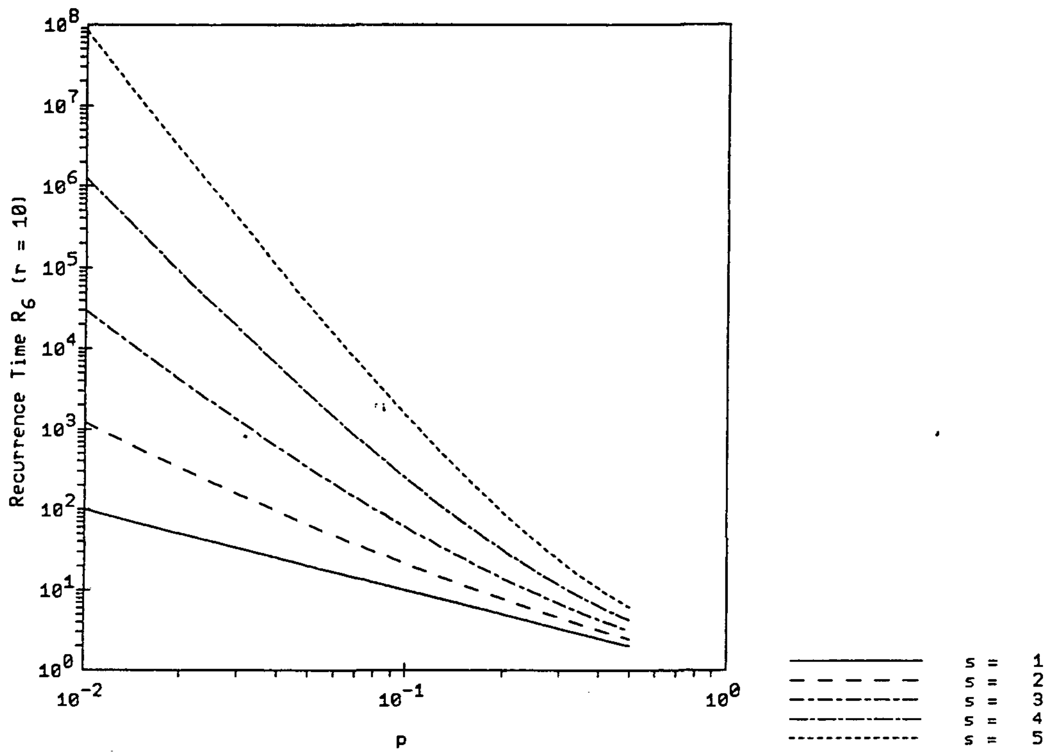


Fig.6. Recurrence time R_6 as a function of chance p for r = 10 and five values of parameter s.

3.7 Recurrence time R_7 for event E_7

Event E_7 occurs if position t is an element of an r-window with at least s ones. Of all events dealt with in this report event E_7 agrees probably most with the intuitive idea of recurrence time. This is best seen from the following property. If we meet in the z-series with the pattern

$$. . . . . 1\ 0\ 0\ 0\ 0 . . 0\ 0\ 0\ 0\ 0\ 1 . . . . \tag{29}$$

,i.e., a zero-window of arbitrary length preceded and succeeded by a one, then we may conclude that the left one is the last element of an r-window with at least s ones, while the right one is the first element of such a window.

In case of E_7 the combinatorics is quite complicated and we did not yet derive an analytic expression for R_7. We shall therefore point out how R_7 can be obtained numerically.

To determine the chance $P(z_t = 0)$ we consider all possible patterns of zeros and ones in the (2r-1)-window at positions (t-(r-1) , ... , t+(r-1)) in the x-series. We call such a window **allowed**, if it does not comprise an r-window with s or more ones. $P(z_t = 0)$ is obtained from addition of the chances of the allowed windows. Two allowed patterns with equal numbers of ones have equal chances. So, we have to determine the numbers c_n of allowed patterns containing n ones for $n = 0,......,N \equiv 2r-1$. Because the chance to find a pattern with n ones, and thus N-n zeros, is given by $p^n(1-p)^{N-n}$ we have

