Vol. 25: 3–13, 2003 | CLIMATE RESEARCH
Clim Res | Published October 8

How often can we expect a record event?

R. E. Benestad*

The Norwegian Meteorological Institute, PO Box 43, 0313 Oslo, Norway

ABSTRACT: This study applies a simple framework for analysing the incidence of record events. A test of this method on the global mean temperature yields results consistent with a global warming, where record-warm events are more frequent than for a stationary series. The record event analysis suggests that the number of record-warm monthly global mean temperatures is higher than expected, and that the number of record events in the absolute monthly maximum and minimum temperatures in the Nordic countries is slightly higher than expected from a null hypothesis of a stationary behaviour. Because the different station series are not strictly independent, it is difficult to resolve whether there is a significant trend in the warmest absolute monthly minimum temperatures in the Nordic countries. The behaviour of the maximum monthly 24 h precipitation is not distinguishable from the null hypothesis that the series consists of independent and identically distributed random variables.

KEY WORDS: Record-value statistics · Extremes · Temperature · Precipitation

1. INTRODUCTION

One important question about climate change is whether it results in changes in extreme weather or climate events. Extreme value analysis often involves fitting the tail of distributions such as general extreme value (GEV) and general Pareto (GPD) to empirical data (Coles 1999, von Storch & Zwiers 1999, Wilks 1995). It has been acknowledged that it is extremely difficult to detect trends in extreme weather and climate events because these are rare and because the observational records usually are short (IPCC 2002, Frei & Schär 2001). Small data samples often result in unstable fits and measurement errors can have a strong effect on the results. Frei & Schär (2001) introduced a framework based on event counts for studying trends in extreme rainfall in Switzerland, using a binomial distribution as a reference. They argued that trends in extreme events become increasingly difficult to determine as the events become more rare (e.g. more extreme).

A new approach for studying extreme events in climate research involves rank statistics where the timing of record-setting events is focused upon, and this type of analysis can be considered as a variant of Spearman rank correlation, Kendall's tau (Press et al. 1989, p. 536), or the Kendall *t*-statistic (Mann test) (Sneyers 1990). This approach compliments the count-analysis described by Frei & Schär (2001) which has a limitation regarding the most extreme events (e.g. record-values). An Internet search suggests that record event statistics is rarely used in climatological studies (one study was found in hydrology: Vogel et al. 2001, but there is no reference to this subject in Anderson 1958, Wilks 1995, von Storch & Zwiers 1999), although some literature on this subject does exist in other fields (Feller 1971, Glick 1978, Nevzorov 1987, Nagaraja 1988, Ahsanullah 1989, Ahsanullah 1995, Arnold et al. 1998, Balakrishnan & Chan 1998, Bairamov & Eryilmaz 2000, Raqab 2001). Some of the existing theories are much more complicated than required for studying record events in observed climatological series and some are too theoretical. This paper will introduce a simple probabilistic record event framework for analysing extreme climate and weather events. A null hypothesis assumes a homogeneous (no change in observation practice or instrumentation) and stationary (no long-term trend) series, i.e. a series consisting of independent and identically distributed (iid) random variables (von Storch & Zwiers 1999, Raqab 2001). One

*Email: rasmus.benestad@met.no

advantage of this approach is that it is easy to compare different series as their respective rank statistics are not affected by whether the different series are characterised by different probability distribution functions (p.d.f.).

2. METHODS AND DATA

The probability, $p_n(1)$, that the nth observation of a series $\vec{x} = x_1, x_2, \ldots x_n$ has a higher value than the previous observations $[p_n(1) = \Pr(x_n > x_i|i < n)]$ can be expressed as:

$$p_n(1) = \frac{1}{n} \quad (1)$$

provided the values in series are iid random variables. The present notation uses $p_n(0)$ to denote the probability of no new record event at the end of a series of length n whereas $p_n(1)$ refers to the probability of a new record event.

For many parallel observations, it can be shown that the probability of seeing a new record event depends on the number of simultaneously taken measurements (the symbol 'N' will be used for the number of independent simultaneous observations, whereas 'n' represents the length of each series). If there are N different independent observations, for instance from different locations, made over a time interval $i = 1, 2, \ldots, n$, then the probability for seeing at least one new record event at time n can be estimated by considering the probability that there is not a new record at time n in a single series $p_n(0) = [1 - p_n(1)]$. The probability that there are no new records in N series is $p_n(0)^N$ (this is equivalent to considering a special case of the binomial distribution). Since the inverse of no events in N independent realizations, $1 - p_n(0)^N$ is the probability of at least one record event, we can write the expression for the probability for the values in the nth observation setting at least one new record in N independent locations as:

$$p_n^N(1) = 1 - \left(1 - \frac{1}{n}\right)^N \quad (2)$$

Fig. 1 gives a graphical presentation of the solutions of Eq. (2). According to this expression, the probability for setting a new record increases with the number of contemporary independent observations. With more than 100 parallel time series of annual observations, there is still a high probability of seeing new records after 100 years.

This type of record event analysis can be applied to a series using chronological order, but it can also be applied to data using reversed chronological order starting with the most recent observations. The analysis of series with reversed order is in this paper referred to as 'backward', as opposed to a 'forward' analysis. A comparison between the 'forward' and 'backward' analyses can give useful information about the incidence of the record events in case of an extreme record-value in the early part of the series. If x_1 is the maximum value in a series ($\vec{x} = x_1, x_2, \ldots, x_n$, then the 'forward' method yields only one record event and thus suggests a rejection of the null hypothesis (too few records). The 'backward' method, on the other hand, may reveal that the record event statistics nevertheless does follow the null hypothesis and that x_1 being the maximum is due to chance. Similar results from the 'forward' and 'backward' analyses improve the confidence in the conclusion that the data are consistent with the null hypothesis.

It is possible to relate the theoretical probability to the empirical data by utilising the expectation value $E_n = N p_n(1)$. In this paper, we will distinguish between the expectation value E_n for N independent series and the expected number of record events $\varepsilon(n)$ in a single series with n observations. Furthermore, there is a distinction between the observed number of simultaneous record events $\hat{E}_n$ at time n and the theoretical expectation value E_n (similar for the expected number of records: $\varepsilon(n)$ and $\hat{\varepsilon}(n)$). For a number of independent stationary series with length n, a count of the number of new parallel record-values at time n, $\hat{E}_n$, is taken as an estimate for the expectation value E_n. The quantity $\hat{E}_n/N$ is henceforth referred to as the record density, and according to the expression for the expectation

