A Nonparametric Test for Trends in the Occurrence of Rare Events

JOSÉ A. LÓPEZ-DÍAZ

Climatological Techniques Division, Instituto Nacional de Meteorología, Madrid, Spain

ABSTRACT

A nonparametric test for trends in the occurrence of rare events, based on the average position that the events occupy in the series, is presented. This test is formally identical to the Wilcoxon–Mann–Whitney test for the difference of means between two samples. In the present application, however, the range of values of the length N of the series for which accurate critical values are available has to be expanded considerably. Exact formulas for the p value of the test on the average position for number of events $m = 2, 3$, and 4 are given, as well as a recursive relation for general m. Since this procedure cannot in practical times be carried out beyond a small m, a combinatorial technique that allows the p-value computation with great accuracy is explained. For m greater than around 20 it is shown that the convergence to the normal distribution is good. The power of this nonparametric test is shown to be superior to that of tests based on the interevent times. The test is applied to a series of annual minimum temperatures and to a series of seasonal precipitation totals, thereby illustrating the practical advantages of this approach.

1. Introduction

Interest in climate extremes has been growing in recent years, since changes in the magnitude and frequency of climatic extremes will have important environmental and socioeconomic consequences. Numerous task groups, including the Intergovernmental Panel on Climate Change, have identified the detection of trends and variability in extreme temperatures as critical factors toward an improved understanding of past and potential future global change. Climate simulations with coupled models have forecast relative changes in extreme precipitation that are larger than those in total precipitation (Zwiers and Kharin 2000). In the context of climate change detection, existing data indicate that the climate is becoming more extreme in some areas and for some variables, but at this time it is difficult to link such changes with anthropogenic effects (Karl and Easterling 1999). Emphasis has been laid recently on the importance of developing indicators and indices of climate extremes with a view to monitoring changes in the climate extremes (Nicholls and Murray 1999; Folland et al. 1999).

The final aim of the study of these series of indicators or other extreme-related parameters is the detection of a trend in the frequency or magnitude of extreme events in a time series, be it historical or coming from model simulations. Two broad approaches can be distinguished in this connection: either the inference about the extremes is based on modeling the entire time series, or it is circumscribed to the extreme events themselves (Solow 1999). The first of these is based on parametric tests and allows maximum probabilities in the detection of trends, but this is at the cost of assuming a number of distributional hypotheses about the underlying process as well as a particular form of the trend under the alternative hypothesis. These assumptions, if ever delicate, can become very difficult to justify for the tails of the distributions (which determine the extreme behavior) due to the limited availability of observations in this range. On the other hand, the extreme-centered approach using nonparametric tests, as the one considered in this paper, offers the undeniable advantage of its superior robustness, which may well outweigh the loss in power with respect to the other approach.

The test presented here allows testing the hypothesis of a trend in the frequency of events in a series where no assumption is made with respect to their distribution except that they come from a random sampling (i.e., independence and identical distribution). It has to be borne in mind, however, that the presence of serial correlation or of low-frequency climate variability signals (e.g., El Niño), to cite two quite common circumstances in climatological series, may call into question the validity of this null hypothesis. With a choice of a threshold the test can be applied to any series, with events defined as values below or above the threshold.

This paper is structured as follows. Section 2 presents the basic definitions of the test based on the sum of the

Corresponding author address: José Antonio López-Díaz, Head of Climatological Techniques Division, Instituto Nacional de Meteorología, Camino de las Moreras s/n, 28040 Madrid, Spain.
E-mail: jalopez@inm.es

event positions and its relation with other procedures. In section 3 procedures for the calculation of the *p* value of the test are explained, first exact formulas (for number of events $m = 2$, 3, and 4), then the combinatorial approximation and finally the normal approximation, which is valid for *m* greater than around 20. The question of the power of this test is dealt with in section 4, which offers results of Monte Carlo simulations for the comparison of the detection probabilities with other parametric and nonparametric tests. Section 5 shows by means of two examples of application (to a minimum temperature series and to series of seasonal precipitations) some of the possibilities of the test and a suitable graphical procedure for the presentation of its results. Finally section 6 contains the conclusions.

2. The test based on the sum of the positions

Suppose we have a series of independent observations $x_1, x_2, \ldots x_N$ from a binary random variable, with 0 denoting the absence of the event and 1 its occurrence. In these conditions the positions that the 1's take in the series are obviously randomly distributed. More precisely, the probability of an *m*-tuple of positions $(o_1, o_2, \ldots, o_m)$, $o_i \in \{1, 2, \ldots, N\}$, is the same for all the $N(N - 1) \cdots (N - m + 1)$ possible *m*-tuples. On the other hand, if the occurrence of the event has an increasing trend, one expects that the 1's should be pushed to the right in the series, and conversely for a decreasing trend.

In order to design an objective test a natural choice as statistic is the sum of the *m* positions $S = \Sigma \, o_i$ (or equivalently their average). Then we reject the null hypothesis of randomness in the positions in favor of a decreasing/increasing trend in their frequency if *S* is below/above a certain critical value. The *p* value α of this test in one-sided form can be determined if we are able to find the number $F(N, m, S)$ of *m*-tuples of positions that satisfy the condition $\Sigma \, o_i \leq S$, for then we get simply (obvious modifications apply for the case where we want to reject if *S* exceeds a certain value),

$$\alpha = F(N, m, S)/[N(N - 1) \cdots (N - m + 1)]. \quad (1)$$

This test is closely related to the Mann–Whitney test, also known as Wilcoxon test (see, e.g., Conover 1999). This is a nonparametric test for testing the difference in the means of two independent samples. Both samples are combined into a single-ordered sample and then ranks are assigned to the sample values from the smallest value to the largest, without regard to which population each value came from. Then the test statistic is the sum of the ranks assigned to those values from one of the populations. Consequently the test statistic and its distribution under the null hypothesis are the same as for the test on average positions considered here, with *N* equal to the sum of lengths of the two samples.