$$P(z_t=0) = \sum_{n=0}^{N} c_n\, p^n(1-p)^{N-n} \tag{30}$$

To determine $P(z_t = 1,\ z_{t+1} = 0)$ we apply the same strategy. Now we consider al possible patterns of zeros and ones in the 2r-window at positions (t-(r-1), ... , t+(r-1), t+r) in the x-series. Allowed patterns in this window are those which contain at least s ones at the first r positions (t-(r-1),...., t), and less than s ones in all other r-windows around positions t and t+1. From enumeration we may determine numerically the numbers d_n of allowed patterns with n ones. In terms of the coefficients d_n the denominator in (4a) is given by

$$P(z_t=1,\ z_{t+1}=0) = \sum_{n=s}^{N+1} d_n\, p^n(1-p)^{N+1-n} \tag{31}$$

From (30) and (31) we obtain an expression for R_7 in terms of the coefficients c_n and d_n . To this end it is convenient to introduce the polynomials

$$P_1(y) = \sum_{n=0}^{N} c_n\, y^n \tag{32}$$

and

$$P_2(y) = \sum_{n=s}^{N+1} d_n\, y^n \tag{33}$$

R_7 is then given by

$$R_7(p,r,s) = \frac{P_1(p/(1-p))}{P_2(p/(1-p))} \frac{1}{1-p} \tag{34}$$

Figure 7 shows R_7 for r = 10 and several values of s :

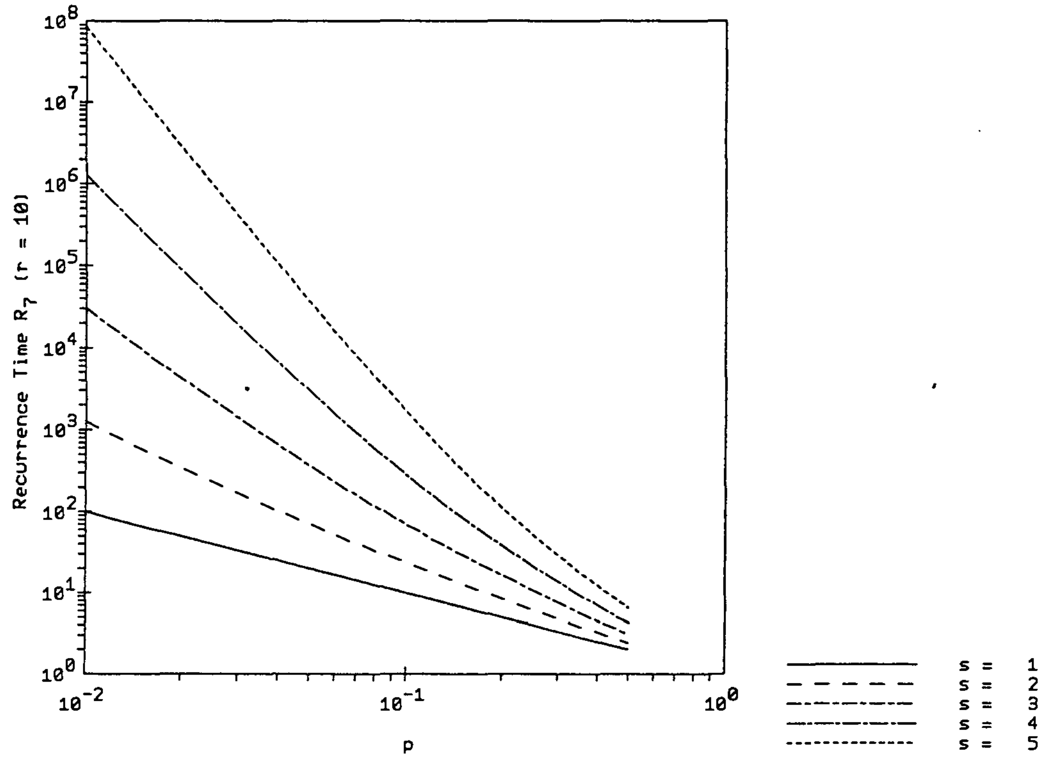


Fig.7. Recurrence time R_7 as a function of chance p
for r = 10 and five values of parameter s.