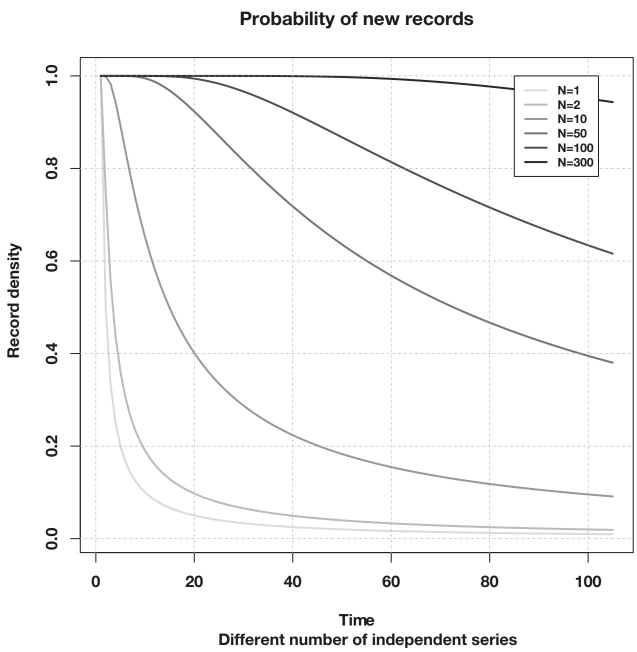

Fig. 1. Probability as a function of observation time of seeing at least one record event in N independent series (Eq. 2)

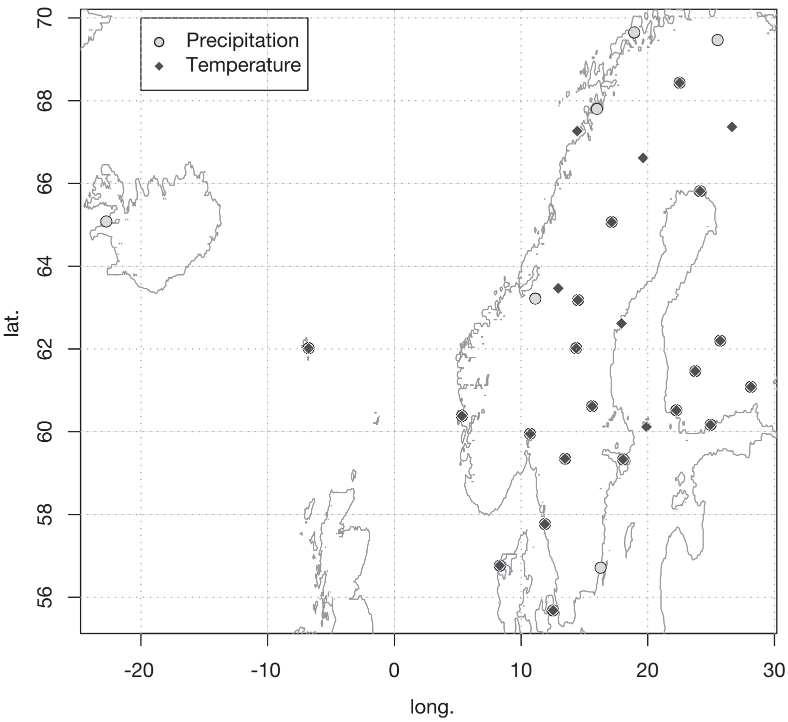

Fig. 2. Locations of the absolute monthly maximum, minimum temperature measurements (◆) corresponding to the Nordklim codes 112 and 122. The temperature data were from: Oslo, Bergen, Bodø, Værnes (Norway); Copenhagen, Vestervig (Denmark); Turku, Helsinki, Tampere, Jyväskylä, Lappeenranta, Sodankylä, Mariehamn (Finland); Torshavn (Faeroes); Stockholm, Gothenburg, Stensele, Karesuando, Falun, Karlstad, Sveg, Östersund, Härnösand, Haparanda, Jokkmokk (Sweden). (◯) Locations where the maximum monthly 24 h precipitation (Nordklim code 602) were taken from: Oslo, Tromsø, Bergen, Kråkmo, Karasjok, Lien (Norway); Stockholm, Haparanda, Gothenburg, Stensele, Karesuando, Falun, Karlstad, Sveg, Östersund, Kalmar (Sweden); Helsinki, Turku, Jyväskylä, Lappeenranta (Finland); Copenhagen, Vestervig (Denmark); Torshavn (Faeroes); Stykkisholmur (Iceland)

value, it can be compared directly with the probability of seeing new records.

In this study the monthly series are stratified into the 12 calendar months in order to avoid comparing January anomalies characterised by large magnitudes with July anomalies that are characterised by smaller deviations from climatology. In other words, each location is represented by 12 different series. Since the month-to-month persistence in northern Europe is low, these can be treated as 12 independent series.

Because of the $^1\!/_n$ dependency in Eq. (1), the probabilities are expected to converge towards zero with time, and it becomes difficult to compare the numbers visually for long time-series. However, this convergence follows a linear trend on a logarithmic scale:

$$\ln[p_n(1)] = -\ln(n) \qquad (3)$$

It is important to keep in mind the distinction between the probability for a new record event $p_1(n)$, the p.d.f. describing distribution of record values (e.g. see Bairamov & Eryilmaz 2000), and the p.d.f. for each of the iid random variables (e.g. Gaussian or EVD). For single series with n random values drawn from the same population, the expected number of record-values is $\varepsilon(n) = \sum_{i=1}^{n} p_i(1) = \sum_{i=1}^{n} 1/i$.

Monte-Carlo simulations were used in this study to estimate confidence limits, and 2 slightly different approaches were used for this means to reflect the different types of tests. The approach used for the case of parallel series (henceforth referred to as 'MC I') involved series of normally distributed random numbers with no time dependency [i.e. a 'white noise' process (Wilks 1995, Wallace 1996, von Storch & Zwiers 1999)], unless otherwise stated. The estimation of confidence intervals involved Monte-Carlo simulations with 1000 data matrices similar to the actual data (size $N \times n$, i.e. $1000 \times 25 \times 12$ series), and the results consisted of 1000 surrogate record densities, hence the results consisted of fractional numbers. The derived null-distribution for the expected number of records was taken as the sum of record densities: $\hat{\varepsilon}(n)_{MC} = \sum_{i=1}^{n} \hat{E}_i$.

In order to verify that the expected number of records $\varepsilon(n) = \sum_{i=1}^{n} 1/i$ gives a good estimate of the actual number of records, a different set of Monte-Carlo integrations (henceforth referred to as 'MC II') was carried out for 1000 stochastic [white noise and AR(1) process (Wilks 1995)] series with different lengths n. Although this set of simulations only involved 1000 counts for each series x_1, x_2, … x_n of iid random values, the counting process was repeated 40 (n = 1,5, ..., 200) times. The number of record events for each of the 1000 simulation and each value of n were counted and stored in a matrix $\mathfrak{R}$. Thus MC II results consisted of discrete numbers, and the smaller sample (1000) is expected to yield a greater spread than the MC I simulations (300 000).

A robust and unbiased estimation of the record-density requires N independent series, each containing n independent realizations. Hence, to execute the objective tests, each station series should have zero serial correlation and the stations should be independent of each other. Over northern Europe, the month-to-month and year-to-year correlations are low, hence each series can be treated as consisting of independent realizations. Although the short-term variations exhibit a low degree of persistence, long-term trends may still affect the results from these tests.

A Pearson's χ^2-test (Wilks 1995, p. 133, Eq. 5.18) was applied to $p_i(1)$ and $\hat{E}_i/N$; $i = 1, 2, \ldots, n$ in order to examine whether the observed record-densities were significantly different to expected probabilities of seeing new record events. The 95% confidence limit for the χ^2-test values was estimated repeating the test 1000 times using iid random numbers instead of the actual observations (MC I).