In the context of trend analysis in Poisson processes, which are the continuous-time analog of the binary var-

iable random sampling considered in this paper, the statistic $S^* = \Sigma_{i=1}^{m} T_i$ (with T_i being the time of occurrence of the *i*th event) has been proposed before (Cox and Lewis 1966).[1] Under the hypothesis of an exponential trend in the rate $\lambda(t)$ of the process, that is, $\lambda(t) = \exp(\alpha + \beta t)$, they show that the log-likelihood of observing *m* events at the times T_i, $i = 1, 2, \ldots, m$ during the period of time $[0, T]$ is

$$l(\alpha, \beta) = m\alpha + \beta S^* - e^{\alpha}(e^{\beta T} - 1)/\beta. \quad (2)$$

This expression shows that for fixed *m* the slope β depends on the observations only through the statistic S^*. Under the null hypothesis $\beta = 0$, S^* is distributed as the sum of *m* random variables uniformly distributed on $[0, T]$ and so they propose using this fact to test this null hypothesis against the alternative $\beta \neq 0$. The distribution of S^* under H_0 is given by Eq. (A2) of appendix A (with $U_i = T_i/T$). The authors give the normal approximation $S^* \sim N(mT/2, 1/12 \, mT^2)$ toward which the convergence is rapid. Dividing the interval $[0, T]$ in *N* equal parts, as *N* tends to infinity this test and the test on the average positions become identical.

Finally, the statistic of the test presented here, the sum of the positions, can be regarded formally as an extension, for series obtained from sampling of a binary random variable, of Mann–Kendall's test (Mann 1945; Kendall 1955; Sneyers 1990), which is considered one of the best nonparametric tests available for testing trends in series with no (or only a few) repeated elements (see section 4b below). It can also be obtained by applying Spearman's trend test, another nonparametric trend test widely applied in climatology, to the binary series considered here. This test is based on the Pearson's correlation coefficient between the series of ranks of the observations and the series of the natural numbers. In the binary series there are only two ranks, say 0 and 1, and the variances in the denominator of the correlation coefficient only depend on *N* and *m*, while the covariance in the numerator depends on the observations only through *S, N,* and *m*. So for fixed *m* (and *N*) we get an equivalent test to the one presented here.

3. Computation of the probabilities for the test on the sum of positions

Since, as explained in the preceding section, the distribution law of the sum of positions under the null hypothesis of random binary sampling (for fixed *m*) is the same as that of Mann–Whitney's statistic, the tables for the latter can be used for the test on sum of positions. zFor example, von Storch and Zwiers (1999) give detailed tables for $N \leq 64$ and $m \leq 14$ computed using Monte Carlo techniques. However, for the present application it is necessary to extend the range of *N,* the length of the series of events.

---

[1] I am indebted to one of the anonymous reviewers for having pointed out this fact to me.

TABLE 1. Values of the relative error in % of the combinatorial approximation of $F(N, m, S)$ with $N = 50$ and different orders of approximation up to the full-order approximation (see appendix A).

p value	Avg position $S/[m(N + 1)]$	Approximation order				
		m = 3				
		1	2	3		
0.394	0.456	0.5	−0.009	0.004		
0.193	0.358	2.6	−0.04	0.03		
0.0485	0.230	8.1	−0.2	0.1		
0.0104	0.144	18.6	−0.8	0.2		
0.001 17	0.78	51.2	−4.9	0.3		
		m = 5				
		1	2	3	4	5
0.40	0.466	1.0	−0.16	0.01	0.005	0.005
0.197	0.394	5.1	−0.9	0.07	0.03	0.03
0.05	0.298	15.9	−3.3	0.3	0.08	0.09
0.01	0.223	33.5	−9.0	0.8	0.17	0.18
0.001	0.153	76.1	−28.6	3.1	0.35	0.42
		m = 10				
		2	4	6	8	10
0.026	0.345	−480	−9.5	0.03	0.1	0.1
0.0094	0.313	−830	−19.3	−0.03	0.15	0.15
0.000 93	0.258	−2300	−79.5	−0.8	0.3	0.3

a. Exact computation of F(N, m, S)

The direct computation of $F(N, m, S)$ is rather straightforward, but the time consumed increases very rapidly with m. One can show that the function

$$F^*(N, m, S) = F(N, m, S)/m! \qquad (3a)$$

(i.e., F^* gives the number of ordered m-tuples instead of arbitrary m-tuples) satisfies the recursive relation:

$$F^*(N, m, S) = \sum_{i=1}^{[S/m-(m-1)/2]} F^*(N - i, m - 1, S - mi), \qquad (3b)$$

which allows in principle the computation starting with $m = 1$ and increasing the value of m successively. In order to be able to reach in practical times values of m greater than around six we need to find, however, explicit formulas for F^* for the cases $m = 2, 3,$ and 4. These are given in appendix B. Although we could proceed this way for greater m, a glance at the formulas shows that they would have a very complicated structure and a lot of coefficients would need to be determined. As m increases the time consumed blows off (if S is very small the increase is slower), so a new approach is needed that does not suffer from this drastic increase in computing time with m.

b. The combinatorial approximation to F(N, m, S)

Fortunately using combinatorial techniques we can expand the range of values of m for which a very good approximation to $F(N, m, S)$ can be calculated in little time. An outline of this method can be found in appendix A. To give an idea of the accuracy of this combinatorial approximation, Table 1 lists the relative deviations $100[F_n^0(N, m, S)/F(N, m, S) - 1]$, in percentage, of the combinatorial approximation $F_n^0(N, m, S)$ to $F(N, m, S)$ for different orders of approximation n to the true value and fixed $N = 50$. In all cases we can see that the full-order approximation with $n = m$, or near it, is excellent. As a general rule, for fixed N and m the combinatorial approximation converges more rapidly as S increases. Note that, for $m = 10$, only values for small p value were included since the time taken by the computation of the exact value of F for greater p values was prohibitive.

Table 2 contains the critical values of average position in thousandths, $1000 S/[m(N + 1)]$, for different values of N between 40 and 200, m between 3 and 20 and of the p value (one-sided form).

c. The normal approximation

For values of m greater than around 20 even the combinatorial approach requires a lot of computations. But then the distribution of S comes close enough to the normal distribution. The mean and variance of S are the same as for the Mann–Whitney test (see section 2) and have the expressions:

$$E[S] = \frac{1}{2}m(N + 1), \qquad (4a)$$

$$\text{var}[S] = \frac{1}{12}m(N - m)(N + 1). \qquad (4b)$$

For the average position $s = S/[m(N + 1)]$, whose critical values are given in Table 2, we get the simple expression $E[s] = 1/2$.