4. Practical considerations

In the preceding sections we considered seven types of events. In E_1 - E_4 we studied the occurrence of r consecutive ones, and in E_5 - E_7 the occurrence of a pattern with s ones at r consecutive positions. From the quartet E_1 - E_4 event E_4 is the one most close to the intuitive notion of such an event. The same holds for E_7 in the triple E_5 - E_7. However, in practice one is mostly interested in the recurrence time of rare occurrences, i.e., if $p \ll 1$. For small p we have

$$R_1 \approx R_2 \approx R_3 \approx R_4 \approx \frac{1}{p^r} \tag{35}$$

and

$$R_5 \approx R_6 \approx R_7 \approx \frac{1}{\binom{r-1}{s-1}} \frac{1}{p^r} \tag{36}$$

From (36) we see that the relative difference between R_6 and R_7 vanishes in the limit $p \downarrow 0$. So,

$$\lim_{p \downarrow 0} \frac{R_6 - R_7}{R_6 + R_7} = 0 \tag{37}$$

However, the coefficients of the first terms of the Laurent series of R_6 and R_7 are equal, but not the other terms. This implies that the absolute value of the difference of R_6 and R_7 will not vanish for small values of p :

$$\lim_{p \downarrow 0} (R_6 - R_7) = \infty \tag{38}$$

This effect is, e.g., present in the last two lines of table 3.

5. Applications

5.1 Northern Hemisphere Temperatures

In Kema (1992) a temperature series averaged over the Northern Hemisphere has been analyzed. This series contains annual mean data for the period 1881 - 1990. In Fig. 8a the series itself and in Fig. 8b the corresponding series of residuals is given. The latter series has been obtained by removing the trend calculated from the data in the period 1881 - 1980 and standardizing to unit variance. The residuals appear to be normally distributed, stochastically independent, and the variance of the series is homogeneous. In Fig. 8a this trend is presented together with its extrapolation over the decade 1981 - 1990. It is clear that this decade contains a large number of remarkably warm years. To establish how rare the occurrence of such a pattern is, we apply the theory of the preceding sections. The values of the residuals in the last decade and their percentiles are given in Table 2 in KEMA (1992). These percentiles are calculated from the residual series 1881 - 1980. These residuals are independently and normally distributed with zero-mean and constant variance. If we prescribe a certain percentile value, say q, a year is considered to be extreme if its temperature exceeds the temperature corresponding to the q percentile. This temperature level corresponds to the level L in formula (1). We calculated the recurrence times R_1 - R_7 for six values of q. The results are presented in Table 3. The value of r is set at 10 in case of R_5 -R_7, because we consider a decade. The values of p and s and the remaining values of r follow directly from Tables 1 and 2 in KEMA (1992).

Table 3. Recurrence times R_1 - R_7 in years for the Northern Hemisphere series of annual-mean temperatures for six threshold values L. The corresponding percentiles are given in the q row and the corresponding chances of crossing the threshold in the p row. The s value is the total number of crossings in the decade 1981 - 1990. The r value is the biggest number of successive crossings in this decade. Data from KEMA (1992).

L (^{0} C)	0.01	0.08	0.10	0.19	0.24	0.26
q (%)	50	66	75	90	95	97.5
p	0.50	0.34	0.25	0.10	0.05	0.025
r	5	5	4	4	4	3
s	8	7	6	6	6	5
$R_1(p,r)$	62	332	340	11,110	168,420	65,640
$R_2(p,r)$	61	331	339	11,109	168,419	65,639
$R_3(p,r)$	123	500	451	12.342	177,281	67,321
$R_4(p,r)$	58	328	337	11,107	168,417	65,638
$R_5(p,10,s)$	54	117	135	13,322	643,039	830,698
$R_6(p,10,s)$	53	116	134	13,322	643,039	830,698
$R_7(p,10,s)$	62	139	163	14,380	666,505	853,177

From Table 3 we see that the results for R_1, R_2, and R_4 are much alike. The corresponding event definitions do not differ much. The values of R_3 are somewhat different. The relative differences between the results for R_5, R_6, and R_7 are very small.

From the fifth column we learn that the occurrence of four consecutive crossings of the 95 percentile has a recurrence time of about 168,000 years. See the values of R_1 - R_4. This rare event happened in the decade 1980 - 1990. In the same decade six (non-consecutive) crossings of the 95 percentile occurred. This event is even more rare : the corresponding recurrence time, measured by R_5, R_6, and R_7, is about 650,000 years.