The data examined here included the monthly global mean temperature from Jones et al. (1998a,b) and absolute monthly maximum and minimum temperature as well as 24 h accumulated precipitation from the high-quality homogenized Nordklim data set (Tuomenvirta et al. 2001). The global mean temperature spanned 1856–2002 (147 yr), the monthly maximum/minimum temperature covered 1908–1999, and the precipitation was taken from the period 1895–1999. The locations of the Nordklim stations used in this study are shown in Fig. 2.

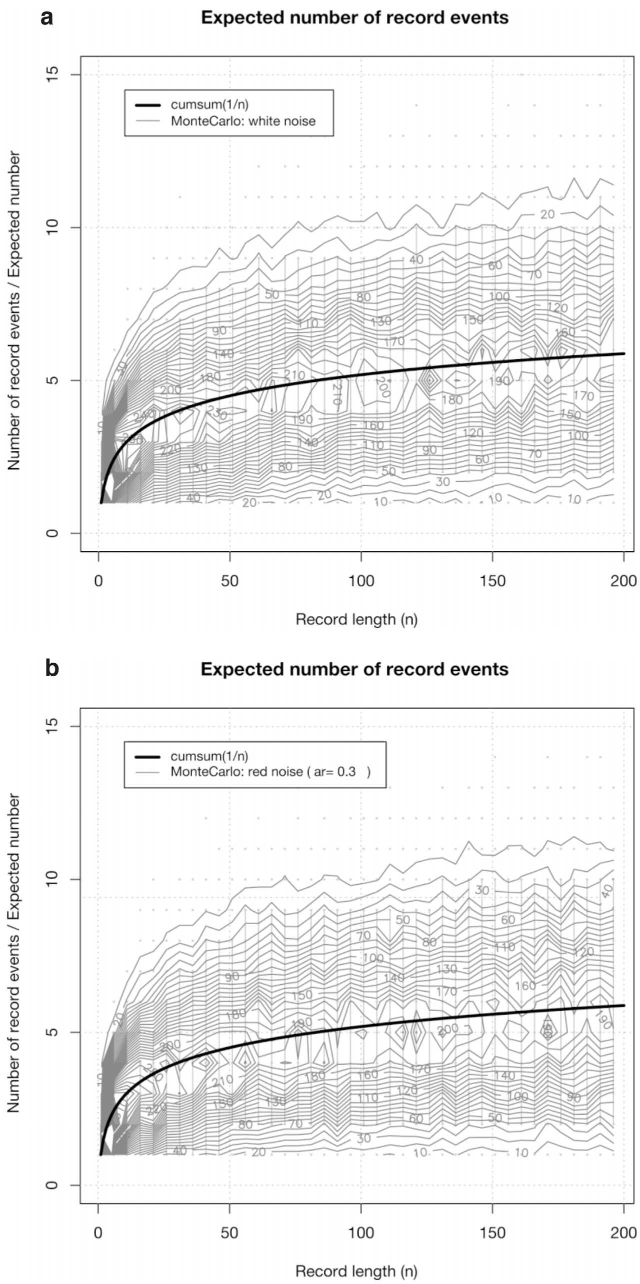

Fig. 3. Comparison between the expected number $\varepsilon(n)$ and the results from Monte-Carlo (MC II) simulations for a single series assuming a white-noise process (a) and an AR(1) process process with a serial correlation of 0.3 (b). The thick solid line indicates the expected number of records $\sum_i(1/i)$ in a series of iid random variables. The contours show the number of times that a series of length n (x-axis) produced m (y-axis) record events in 1000 Monte-Carlo simulations (i.e. the contours show the counts stored in in the matrix $\Re$)

3. RESULTS

3.1. Expected number of record events

In order to ensure that the $\varepsilon(n) = \sum_{i=1} 1/i$ gives a good estimate of the expected number of record events, 2 sets of MC II Monte-Carlo experiments were carried out (Fig. 3). The results from these simulations suggest that $\varepsilon(n)$ gives a good description for series with iid random variables, regardless of the length of the series. The probability distributions of the Monte-Carlo integration results are skewed with longer upper tails, and the mean number of records is expected to be slightly higher than the location of the maximum probability. The distribution of the Monte-Carlo results is not significantly affected by the serial correlation.

3.2. Inter-station dependencies and temporal autocorrelation

Dependency between different climate station records presents a concern, as inter-station dependencies will have similar effects on the test results as so-called 'overdispersion' discussed in Frei & Schär (2001). To examine the effects of serial and spatial correlation on the estimation of the confidence limits, a set of MC I Monte-Carlo simulations was used to simulate both the effect of temporal autocorrelation (assuming an AR(1) process with $\rho = 0.2$) as well as inter-station dependencies. For 300 truly independent series of length $n = 92$, cases with $\hat{E}_n = 4$ are more frequent than cases with $\hat{E}_n = [1, 2, 3]$. Fig. 4a shows results from these simulations for 25 stations where different linear trends have been added to series of white noise. The histograms show that the statistics of the number of coinciding events $\hat{E}_n$ are indeed affected by the trends. Since the purpose of this study is in fact to detect such trends, the test results shown in Fig. 4a suggest that the method is able to identify long-term changes in the record event statistics.

Inter-station dependencies affects the distribution of the number of parallel record events $\hat{E}_n$ (Fig. 4b). Fig. 5 shows a similar analysis for 2 cases where the trend in the monthly maximum temperature was retained and subtracted. The histogram showing the record event statistics for monthly maximum temperature shows a high frequency of $\hat{E}_n = [1, 2, 3]$, both for the original data as well as for de-trended data. There are some indications that the trend may account for some of the $\hat{E}_n = [1, 2, 3]$ cases, but the results for the de-trended series suggests there are also inter-station dependencies.

One set of Monte-Carlo experiments was carried out to examine how the χ^2-statistics and $\hat{\varepsilon}(n)$ are affected by inter-station dependencies. One test was for the extreme case that all 25 stations were inter-dependent, and this set of experiments consisted of repeating a set of 12 independent series of iid random values 25 times (Fig. 6). Another less severe case was also examined, corresponding to 5 stations being completely independent of each other (Table 1). Alternatively, the inter-dependency can be accounted for by imposing non-zero spatial correlations. However, the advantage of the former approach is that it is simple and ensures exactly the same timing of record events in the affected series (here we are only interested in the timing of record events). Furthermore, an inter-series correlation does not represent the aggregated series well, since the zero-lag (simultaneous) observations show much higher crosscorrelations than the lagged series do (Table 2). Table 1 lists the number