TABLE 2. Critical values of the average position (in thousandths) 1000 $S/[m (N + 1)]$ for different one-sided p values (shown in the first row after the number of events m).

N					$m = 3$				
	0.2	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
40	357	288	233	189	145	120	99	79	68
60	356	286	230	185	140	114	94	73	61
80	356	285	228	183	138	112	91	70	57
100	356	284	227	182	136	110	89	67	55
150	355	283	226	180	134	108	87	65	53
200	355	283	225	180	133	107	86	64	51

N					$m = 4$				
	0.2	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
40	380	320	274	235	193	166	145	121	107
60	378	317	270	230	187	160	138	113	99
80	377	316	268	227	184	157	134	109	94
100	377	315	266	226	182	155	132	107	92
150	376	314	265	224	180	152	129	104	89
200	376	313	264	223	178	151	128	103	87

N					$m = 5$				
	0.2	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
40	395	342	301	266	227	203	181	157	142
60	393	339	296	260	221	195	173	148	132
80	392	337	294	257	217	191	169	144	127
100	391	336	292	256	215	189	167	141	124
150	390	335	291	254	213	186	163	137	121
200	390	334	290	253	211	185	162	136	119

N					$m = 6$				
	0.2	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
40	406	359	321	289	254	231	210	186	171
60	403	355	316	283	246	222	201	177	161
80	402	353	313	280	243	218	197	172	155
100	402	352	312	278	240	216	194	169	152
150	401	350	310	276	237	213	191	165	148
200	400	349	309	274	236	211	189	163	146

N					$m = 8$				
	0.2	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
40	421	382	350	323	292	272	254	232	218
60	418	377	344	315	283	262	243	221	205
80	417	375	341	312	279	257	238	215	199
100	416	373	339	309	276	254	235	212	196
150	415	372	336	307	273	251	231	207	191
200	414	371	335	305	271	249	229	205	189

N					$m = 10$				
	0.2	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
40	432	398	370	346	320	302	285	266	253
60	429	392	363	338	309	290	273	253	239
80	427	390	360	334	304	285	267	246	232
100	426	388	358	331	302	282	264	242	227
150	425	386	355	328	298	277	259	237	222
200	424	385	354	327	296	275	257	235	220

N					$m = 15$				
	0.2	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
40	450	424	403	385	365	351	339	324	314
60	445	417	394	374	352	336	323	306	295
80	443	414	390	369	345	330	315	298	286
100	442	412	387	366	342	325	311	293	281
150	440	409	384	362	337	321	305	287	274
200	439	408	382	360	335	318	302	284	271

N					$m = 20$				
	0.2	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
40	461	441	425	411	395	384	375	363	355
60	455	432	413	397	378	366	355	341	331
80	453	428	408	391	371	358	345	331	321
100	451	426	405	387	366	353	341	325	315
150	449	423	401	383	361	347	333	318	307
200	448	421	399	380	358	344	331	315	303

For fixed m, the dependence of the relative errors of this normal approximation with N is very weak. With increasing m the accuracy of the normal approximation improves, for $m = 3$ we have relative errors for the critical values of the test not surpassing 10% for one-sided p values not below 2.5%, for $m = 8$ this limit is pushed to 1% and for $m = 15$ to 0.5%. By a happy coincidence the normal approximation gives critical values for the common 2.5% one-sided p value, that is, 5% for the two-sided test, with relative errors of only a few percent across all the N and m of Table 2.

4. Considerations on the power of the test

Nonparametric tests have the advantage of their wider range of applicability and the robustness of their conclusions, a fact derived from the simplicity of the hypotheses that have to be assumed in their application. However, a price to pay for this is that usually they are less powerful than similar parametric tests. This means that they can be less able to detect deviations from the null hypothesis, or stated in terms of the usual classification of error types in hypothesis testing, that the type II error (conserving the null hypothesis when it is the case that it is not satisfied) can be greater than for similar parametric tests. The ability of a test to detect that the null hypothesis is not satisfied is called the power of the test.

a. The test based on logistic regression

Frei and Schär (2001) use a method for the assessment of trends in the frequency of rare events in long-term records based on the stochastic concept of binomial counts. They employ logistic linear regression, a particular case of generalized linear models (see McCullagh and Nelder 1989), that postulates a linear relation between the logit of the event probability and time, that is, denoting the probability of the event at the ith position by p_i

$$\log\left(\frac{p_i}{1 - p_i}\right) = \alpha + \beta i. \tag{5}$$

This test, when applied to binary random variables, the case considered in the present paper, is closely related to the test on the average positions. This can be seen from the expression for the log-likelihood of N observations from variables Bi $(1, p_i)$

$$l(\alpha, \beta) = m\alpha + \beta S - \sum_{i=1}^{N} \log(1 + e^{\alpha+\beta i}). \tag{6}$$

The Frei–Schär method consists of testing the null hypothesis $\beta = 0$ against the alternative $\beta \neq 0$. Equation (6) is very similar in structure to Eq. (2), so that for fixed N and m again the dependence of β on the observations is only through S. The p value of this test uses the asymptotic result that under H_0 the difference between $1(\alpha, \beta)$ of (6) maximized over α and β and $1(\alpha, 0)$ maximized over α is distributed as a $\chi^2(1)$ random variable in the limit as N and m tend to infinity.

In order to compare the power of detection of both tests of a linear trend, a Monte Carlo simulation with 2000 samples, each with $N = 100$, was carried out. The central value of p was set to 0.1, with $\Delta p = 0.1$ from the beginning to the end of each simulated sample. The linear trend was detected in 16.5% of the cases using the logistic regression, while the test on the average positions detected 14.1% of the cases, both at the 5% significance level. However this difference of power appears to be due to the fact that the p values provided by the logistic regression are underestimated, that is, the test tends to overreject the null hypothesis when it is true. In fact the p value of this test corresponding to a rejection region of probability 5%, found by another simulation with $p = 0.1$ and no trend (2000 random series also), was 4.1%. If this p value is used as a criterion for rejection at the nominal 5% level in the preceding simulation with trend, then the trend is detected in 14.5% of the cases, practically identical to the test on the average position.