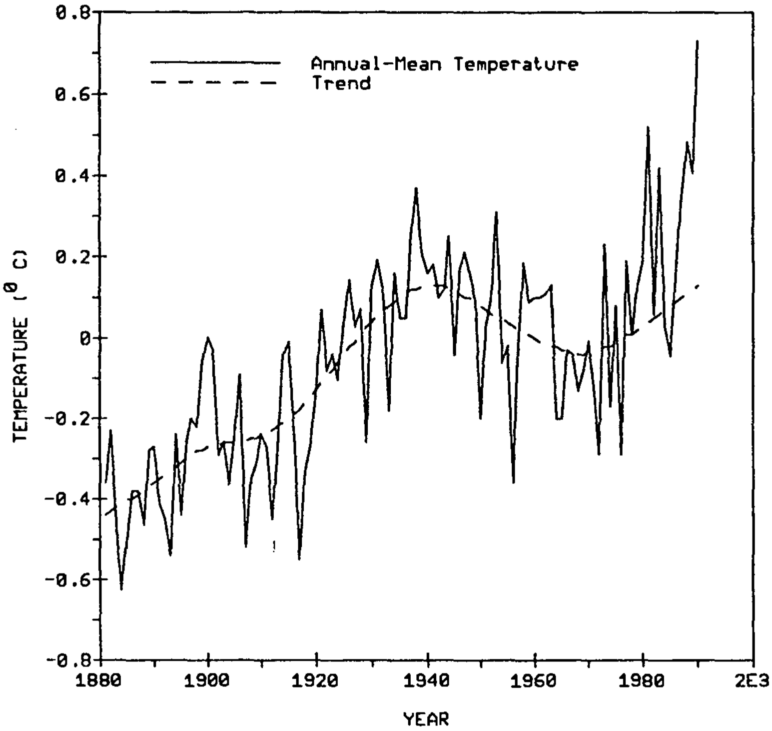


Fig.8a. Annual-mean air temperatures averaged over the Northern Hemisphere, 1881 - 1990. The calculation of the trend is dealt with in KEMA (1992).

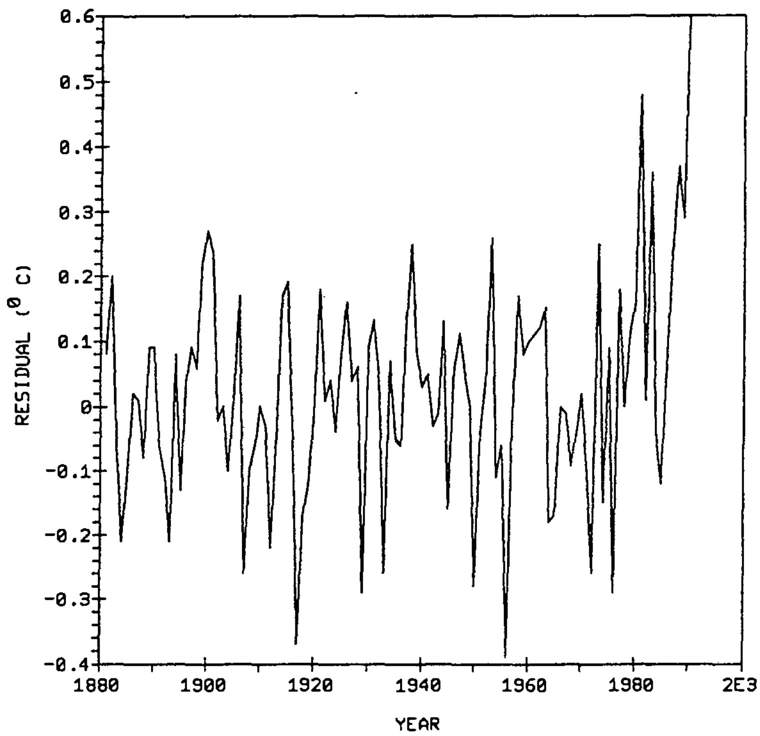


Fig. 8b. Residuals of the temperature series in Fig. 8a with respect to the trend drawn.

5.2 Temperatures from The Netherlands

Another series example, also analyzed in Kema (1992), is a very long series of annual mean temperatures in The Netherlands covering the period 1706 - 1990. The series is a reconstruction of temperatures at De Bilt. In Fig. 9a the series itself is given together with the trend, which is determined from the period 1706 - 1980 and extrapolated over the period 1981 - 1990. After removal of the trend the residual series in Fig. 9b results. In Table 4 the temperatures, the trend values, and the residuals in the last decade are given.