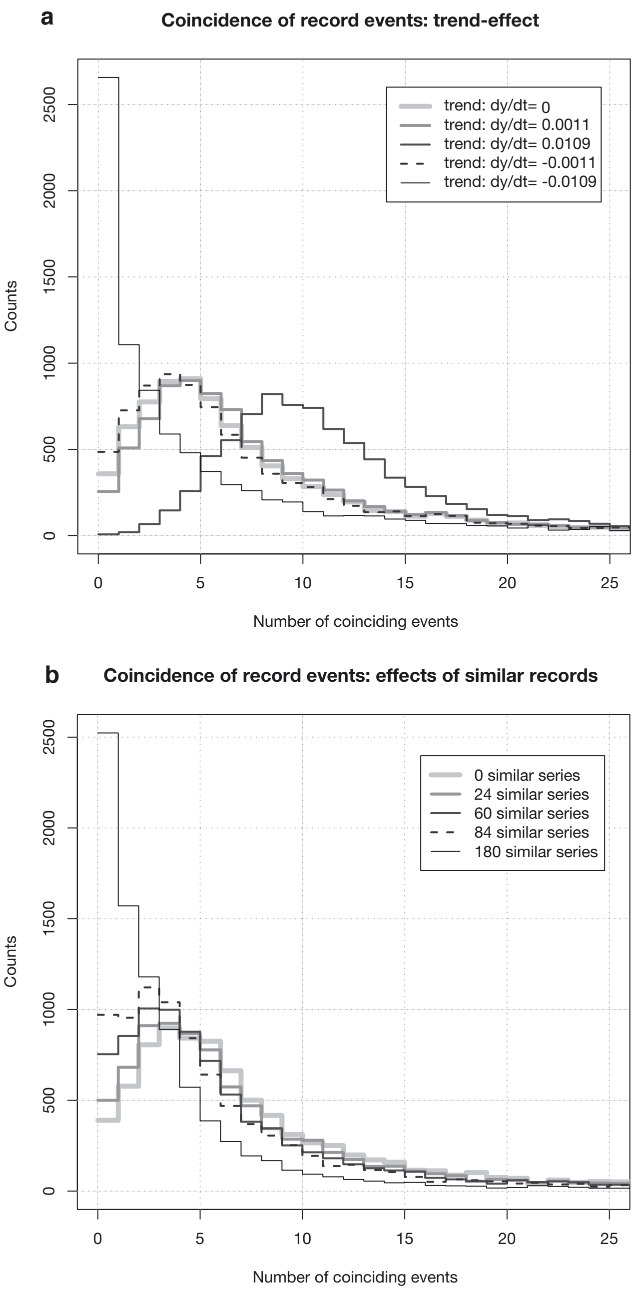

Fig. 4. Histogram showing the frequencies of $\hat{E}_n = 1, 2, 3...$ for different Monte-Carlo experiments where a trend has been imposed onto the random numbers (a) and where subsets of the series are identical (b). Random incidences for n = 92, N series = 300, N tests = 100

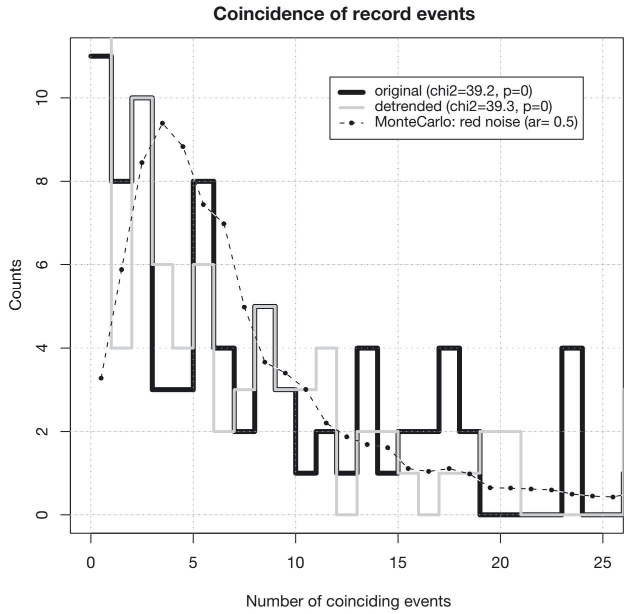

Fig. 5. Histogram similar to Fig. 4 for the monthly maximum temperature, comparing $\hat{E}_n$ derived from the original data and corresponding de-trended series. Empirical data vs random incidences for n = 92 and 300 series

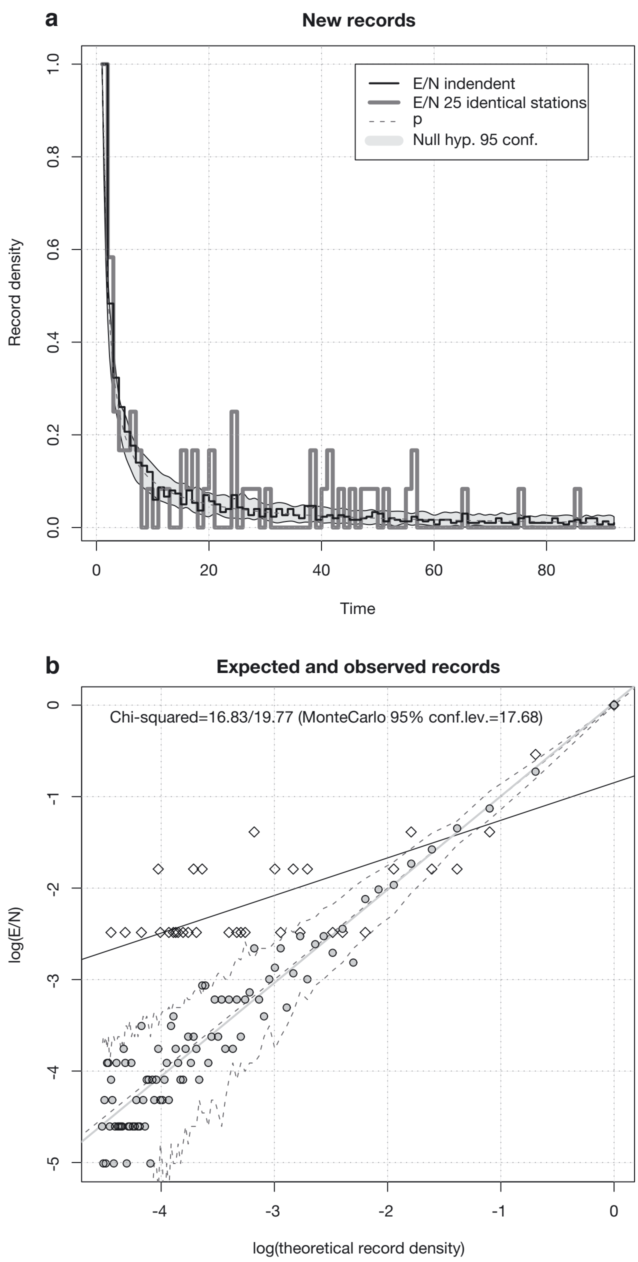

Fig. 6. Record-value statistics for the MC I Monte-Carlo simulations where one set represents 25 × 12 truly independent series (independent) and another set where 50 of the series are identical (dependent). Panel (a) shows the record-density $\hat{E}_n/N$ plotted with $p_n(1)$. The grey curve shows the record-value analysis applied to the same record, but starting at the end and going backward in time. Panel (b) shows a scatter-plot where ln(p) is plotted along the x-axis and ln($\hat{E}_n/N$) on the y-axis. (◇) Results from the 'dependent' test; (○) iid case. The grey region in panel (a) and dashed lines in (b) show the 95% confidence region derived from Monte-Carlo (MC I) simulations. A straight dashed line in (b) also marks the diagonal $y = -x$, and the black and grey straight lines show the best linear fit between log($\hat{E}_n/N$) and n