The problem comes from the fact the null hypothesis of the logistic regression is composite, containing a nuisance parameter α. The distribution of the test statistic should ideally be independent of this parameter, but in this case it is so only in an asymptotic sense (i.e., the test is not similar in the language of statistical theory). A standard way out of this difficulty (see, e.g., Cox and Hinkley 1990) is to build a statistic by analyzing the distribution of the data after conditioning on a sufficient statistic for the nuisance parameter under H_0. From (6) we can see that m is such a sufficient statistic, and so the test on average positions, being based on the distribution of S for fixed m, does not suffer from this problem. Furthermore, since the p value of the logistic regression depends only on m and S, it can be shown that if we insist on demanding that the distribution under H_0 of the p value be independent of the nuisance parameter, then the p value of the logistic regression cannot but be identical to that of the test on average positions.

b. Tests based on interevent times

Another way to construct tests for trends in the occurrence of events is through consideration of the interevent times (or interarrival times in the language of counting processes). But these tests have some drawbacks. First, there is a loss of information in the process of differentiation that the interevent times imply, which can seriously weaken them. This is most easily seen through an extreme example: suppose that we knew that in a series of length N the m events occupied just the last m positions. A nonparametric test applied to the interevent times would be unable to distinguish if the

series had $N = 2m$ or $N = 10m,$ say, since in both cases it would be applied to a series of interevent times of the form $(t, 1, 1, \ldots 1)$ with m 1's and $t > 1$, and no matter how great t is the p value would be exactly the same. Second the presence of ties in the interevent times weakens this type of test.

Keim and Cruise (1998) consider, for continuous-time events, the Poisson process (as in the Cox and Lewis procedure, see section 2 above) and propose testing the significance of the linear regression between the logs of the interarrival times of the Poisson-assumed process and the times of the events since the beginning.

Another possibility is to use the nonparametric Mann–Kendall's trend test applied to the interevent times. Since ties in the interevent times are bound to occur, the version of Mann–Kendall's test that takes into account the reduction in variance that ties cause is preferable (see, e.g., Hirsch et al. 1993). This uses as statistic, for a series of observations $x_1, x_2, \ldots, x_N,$

$$ S = \sum_{k=1}^{N-1} \sum_{j=k+1}^{N} \operatorname{sgn}(x_j - x_k), \qquad (7a) $$

with $\operatorname{sgn}(x)$ the sign function (0 for ties). Under the null hypothesis of identical distribution of the x_i, S has null mathematical expectation and variance given by

$$ \operatorname{var}[S] = \frac{1}{18}\Bigg[N(N-1)(2N+5) $$

$$ - \sum_t t(t-1)(2t+5)\Bigg], \qquad (7b) $$

where t is the length of any tie and the summation is extended over all ties. In computing the standardized variable used for testing we use, instead of S, $S - 1$ if $S > 0$, and $S + 1$ if $S < 0$.

Note that for binary data a test based on the statistic (7a) is equivalent to the test on the average positions. To see this consider the ith 1 in the series, in position o_i, having then $o_i - i$ 0's to the left, and $N - m - (o_i - i)$ to the right, giving a net contribution to the statistic of $2(o_i - i) - N + m$. Summing over all i the Mann–Kendall statistic depends on the random quantities only through $\sum o_i$, the same as the test on the average positions. But note that Mann–Kendall's test applied to the interevent times is certainly not equivalent to the test on average positions.

Table 3 compares the test on the average positions, Mann–Kendall's (on interevent times) and the log-regression test (with no grouping of interevent times since this was found to be optimal in these 2000 simulations) The table shows a decrease in the detecting power of both the test on the average position and the tests on interevent times with increasing p for fixed N and Δp. We may understand qualitatively the reason by first putting S in the form

TABLE 3. Detection probabilities at the 5% two-sided level of linear trend ($\Delta p = 0.2$) in the occurrence probability p for the test on the average positions, Mann's test on the interevent times, and log-regression for $N = 200$. The last row shows the ratio of the first two rows.

p (center)	0.2	0.3	0.4
Avg pos. (1)	53.7	40.5	39.0
Mann (2)	36.7	27.8	28.2
Log-reg	34.5	26.5	28.0
(1)/(2)	1.47	1.46	1.38

$$ S = \sum_{i=1}^{N} iI_{(x_i=1)}, \qquad (8) $$

where the indicator function of an event A, I_A, is a random variable that is 1 if A is true, 0 otherwise. From this we deduce, for the expected values of S for a given Δp:

$$ p_i = \left(p - \frac{\Delta p}{2}\right) + \frac{\Delta p}{N-1}(i-1) $$

$$ E[S_{\Delta p=0}] = p \sum i $$

$$ E[S_{\Delta p}] = \sum i p_i - E[S_{\Delta p=0}] + \Delta p h(N), \qquad (9) $$

where the function h depends only on N. We may reason then that the change in the mathematical expectation of S for the most probable values of m will also be proportional to Δp. On the other hand, for the variance of S we expect from (4b), putting $m \cong Np$, a dependence on p of the form $p(1 - p)$, so that for the standardized deviation of S we get a dependence (apart from functions of N) of the form $\Delta p/[p(1 - p)]^{1/2}$, which decreases with p for $p \leq 1/2$.

Another conclusion from Table 3 is that the test on the average positions performs consistently better than the other two, the relative increase in detection power being around 40% or greater (for smaller N the relative advantage is even greater). The log-regression and Mann's tests perform similarly, the latter slightly better. A possible explanation is that the log-regression is designed for continuous-time events and so its distributional assumptions are only approximately satisfied here.

Figure 1 depicts the detection probabilities of the test on the average positions at the 5% two-sided level for different values of a linear trend in the occurrence probability. For each simulated series $p = 0$ at the beginning. We notice an increase in the detection probability with the length N of the series, for fixed Δp. This can be explained by noting that the function $h(N)$ in (9) is of order N^2 while the variance of S is of order N^3, resulting in a dependence on N for the standardized deviation of S of order $N^{1/2}$ (for Mann–Kendall's test on interevent times one can show, reasoning along similar lines as before, that the standardized statistic also grows as $N^{1/2}$).
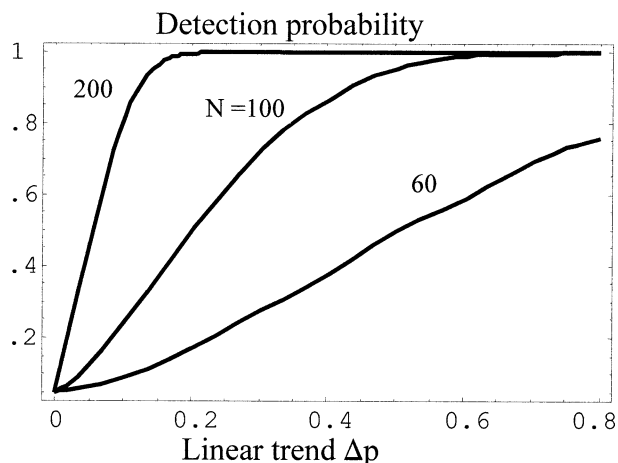
Detection probability

FIG. 1. Detection probabilities of the test on the average positions at the 5% two-sided level for linear trends in the occurrence probability ($p = 0$ at the beginning of the series).