The percentiles q and corresponding chances p are calculated using the information in the period 1706 - 1980. We calculate the recurrence times R_1 - R_7 for one q value, namely q = 90 %. Together with the chance p = 0.1 we determine the 95 % confidence interval around this value. The recurrence times are calculated for p = 0.1 as well as the upper and lower bound of this interval. This gives an impression of the uncertainty in the calculated recurrence times. The results are given in Table 5.

Comparison of Tables 3 and 5 learns that the extreme events in the Northern Hemisphere series are much more outspoken than those in the The Netherlands series. This can be understood from the fact that temperature variations in The Netherlands are considerably tempered by the near presence of the sea.

Table 4. Temperatures, trend values, and residuals at De Bilt (The Netherlands) in the period 1981 - 1990.

year	temperature (^{0} C)	trend (^{0} C)	residual (^{0} C)
1981	9.21	9.31	-0.10
1982	10.05	9.31	0.74
1983	10.08	9.32	0.76
1984	9.45	9.32	0.13
1985	8.53	9.32	-0.79
1986	8.97	9.33	-0.35
1987	8.85	9.33	-0.48
1988	10.34	9.33	1.01
1989	10.74	9.33	1.41
1990	10.90	9.34	1.56

Table 5. Recurrence times R_1 - R_7 for annual-mean temperatures in The Netherlands for the threshold value L corresponding to the 90 % percentile. Between brackets the 95 % confidence intervals in chance p and thus in the recurrence times are given.

L (^{0} C)	0.82
q (%)	90
p	0.1 (0.07 - 0.14)
s	3
r	3
R_1(p,r)	1105 (3036 - 411)
R_2(p,r)	1103 (3035 - 410)
R_3(p,r)	1226 (3265 - 477)
R_4(p,r)	1102 (3034 - 409)
R_5(p,10,s)	61 (137 - 30)
R_6(p,10,s)	60 (137 - 28)
R_7(p,10,s)	70 (156 - 34)

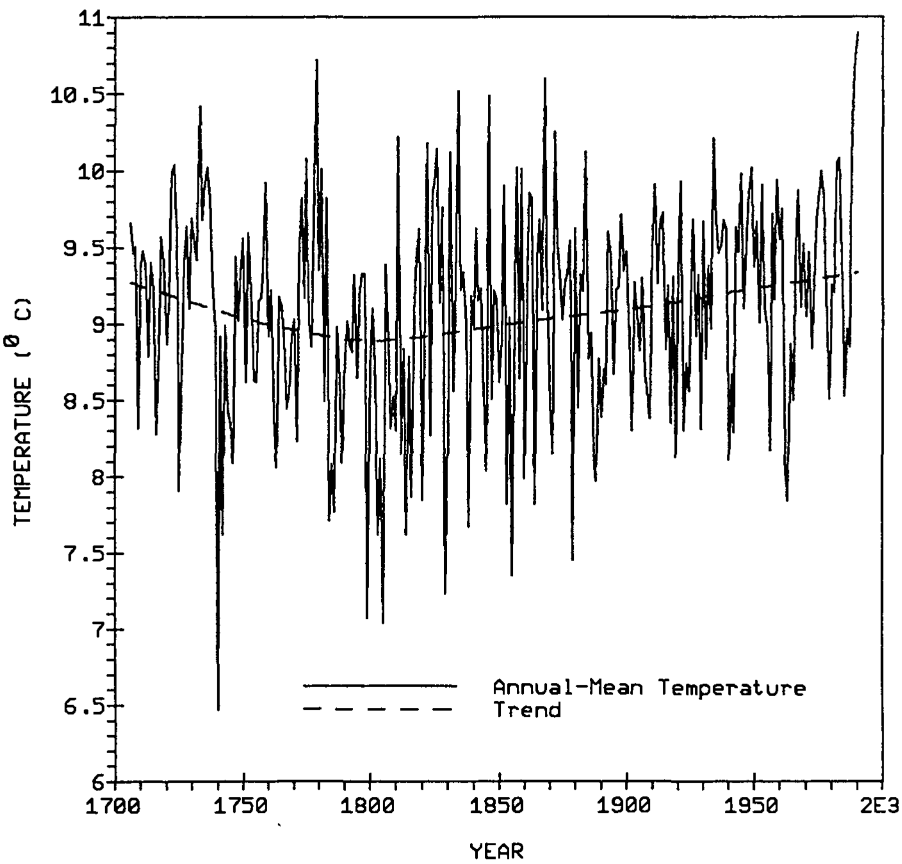


Fig.9a. Annual-mean air temperatures in The Netherlands, 1706 - 1990. The calculation of the trend is dealt with in KEMA (1992).