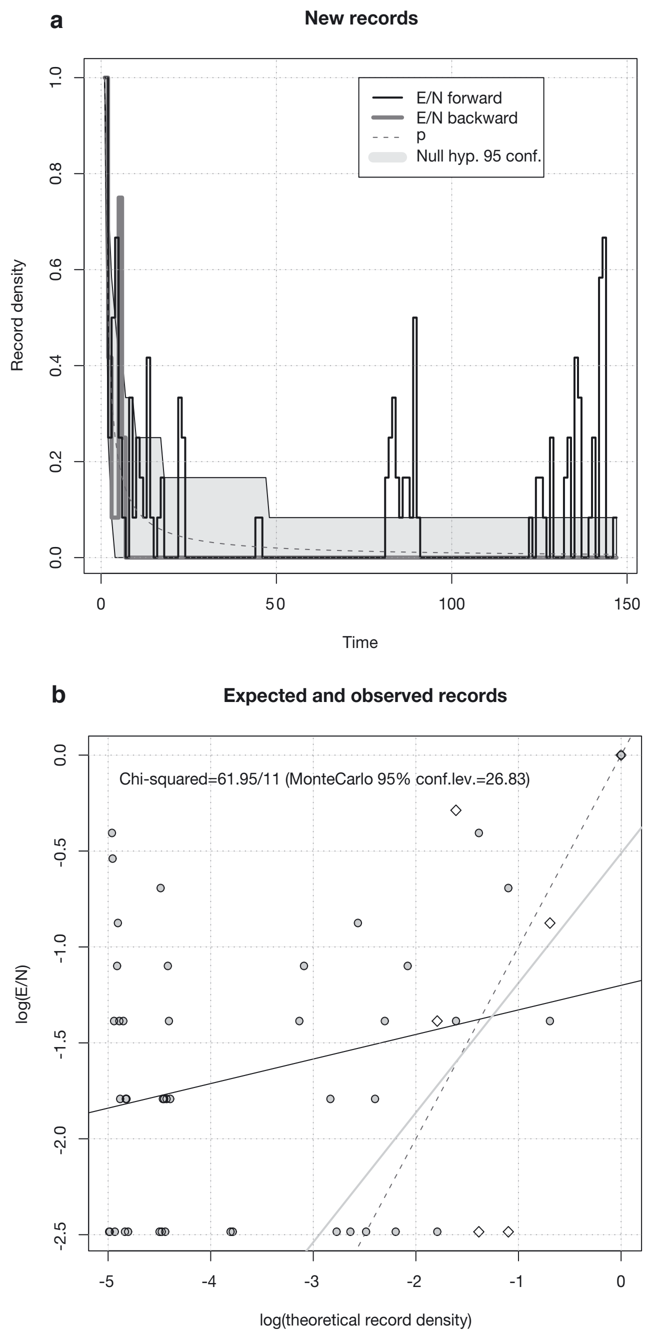

Fig. 7. Record-value statistics for the Jones et al. (1998a,b) global mean temperature. Panel (a) shows ($\hat{E}_n/N$) plotted with $p_n(1)$. The grey curve shows the record-value analysis applied to the same record, but starting at the end and going backward in time. Panel (b) shows a scatter-plot where ln(p) is plotted along the x-axis and ln($\hat{E}_n/N$) on the y-axis. (◇) Results from the 'backward' test; (○) forward case. The grey region in panel (a) and dashed lines in (b) show the 95% confidence region derived from Monte-Carlo (MC I) simulations. A straight dashed line in (b) also marks the diagonal $y = -x$, and the black and grey straight lines show the best linear fit between log($\hat{E}_n/N$) and n. 12 monthly temperature series (Jan–Dec)

Table 1. Mean observed and expected number of record events. The length of the temperature series is 92 and the precipitation analysis covers 105 yr. The second column gives the mean observed number of record events, the third the 95% confidence region derived through Monte-Carlo simulations, the fourth the expected value, and the fifth the results from a Pearson's χ^2-test. The numbers in the parenthesis indicate the null hypothesis 95% confidence level derived from a Monte-Carlo simulations using ($N = 1000 \times 300$) normally distributed random numbers

	$\hat{\varepsilon}(n) = \sum_i^n \hat{E}_i/N$	95% CI	$\sum_{i=1}^n 1/i$	χ^2
Global mean temperature	10.8	4.5–6.8	5.6	62.0 (23.8)
Global mean temp. (backward)	2.6	4.5–6.8	5.6	11.0 (23.8)
Abs monthly max temp	5.3	4.9–5.3	5.1	18.0 (17.5)
Abs monthly max temp (backward)	4.3	4.9–5.3	5.1	14.8 (17.5)
Abs monthly min temp	4.9	4.9–5.3	5.1	16.4 (17.5)
Abs monthly min temp (backward)	5.7	4.9–5.3	5.1	19.1 (17.5)
Max month 24 h precip	5.2	5.0–5.4	5.2	17.9 (18.6)
Max month 24 h precip (backward)	4.8	5.0–5.4	5.2	16.6 (18.6)
MC I all indep. autocor $\rho = 0.0$	5.1	5.0–5.4	5.1	16.5 (17.4)
MC I all dep. autocor $\rho = 0.0$	5.5	5.0–5.4	5.1	18.4 (17.4)
MC I all indep. autocor $\rho = 0.2$	5.1	5.0–5.4	5.1	16.8 (17.7)
MC I all dep. autocor $\rho = 0.2$	5.5	5.0–5.4	5.1	19.8 (17.7)
MC I 5 indep. autocor $\rho = 0.0$	5.3	5.0–5.4	5.1	17.8 (17.5)
MC I 5 indep. autocor $\rho = 0.2$	5.3	5.0–5.4	5.1	17.5 (17.5)

of records obtained with these simulations as well as the estimated confidence intervals. The experiments with the Monte-Carlo simulations suggest that a modest serial correlation does not have any significant effect on the tests and that the χ^2 statistic is affected, but is not very sensitive to inter-station dependence. The values for $\hat{\varepsilon}(n)$ do not appear to be very sensitive to inter-station dependence.

3.3. Global mean temperature

The record event analysis was applied to the global mean temperature record. Here the global mean temperature has been split (stratified) into 12 parallel series representing each calendar month. According to the Monte-Carlo simulations, one would expect to see a new record being set 4.5 to 6.8 times (Table 1). The number of record events statistics (here using the record-density: $\hat{\varepsilon}(n) = \sum_i^n \hat{E}_i/N$), on the other hand, suggests that new records have been set 11 times in the 'forward' analysis, but only 3 in the 'backward' analysis; both estimates are well outside the 95% confidence region. Hence, the global mean temperature development since 1856 is not consistent with the null hypothesis. This discrepancy can be explained in terms of the global warming or interdecadal variability and an associated shift in the record-warm global mean temperature statistics. (This analysis does not answer whether this non-stationarity is anthropogenic or related to natural causes, or if there will be more records in the future.)