5. Examples of application

a. The series of minimum annual temperature of Villanubla

The Villanubla Observatory is situated near the airport of the city of Valladolid in the northern half of the Iberian Plateau. Since the station is a few kilometers away from the urban center the series should in principle not show an appreciable urban effect.

Figure 2 is a plot of the annual absolute minimum temperature series at Villanubla. Table 4 contains the p values of Mann's tendency test (in two-sided form, with sign indicating sign of trend, note that it does not necessarily coincide with the sign of the slope) and of the linear regression with respect to the year (with plus sign for increasing tendency and vice versa, as in what follows). These are applied to the series of absolute annual minimum temperature of Villanubla (second column), to the series of number of days each year with minimum temperature below the first decile of the series corresponding to that calendar day and number of days above the last decile (third and fourth columns) (Nicholls and Murray 1999). The last row gives the slopes of the fitted lines (in degrees or days per decade, respectively). The annual minimum temperature shows an increasing trend significant at the 5% level (and nearly 1% for Mann's test). However this trend does not exist in the second column of number of days below the first decile (nor in the last column).

The results of applying the test on the average position to extremal sets are listed in Table 5, for sizes m of the extremal sets ranging from 2 to 20. For each m an event is said to occur if the value is among the m greatest/smallest values in the series. Note that since the test on the average positions is not a simultaneous test over all m, applying it to different values of m on the same series increases the likelihood
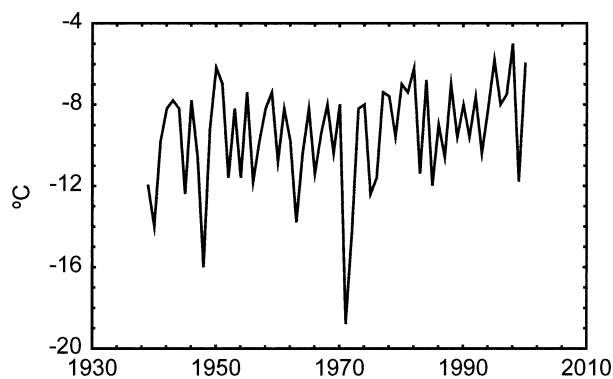

FIG. 2. Plot of the absolute minimum annual temperatures (in °C) at Villanubla (1939–2000).

of getting a significant result by pure chance. With the present test, however, the tests for different m's are far from independent. Successive values of the average position are strongly correlated, with the correlation coefficient increasing with m up to its maximum for m next to $N/2$ {it can be shown that, under H_0, $r(S_m, S_{m+1}) = mm'/[(m + 1)(m' + 1)]^{1/2}$, $m' = N - m - 1$}.

The first horizontal half of the table corresponds to the smallest values, the second to the greatest ones. For each of these the test on the average position using the exact and combinatorial formulas has been applied, in two-sided form.

Looking at the average position p values we see that the lowest minimum temperatures do not show significant values at the 5% level for any value of m, but from $m = 6$ onward hover around the 10% level. On the other hand the highest minimum temperatures do have average positions shifted to the more recent years generally significantly at the 5% level, and the three highest values much more so ($\alpha = 0.0006$). For values of m greater than 10, where its application is warranted, Mann–Kendall's test applied to the times between events does not yield significant values (not represented).

Instead of finding the p value of the average position test for different values of m as before, we can get rapidly an idea of the results of the test by plotting for each m the value of the average position. More

TABLE 4. Two-sided p values of trend tests (sign according to trend) and slope for the series of minimum annual temperatures of Villanubla (1939–2000), the series of number of days with T_{min} below the first decile and number of days with T_{min} above the ninth decile.

	Annual T_{min}	No. days $\leq Q_{0.1}$	No. days $\geq Q_{0.9}$
p-value Mann	0.016	0.94	−0.36
p-value regr.	0.037	0.90	−0.41
Slope	0.38°C decade^{-1}	0.13 days decade^{-1}	−0.69 days decade^{-1}

TABLE 5. Two-sided p values of the test on the average of positions applied to the minimum annual temperatures of Villanubla (1939–2000). The event sets are defined as those containing the smallest or greatest values, respectively, with increasing size m. Bold entries are significant at the 5% level, bold italic at the 1% level.

	Smallest values										
m	2	3	4	5	6	8	10	12	14	16	20
p value	0.47	0.60	0.19	0.18	0.14	0.10	0.13	0.16	0.10	0.16	0.15
	Greatest values										
m	2	3	4	5	6	8	10	12	14	16	20
p value	**0.017**	***0.0006***	**0.017**	**0.044**	**0.027**	**0.037**	0.053	0.082	**0.037**	**0.012**	0.060

precisely, we plot the minimum of both $S/[m(N + 1)]$ (as in Table 2 of critical values) and $1 - S/[m(N + 1)]$ and indicate by a symbol, in this case a triangle, the sign of the trend (for greatest values, the first possibility corresponds to decreasing trend, the second to increasing trend, and vice versa for smallest values). The plot corresponding to Villanubla is in Fig. 3. An upward-pointing triangle denotes a shift in positions in accordance with an increasing tendency (downward pointing for a decreasing tendency). The full line corresponds to greatest values, and the broken line to smallest ones. Besides, the critical values for the 5%, 1%, and 0.1% levels (two-sided) are depicted as broken smooth lines (without triangles), the lesser the p value the nearer to the x axis the line.