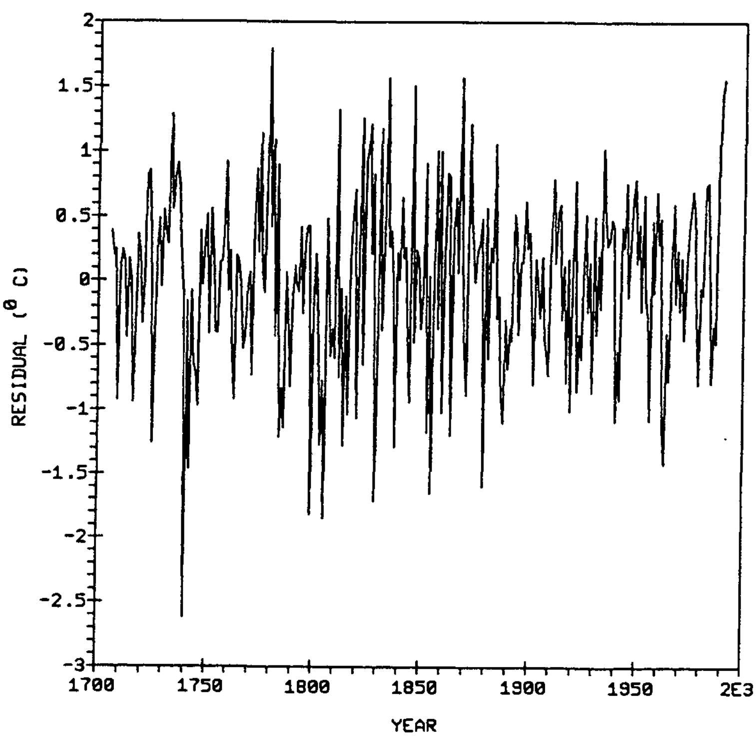


Fig. 9b. Residuals of the temperature series in Fig. 9a with respect to the trend drawn.

6. Conclusions

The notion of recurrence time is quite appropriate to quantify the rareness of the occurrence of a given pattern - a so-called **event** - in series of Bernoulli trials. The results depend on the event definition used. In our opinion event definition E_7 describes best the intuitive notion of recurrence time of rare events in series of Bernoulli trials. No analytical expression for the corresponding recurrence time R_7 is available, because the required combinatorics is too difficult. However, we have found that the numerical results for R_6 and R_7 are so similar, that in practice one can reliably use expression (26) for both R_6 and R_7.

The theory is illustrated by applying it to climatical series of annual-mean temperatures. It is shown that the series of temperatures averaged over the Northern Hemisphere shows a pattern with relatively high temperatures over the last decade. The recurrence times of the these patterns is in the order of 100,000 years, which is extremely long. These findings are a clear indication that the global warming observed for the last century has got a much more pronounced character over the last decade. The annual-mean temperatures in The Netherlands do not exhibit the same extremal behaviour. In these data a steady warming is found too, but in a much more tempered fashion, due to the nearby presence of the sea.

References

Feller, W., 1957, An introduction to probability theory and its applications. Vol. 1, Wiley and Sons.

KEMA,1992,(H.Visser and J.Molenaar),Trends and extreme values in climatological time series : what do they tell about climate change ?, report 92252-MOF 92-3683.

Kottegoda, N.T., 1980, Stochastic water resources technology, MacMillan Press Ltd., London, 1980.

Lawrence, A.J., and Kottegoda, N.T.,Stochastic modelling of river flow time series (with discussion), J.Roy.Stat.Soc. A, 1977, Vol. 140, Part 1, pp 1 - 47.

Visser, H., and Molenaar, J., 1990, Estimating trends in tree-ring data, Forest Science, Vol. 36, 87-100.

Visser, H., and Molenaar, J., 1993, Trend estimation and regression analysis in climatological time series : an application of structural models and the Kalman filter, Submitted to J. of Climate.

Wigley, T.M.L., and Barnett, T.P., 1990, Detection of the greenhouse effect in the observations. Climate change, The IPCC Scientific Assessment, J.T. Houghton, G.J. Jenkins, and J.J. Ephraums (eds), Cambridge University Press, 245-255.