Fig. 7a shows a comparison between the probability of seeing a new record and the record-density derived from monthly mean global mean temperature. There is a long-term upward trend in the global mean temperature which results in more frequent incidents of new record events than if there were no trend.

Fig. 7b shows a comparison between $\ln[p_n(1)]$ and $\ln(\hat{E}_n/N)$, and a straight diagonal line with unit slope (shown as a dashed line) through the origin would indicate a good agreement between the values expected under the null hypothesis and the empirical values. The best-fit to the data based on Eq. (3) is shown as a black solid line for the 'forward' and a grey line for 'backward' analysis. The clear deviation of these fitted lines from the diagonal and the large scatter in Fig. 7b, however, indicate that the empirical data do not agree well with the null hypothesis based on the assumption of data being iid.

3.4. Nordic station data

Fig. 8 shows the record-value analysis applied to the absolute monthly maximum temperature from the high-quality homogenized Nordklim data set (Tuomenvirta et al. 2001). The analysis was applied to the 12 calendar months from 25 locations in the Nordic countries (Fig. 8), i.e. $N = 300$ different series. A large sample of independent series gives a better estimate of $\hat{E}_n$ than a single series (central limit theorem), and in this case, the sample size is limited by non-zero correlation between series from nearby locations (Fig. 2).

The comparison between $p_n(1)$ assuming iid random values and the empirical record events in Fig. 8a indicates a better agreement than for the global mean temperature. The scatter-plot in Fig. 8b shows that the points are close to linear, hence a good agreement

Table 2. Inter-station and inter-monthly cross-correlations. Columns show the absolute cross-correlation minima, the first quartile (25% percentile), the median, mean, the third quantile (75% percentile), and absolute maxima

	Min.	1st quartile	Median	Mean	3rd quartile	Max.
Abs monthly max temp						
All	–0.83	–0.03	0.05	0.07	0.14	0.96
Lag-0 cross-correlations:						
January	–0.09	0.34	0.55	0.51	0.69	0.93
February	–0.17	0.45	0.6	0.56	0.71	0.94
March	–0.27	0.5	0.59	0.54	0.67	0.88
April	–0.04	0.33	0.51	0.47	0.62	0.86
May	–0.25	0.29	0.45	0.41	0.59	0.86
June	–0.14	0.31	0.49	0.46	0.63	0.86
July	–0.08	0.2	0.35	0.36	0.53	0.83
August	–0.13	0.27	0.42	0.43	0.63	0.96
September	–0.18	0.24	0.44	0.41	0.58	0.86
October	–0.17	0.27	0.43	0.4	0.55	0.83
November	–0.14	0.22	0.42	0.39	0.55	0.96
December	–0.07	0.3	0.47	0.46	0.65	0.93
Serial correlation:						
Oslo	–0.2	–0.04	0.035	0.06	0.18	0.38
Abs monthly min temp						
All	–0.89	–0.01	0.07	0.10	0.17	0.92
Lag-0 cross-correlations:						
January	–0.3	0.45	0.58	0.53	0.69	0.89
February	–0.24	0.46	0.63	0.54	0.72	0.9
March	–0.2	0.45	0.6	0.53	0.7	0.88
April	–0.09	0.35	0.5	0.46	0.62	0.85
May	–0.01	0.34	0.47	0.44	0.56	0.83
June	–0.08	0.27	0.39	0.37	0.5	0.74
July	–0.25	0.14	0.29	0.28	0.4	0.69
August	–0.21	0.2	0.34	0.32	0.46	0.75
September	–0.28	0.3	0.41	0.38	0.53	0.84
October	–0.02	0.43	0.56	0.53	0.69	0.89
November	–0.07	0.44	0.53	0.51	0.64	0.89
December	0	0.5	0.62	0.58	0.7	0.92
Serial correlation:						
Oslo	–0.17	0.01	0.12	0.11	0.2	0.45
Max monthly 24h precip						
All	–0.39	–0.06	0.01	0.01	0.08	0.59
Lag-0 cross-correlations:						
January	–0.32	0.01	0.11	0.11	0.21	0.58
February	–0.3	0.02	0.12	0.12	0.23	0.54
March	–0.22	–0.02	0.07	0.079	0.17	0.57
April	–0.22	–0.023	0.07	0.073	0.16	0.53
May	–0.24	–0.02	0.08	0.077	0.17	0.55
June	–0.26	–0.03	0.05	0.052	0.13	0.38
July	–0.28	–0.02	0.05	0.053	0.14	0.41
August	–0.23	–0.02	0.05	0.058	0.15	0.42
September	–0.3	–0.03	0.05	0.062	0.15	0.44
October	–0.31	0	0.12	0.12	0.23	0.57
November	–0.3	–0.03	0.06	0.071	0.18	0.57
December	–0.24	–0.01	0.08	0.088	0.18	0.59
Serial correlation:						
Oslo	–0.24	–0.03	0.025	0.021	0.07	0.26

between the null hypothesis and the empirical data. However, there is a slight tendency of more record events towards the end of the observations in Fig. 8a and a greater spread towards the lower left corner in Fig. 8b. The scatter-plot shows the difference between the 'forward' and the 'backward' analyses, revealing that the 'forward' analysis gives slightly higher values of $\hat{E}_i/N$ for higher values of n, i.e. later times. Although this difference may not be statistically significant, it is consistent with a small increase in the frequency of incidence of record events. Table 1 lists the mean observed and expected number of record events for the 92 yr series. According to the table, the observed number of events is higher than the expected value