Finally, a comment about the relative sizes of m and N. For m equal to its maximum value ($N/2$ supposing N even) we are testing the average position of the 50% elements above the median of the series, and so also of the 50% below it, so this test could in theory be used as a global trend test. But, since we are throwing away all the information that the intraclass variability (i.e., relative ordering of the positions of the $N/2$ values above or below the median) of the values

in both classes has, this would be a poor test for such a purpose.

b. Seasonal precipitation in San Fernando

San Fernando has the longest instrumental record of precipitation in Spain. Over the analyzed period 1839–2000, the series of annual (September–August) totals can be considered homogeneous (Almarza et al. 1996) and does not show a significant trend (Mann's p value 0.84). Nor does the average position of extremal elements show any trend either. A more interesting picture emerges from the analysis of each of the yearly seasons autumn September–November (Mann's two-sided test p value -0.34, sign according to tendency), winter (0.071), spring (-0.23), and summer (-0.31), none of them below the 5% threshold, although winter is close to it. Figure 4 contains the plots of the average positions following the model of Fig. 3.

The plot corresponding to autumn shows that the more dry autumns tend to appear more frequently in recent years, for values of m greater than 10 they are closely aligned to the 5% limit. None of this happens with the more humid falls. In winter, by contrast, it is the more rainy ones that give significant results, remaining practically all up to $m = 20$ in the significant zone, and for $m = 3$, 4, and 5 even at the 1% level. The modest tendency that Mann's p value suggests can thus be attributed to these anomalously humid winters that tend to be more frequent in recent years. In the spring we notice that for m near 20 the largest values get pretty close to the 5% confidence limit, without reaching it. In summer it is worthwhile noting that the horizontal line representing the average positions of the smallest values at the beginning is due to the years with no summer precipitation at all, since in order to avoid unjustified arbitrariness each instance of a repeated value in the series has to be assigned a position equal to the average of all positions that the repeated value takes.
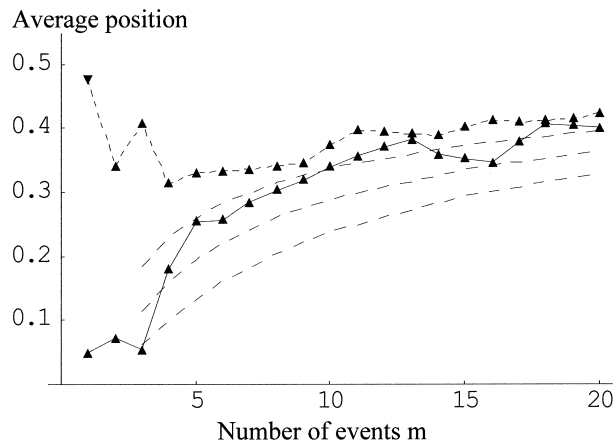

Average position

FIG. 3. Plot of the average positions $S/[m(N + 1)]$ of event sets of greatest (full line) and smallest (broken line) values of size m for the series of Fig. 2. Upward-pointing triangle for increasing trend, downward for decreasing trend. The smooth broken lines are the critical values at the (two-sided) 5%, 1%, and 0.1% levels, the lower the level the nearest to the x axis.

6. Conclusions

In this study a nonparametric test for trends in the occurrence of rare events based on their average position has been presented. This choice of statistic seems quite
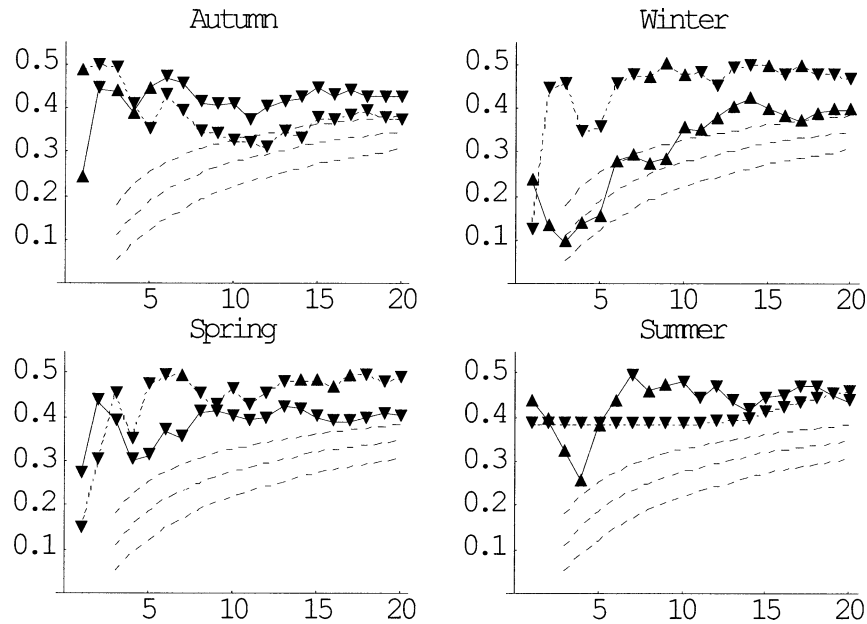
FIG. 4. Same plot as in Fig. 3 for the seasonal precipitation totals at San Fernando.

natural on different grounds, and is closely connected to Mann–Kendall's widely used trend test. The test statistic and its distribution under the null hypothesis are identical to those of a Mann–Whitney test on the difference of means between two samples. Exact formulas for the p-value probabilities have been given for a number of events $m = 2, 3$, and 4. The difficult problem of calculating the exact probabilities for greater m has been circumvented by a combinatorial technique that provides an excellent approximation to the exact values. This has allowed us to provide accurate critical values of the test for $N \leq 200$, $m \leq 20$. For m greater than around 20 it has been shown that the normal approximation can be safely used.

The detection probability of the test for linear trends in the event probabilities has been compared, using Monte Carlo simulations, to that of two parametric tests presently being used for similar purposes, and also to that of the test of Mann–Kendall test applied to interevent times. It has been shown that the test on the average positions is essentially equivalent to the test based on logistic regression when applied to binary random series, the case considered in this paper. When compared with the tests based on interevent times the performance of the test on the average positions is clearly better, although in the context of the general low detection probabilities that not too long series (e.g., N less than 200) of rare events permit.

As practical examples, the test on the average positions has been applied to the series of annual minimum temperatures of Villanubla and to the seasonal rainfall totals in San Fernando. For successively greater m the

events have been defined as those values occupying the extreme positions in the ordered series.