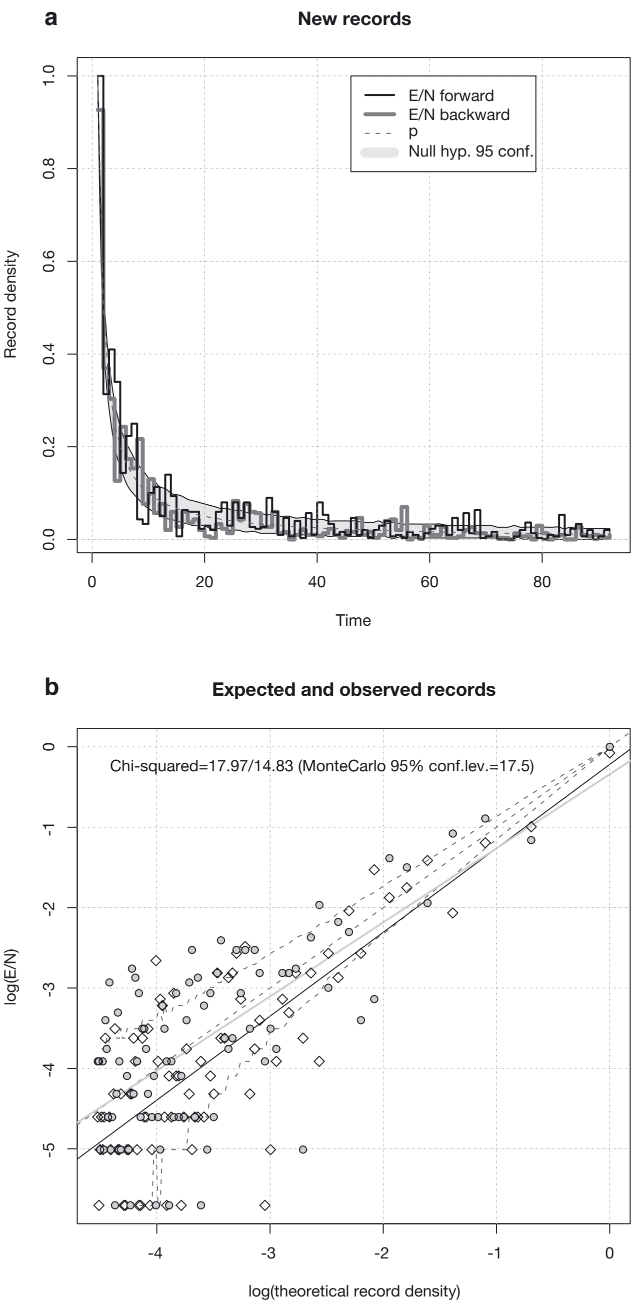

Fig. 8. Record-value statistics like Fig. 7, but for the absolute monthly maximum temperature in the Nordic countries from the Nordklim data set (Tuomenvirta et al. 2001) (code 112). (The locations are shown in Fig. 2.) The additional dashed lines shown in panel (b) mark the 95 % confidence region corresponding to the shaded area in panel (a). 25 × 12 monthly maximum temperature series

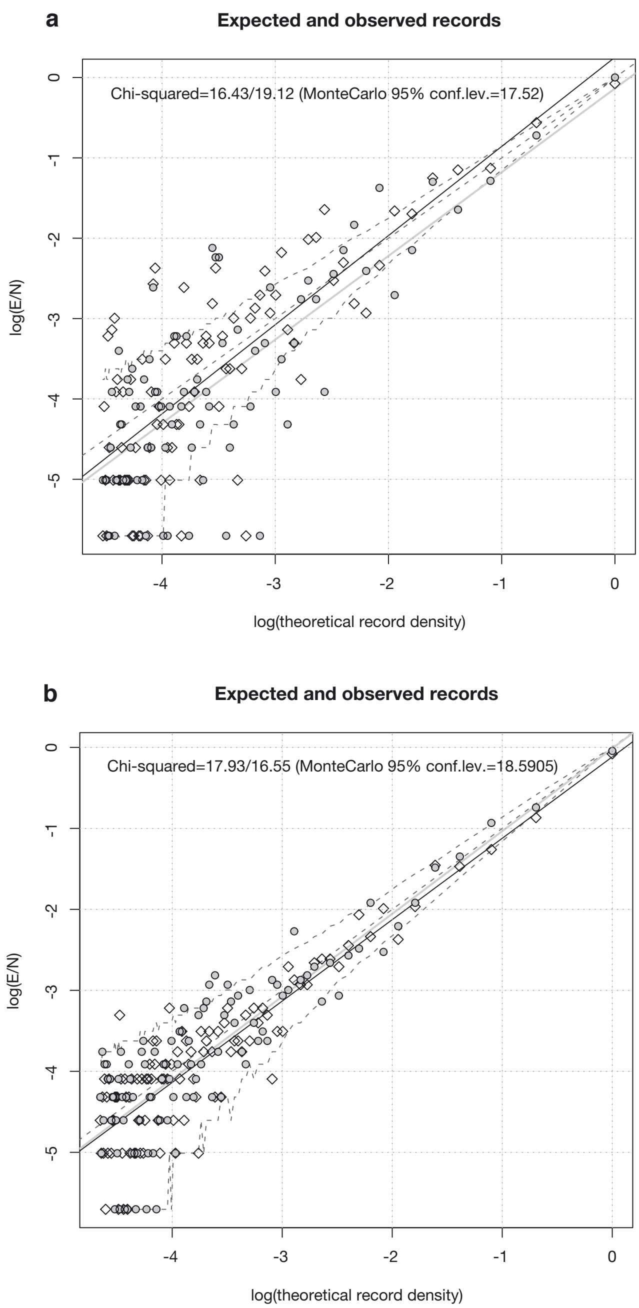

Fig. 9. Record-value statistics like Fig. 8, but for the absolute monthly minimum temperature (a) and maximum monthly 24 h precipitation (b) (Nordklim codes 122 and 602). (The locations are shown in Fig. 2)

derived using the expression $\varepsilon(n) = \sum_{i=1}^{n} 1/i$. The confidence interval for the record-density according to the Monte Carlo simulations is estimated to be 4.9 to 5.3. The observed sum of record events for temperature is near the upper confidence level, and the 'backward' analysis yields significantly fewer than expected record events (outside the confidence interval). A χ^2-test between the record-density and $p_n(1)$ suggested that the deviation from the null hypothesis is statistically significant at the 5 % confidence level.

Fig. 9a shows the results from the record event analysis applied to the absolute monthly minimum temperature. The analysis was identical to the one for absolute monthly maximum temperature, except that the absolute monthly minimum temperature was multiplied by –1 prior to the analysis. The χ^2-test indicates a good agreement with the null hypothesis in the forward analysis, suggesting that the new cold record events are not taking place more often than expected from the null hypothesis. The diamonds in Fig. 9a corresponding to the 'backward' analysis, on the other hand, are biased to higher empirical values for large n than the results from the 'forward' analysis. The results from the χ^2-test on the 'backward' analysis exceed the 95% confidence level, and the mean observed number of record events $\hat{\varepsilon}(n)$ is either on the verge of, or outside the 95% confidence interval for the 'forward' and 'backward' analyses (Table 1).

The record event analysis was applied to the maximum 24 h precipitation from each month (Fig. 9b, the locations of the precipitation stations are shown in Fig. 2 and the length of the series was 105), and the results in general indicate that the empirical data are consistent with the null hypothesis of stationary conditions. The exception is for the 'backward' analysis, which gives slightly too few record events.

4. DISCUSSION

Vogel et al. (2001) examined 1571 flood records in the US and concluded that the average number of record-breaking flood events over 10 to 80 yr intervals behaved as if the flood series were serially independent in the whole of the US when spatial correlation was taken into account. On the other hand, they found that if the spatial correlation was ignored, the flood records did not appear to be serially independent in parts of the US. Hence, their study points to the importance of accounting for the spatial correlation structure when performing regional hypothesis tests. The fact that the record events occurrence follows the theoretical values as closely as they do in Figs. 8 & 9 suggests that record density is not strongly influenced by the correlation between the stations used in this study. The results from the Monte-Carlo experiments with inter-station dependences in Table 1 also suggest that the tests are only weakly sensitive to the correlation among the stations. It appears that the χ^2-statistics and the related confidence limits are biased and unreliable for the extreme case where all station series are identical due to inflated and deflated estimates of the record densities (Fig. 6a). For less extreme cases, such as 5 of 25 stations being truly independent, the χ^2-test is less affected, but still near the confidence limits. The estimated number of records $\hat{\varepsilon}(n)$, on the other hand, appears to be more reliable for inter-station dependences, since the number of independent series affects the robustness, but does not bias the estimates.

Too few records in the 'forward' analysis suggest a negative trend, whereas a small $\hat{\varepsilon}(n)$ value in a 'backward' analysis (Table 1) can be associated with a positive trend (e.g. a global warming). According to the estimated number of records in Table 1, the 'backward' analysis points to values for $\hat{\varepsilon}(n)$ outside the 95% confidence region for maximum and minimum temperatures, whereas the results from the 'forward' analysis verge on being statistical significant at the 5% level. Hence the temperatures do not appear to be iid. For the 24 h precipitation, 'backward' analysis gives slightly too few records, whereas the 'forward'-based $\hat{\varepsilon}(n)$ is well within the confidence interval. This situation may well be due to chance, as described above, where the largest values in a series of iid random variables by coincidence is found among the first entries in a series, as the 'forward' and 'backward' analysis are not consistent with each other. Neither the 'forward' nor the 'backward' analyses produced χ^2-statistics for the precipitation that were significant at the 5% level. Hence, the tests for 24 h precipitation do not discriminate between the precipitation series and the hypothesis of iid random variables.

5. CONCLUSION

According to simple theoretical considerations, the likelihood of seeing new record events diminishes with the length of observational series. However, the probability increases with the number of independent contemporary observations, and it can be shown that it is not unlikely to see new record events in 100 yr long series when there are more than 100 different independent simultaneous observations.

The null hypothesis of stationary series is used as a reference frame for studying the incidence of record events. A comparison with the global mean temperature indicates a poor agreement; this is expected, due to the warming trend in the global temperature series. Rejection of the null hypothesis means rejection of the assumption of iid random variables, which in this case is not equivalent to accepting the notion of anthropogenic change. Comparison between absolute monthly maximum and minimum temperature in the Nordic countries and the null hypothesis is more ambiguous. The mean observed and expected number of record events is near the border of the null hypothesis 95% confidence interval, but the χ^2-test results and $\hat{\varepsilon}(n)$ suggest that the difference is statistically significant. The results for the maximum monthly 24 h precipitation suggests that the incidence of new record events is not extraordinary, but follows the expected behaviour of stationary series. Hence, we cannot say that the 24 h rainfall has become more severe in the Nordic countries in the period 1895–1999. These conclusions hinge on the assumption of adequate data quality as well as appropriate and sufficient data coverage for representing the true climatic trends in the Nordic countries.

Due to the central limit theorem, the results of this type of analysis become more robust with the number of truly *independent* series included. However, there is a trade-off between this robustness, and spatial or seasonal resolution. For instance, this study has included all seasons, and may therefore not give a good description of features such as an abnormal incidence of record events only during one particular season. Furthermore, in cases of very localised changes in record event statistics (e.g. on the west coast of Norway), such trends may not be detectable in an analysis based on a network of stations from a large region. Choosing a region that is too small, on the other hand, has implications for inter-station dependences or robustness. This problem is analogous to field significance testing (von Storch & Zwiers 1999, p. 121), which is suitable for detecting a signal in a high-dimensional environment as long as the signal is not localised in a small part of the area examined. The record event statistics may, for instance, have been affected by the presence of a common temperature trend in the Nordic region, with a warm period in the 1920 and 1930s, a cold period in the 1960s, and rapid warming since the 1970s (Benestad 2003). A cooling during 1940–1960 in many of the series may affect the observed number of record events.

Acknowledgements. This work was done under the Norwegian Regional Climate Development under Global Warming (RegClim) programme, and was supported by the Norwegian Research Council (Contract NRC-No. 120656/720) and the Norwegian Meteorological Institute. Part of the analysis was carried out using the R (Ellner 2001, Gentleman & Ihaka 2000) data processing and analysis language (http://www.R-project.org/). Hans von Storch and 3 anonymous reviewers made useful comments on the manuscript.

LITERATURE CITED

Ahsanullah M (1989) Introduction to record statistics. Prentice Hall, Hemel Hempstead

Ahsanullah M (1995) Record statistics. Nova Science Publishers, Commack, NY

Anderson TW (1958) An introduction to multivariate statistical analysis, 1st edn. John Wiley & Sons, New York

Arnold BC, Balakrishnan N, Nagaraja HN (1998) Records. John Wiley, New York

Bairamov IG, Eryilmaz SN (2000) Distributional properties of statistics based on minimal spacing and record exceedance statistics. J Stat Plan Inference 90:21–33

Balakrishnan N, Chan PS (1998) On the normal record values and associated inference. Stat Prob Lett 39:73–80

Benestad RE (2003) What can present climate models tell us about climate change? Clim Change 59:311–332

Coles S (1999) Extreme value theory and applications. Notes from a course on EVT and applications presented at the 44th Reunião Annual da RBRAS e 8th SEAGRO, at Bucato, São Paulo, Brazil, 26–30 July 1999. www.maths.bris.ac.uk/~masgc/extremes/exnewnotes2.ps

Ellner SP (2001) Review of R, Version 1.1.1. Bull Ecol Soc Am 82(April), 127–128

Feller W (1971) An introduction to probability theory & its applications, 2nd edn. John Wiley & Sons, New York

Frei C, Schär C (2001) Detection of trends in rare events: theory and application to heavy precipitation in the alpine region. J Clim 14:1568–1584

Gentleman R, Ihaka R (2000) Lexical scope and statistical computing. J Comp Graph Stat 9:491–508

Glick N (1978) Breaking records and breaking boards. Am Math Mo 85:12–26

IPCC (2002) Intergovernmental panel on climate change, workshop on changes in extreme weather and climate events. Workshop Report WMO/UNEP, Geneva; www.ipcc.ch/

Jones PD, Raper SCB, Bradley RS, Diaz HF, Kelly PM, Wigley TML (1998a) Northern Hemisphere surface air temperature variations, 1851–1984. J Clim Appl Met 25:161–179

Jones PD, Raper SCB, Bradley RS, Diaz HF, Kelly PM, Wigley TML (1998b) Southern Hemisphere surface air temperature variations, 1851–1984. J Clim Appl Met 25:1213–1230

Nagaraja HN (1988) Record values and related statistics—a review. Comun Statist Theo Meth 17:2223–2238

Nevzorov VB (1987) Records. Theo Prob Appl 32:201–228

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1989) Numerical recipes in Pascal. Cambridge University Press, Cambridge

Raqab MZ (2001) Some results on the moments of record values from linear exponential distribution. Math Comp Model 34:1–8

Sneyers R (1990) On statistical analysis of series of observations. Tech note 143. WMO, Geneva

Tuomenvirta H, Drebs A, Førland E, Tveito OE, Alexandersson H, Laursen EV, Jónsson T (2001) Nordklim data set 1.0. KLIMA 08/01. Oslo, Norway (www.met.no)

Vogel RM, Zafirakou-Koulouris A, Matalas NC (2001) The frequency of record breaking floods in the United States. Wat Resour Res 37:1723–1731

von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge

Wallace M (1996) Observed climatic variability: time dependence. In: Anderson DLT, Willebrand J (eds) Decadal variability. NATO ASI series, Vol 44. Springer, Berlin, p 1–11

Wilks DS (1995) Statistical methods in the atmospheric sciences. Academic Press, Orlando, FL

Editorial responsibility: Hans von Storch, Geesthacht, Germany

Submitted: December 4, 2002; Accepted: August 2, 2003
Proofs received from author(s): September 26, 2003