It has been shown that the test applied in this form complements the information supplied by the comprehensive trend tests like Mann–Kendall's by focusing on the behavior of these extremal sets. It permits in particular the differentiation between the greatest and the smallest elements. In the series of San Fernando, for example, we have seen that in one season (winter) it was the more rainy years that showed a trend while in another (fall) the drier ones. In other cases, the test will permit the detection of significant displacements in the position of the greatest or smallest elements that are not strong enough to make the whole series have a significant trend. Thus the method proposed widens our arsenal of techniques in the study of variability and trends of climatological series.

Besides its robustness, an advantage of the test presented in this paper is that its results can be stated in a concise and understandable manner with all precision. The concerned person or the decision-maker should have no problem in fully understanding what is meant by saying that in a series X of given length, under random conditions, one would expect to find the m events considered (be it occurrence of extreme phenomena, or values above or below certain threshold, or greatest or smallest values, etc.) so shifted to the recent years on average in only an $x\%$ of the cases. That is all that the test provides, and a similar exercise in conveying, with the same precision, the meaning of the p value of other tests (specially parametric ones) to the nonspecialist person can prove much harder.

APPENDIX A

The Combinatorial Approach to the Computation of $F(N, m, S)$ for the Test on the Average Positions

The problem consists in finding the number $F(N, m, S)$ of elements of the Cartesian product $[N^m] = \{1, 2, \ldots, N\}^m = \{(X_1, X_2, \ldots, X_m): X_i \in \{1, 2, \ldots, N\}\}$ satisfying $X_i \neq X_j \; \forall i, j$ and the condition $\mathcal{S}$:

$$\sum_{i=1}^{m} X_i \leq S.$$

a. First-order approximation

In the first approximation we compute all the elements of $[N^m]$ that satisfy the condition $\mathcal{S}$, that is, we let elements have repeated coordinates. This we can do by considering, in the Cartesian coordinates $(X_1, X_2, \ldots, X_m)$ each point belonging to $[N^m]$ in the center of a hypercube of unit length in all dimensions, so with unit volume. To estimate the total volume occupied by these hypercubes it is convenient to make a change of coordinates by moving the origin to $(1/2, 1/2, \ldots, 1/2)$. Let us denote the new coordinates by x_i. Then the hyperplanes $x_i = 0$ will just form part of the faces of the innermost hypercubes and will leave no extra room.

We then have that the approximate volume to consider, Ω_1, is that delimited by the $m + 1$ hyperplanes $x_i = 0$, $\Sigma_1^m x_i = S - m/2 + 1/2$, where the term $m/2$ accounts for the shift in coordinates. The $1/2$ comes from the fact that the hyperplane $\Sigma_1^m x_i = S - m/2$ passes exactly through the last layer of points satisfying $\mathcal{S}$, thus cutting the hypercubes of this last layer right through the middle.

On the other hand the volume Ω_1 satisfies

$$\Omega_1 = N^m \text{ prob}\left[\sum_1^m U_i \leq (S - m/2 + 1/2)/N\right],$$

$$U_i \text{ i.i.d. } U[0, 1] \tag{A1}$$

as can be seen from the fact that the joint density of independent identically distributed (i.i.d.) random variables in $[0, 1]$ is unity on the unit hypercube $[0, 1]x[0, 1]x \cdots x[0, 1]$.

The probability in (A1) is given by the formula (see Feller 1971):

$$\text{prob}\left[\sum_1^m U_i \leq s\right] = \frac{1}{m!} \sum_{j=0}^{[s]} (-1)^j \binom{m}{j}(s - j)^m$$

$$\forall s > 0. \tag{A2}$$

Inserting (A2) in (A1) yields the first-order approximation to $F(N, m, S)$, say $F_1(N, m, S)$. But we have

included points with repeated coordinates, and these have to be taken out.

b. The n*th order approximation*

The nth order ($n \leq m$) approximation includes terms that count points with repeated coordinates such that their number contains an N^{m-n+1} instead of N^m as in (A1). Thus the second-order approximation includes terms with (at least) a pair of coordinates identical, and the third-order approximation those with three identical coordinates, and those with two pairs of coordinates with identical values. In order to specify them it is useful to introduce the concept of a sequence of positive integers (a_i) satisfying $a_i \geq a_{i+1}$ and $a_i \geq 2$, and refer to them hereafter simply as monotone sequences. We call I_n the set formed by all the monotone sequences (a_i) with $\Sigma (a_i - 1) = n - 1$. For example, $I_2 = \{(2)\}$, $I_3 = \{(3), (2, 2)\}$, and $I_4 = \{(4), (3, 2), (2, 2, 2)\}$. For completeness set $I_1 = \{( )\}$. Then the nth-order approximation will contain terms related to I_i, $i = 1, 2, \ldots, n$.

For later use let us associate with a given monotone sequence (a_i) two sequences: 1) the sequence $(\mathcal{R}a_i)$ of the number of repetitions, that is, $\mathcal{R}a_1$ is the number of times that the number a_1 appears repeated in (a_i) and so on; for example, the sequence $(a_i) = (4, 4, 3, 2, 2, 2)$ gives $(\mathcal{R}a_i) = (2, 1, 3)$; and 2) the sequence $(\mathcal{U}a_i)$, which is (a_i) stripped of the repeated elements, for example, $(\mathcal{U}a_i) = (4, 3, 2)$.

Then $E[(a_i)]$, the number of points of $[N^m]$ satisfying $\mathcal{S}$ with a group of a_1 coordinates equal, another disjoint group of a_2 coordinates equal, etc., is given by

$$E[(a_i)] = N^{m-n+1}Z[m, (a_i)]P[(a_i), (S - m/2 + 1/2)/N],$$

$$(a_i) \in I_n. \tag{A3}$$

Here $Z[m, (a, b, \ldots)]$ denotes the number of sets that can be formed by disjoint subsets of a set of m elements with respective sizes $a, b, \ldots$. Set $Z[m, ( )] = 1$. It can be shown that

$$Z[m, (a_i)] = \frac{1}{\prod (\mathcal{R}a_i!)} \frac{1}{\prod (a_i!)} \frac{m!}{\left(m - \sum a_i\right)!},$$

$$\sum a_i \leq m$$

$$Z[m, (a_i)] = 0, \qquad \sum a_i > m. \tag{A4}$$

The probability term $P[(a_i), (S - m/2 + 1/2)/N]$ in (A3) is defined by

$$P[(a_i), x] = \text{prob}\left[\sum a_iU_i + \sum_{j=1}^{m-\Sigma a_i} V_j \leq x\right],$$

where U_i and V_j are all independent uniform variables $U[0, 1]$. This is a generalization of (A2) and can be obtained through an application of the inclusion–exclusion principle [see, e.g., Biggs (1993) or, for a detailed treatment, Stanley (1997)]. It can be shown that

$$P[(a_i), x] = \frac{1}{m!\left(\prod a_i\right)} \sum_{z_1=0}^{\mathcal{R}^*a_1} \cdots \sum_{z_i=0}^{\mathcal{R}^*a_i} \cdots (-1)^{\Sigma z_i} \left[\prod_k \binom{\mathcal{R}^*a_k}{z_k}\right] \left[x - \sum (\mathcal{U}^*a_i)z_i\right]_+^m, \qquad m \geq \sum a_i,$$

$$(\mathcal{R}^*a_k) = \left(\mathcal{R}a_k, m - \sum a_i\right), \qquad (\mathcal{U}^*a_i) = (\mathcal{U}a_i, 1), \tag{A5}$$

where the notation $x_+ = \max[x, 0]$ was used [for (a_i) = ( ) the convention $\prod a_i = 1$ applies]. The sequence $\mathcal{R}^*$ is constructed by concatenating to $\mathcal{R}$ the number $m - \sum a_i$ and similarly $\mathcal{U}^*$.

In the general expansion each of the terms of the form (A3) has to be multiplied by an integer coefficient $Q[(a_i)]$. For example, $E[(2)]$ has to be multiplied by -1

since points with a pair of repeated coordinates are counted once in (A1), and $E[(3)]$ by 2 since points with three repeated coordinates are counted in once in (A1) but taken out $C_2^3 = 3$ times in the term $(-1)E[(2)]$. For $Q[(a_i)]$ one can show that

$$Q[(a_i)] = (-1)^{\Sigma(a_i-1)} \prod (a_i - 1)! \tag{A6}$$

Finally the approximation of order n is given by

$$F_n(N, m, S) = \sum_{(a_i)\in I_k, k=1,...,n} (-1)^{\Sigma(a_i-1)} \frac{1}{\prod (\mathcal{R}a_i!)} \frac{1}{\prod a_i} \frac{m!}{\left(m - \sum a_i\right)} P[(a_i), S^*] N^{m-\Sigma(a_i-1)},$$

$$S^* = (S - m/2 + 1/2)/N \tag{A7}$$

[for the term with (a_i) = ( ) the conventions $\Sigma(a_i - 1) = 0$, $\prod(\mathcal{R}a_i!) = \prod a_i = 1$ apply]. For the full-order approximation $n = m$.

c. Final expression

An improvement on the above approximation of nth order can be obtained by noting that for S sufficiently great we must have

$$F(N, m, S) = F(N, m, \infty) = \prod_{i=0}^{m-1} (N - i).$$

The final expression for the combinatorial approximation of nth order

$$F_n^0(N, m, S) = \frac{F_n(N, m, S)}{F_n(N, m, \infty)} F(N, m, \infty) \tag{A8}$$

satisfies both $F_n^0(N, m, \infty) = F(N, m, \infty)$ and $F_m^0(N, m, S) = F_m(N, m, S)$. The latter from $F_m(N, m, \infty) = F(N, m, \infty)$ since the full-order approximation becomes exact when the probability terms are exact, which is the case when $S = \infty$ and all of them take the value 1.

APPENDIX B

Exact Formulas for the Calculation of $F^*(N, m, S)$ for the Test on Average Positions (Section 3a)

• Formula for $F^*(N, 2, S)$:

$$F^*(N, 2, S) = 0, \qquad S < 3$$

$$F^*(N, 2, S) = \begin{cases} S^2/4 - S/2, & S \text{ even} \\ S^2/4 - S/2 + 1/4, & S \text{ odd} \end{cases} \quad \text{if } 3 \leq S \leq N + 1$$

$$F^*(N, 2, S) = \binom{N}{2} - F^*(N, 2, 2N + 1 - S) \quad \text{if } N + 2 \leq S \leq 2N - 1$$

$$F^*(N, 2, S) = \binom{N}{2} \qquad\qquad\qquad\qquad \text{if } S \geq 2N - 1$$

• Formula for $F^*(N, 3, S)$:

$$F^*(N, 3, S) = 0, \qquad S < 6$$

$$F^*(N, 3, S) = A[S \bmod 6] + \frac{5}{12}S - \frac{5}{24}S^2 + \frac{1}{36}S^3 \quad \text{if } 6 \leq S \leq N + 3,$$

where

$$A[k] = 0, -\frac{17}{72}, -\frac{2}{9}, -\frac{1}{8}, -\frac{1}{9}, -\frac{25}{72} \qquad \text{for } k = 0, 1, \ldots, 5;$$

$$F^*(N, 3, S) = F^*(2N, 3, S) + B[T \bmod 2] - \frac{5}{12}T - \frac{3}{8}T^2 - \frac{1}{12}T^3, \quad \text{if } N + 4 \leq S \leq 3N/2 + 2,$$

where

$$T = S - N - 3, \qquad B[k] = 0, -\frac{1}{8} \qquad \text{for } k = 0, 1$$

$$F^*(N, 3, S) = \binom{N}{3} - F^*[N, 3, 3(N + 1) - S - 1] \quad \text{if } 3N/2 + 2 < S < 3(N - 1)$$

$$F^*(N, 3, S) = \binom{N}{3} \qquad \text{if } 3(N - 1) \leq S.$$

- Formula for $F^*(N, 4, S)$:

$$F^*(N, 4, S) = 0, \qquad S < 10$$

$$F^*(N, 4, S) = A[S \bmod 12] + B[S \bmod 2]S + \frac{53}{288}S^2 - \frac{1}{32}S^3 + \frac{1}{576}S^4, \quad \text{if } 10 \leq S \leq N + 6,$$

with

$$A[k] = 0, \frac{145}{576}, \frac{17}{72}, \frac{17}{64}, \frac{1}{9}, \frac{145}{576}, \frac{1}{8}, \frac{217}{576}, \frac{1}{9}, \frac{9}{64}, \frac{17}{72}, \frac{217}{576} \qquad \text{for } k = 0, 1, \ldots, 11 \quad \text{and}$$

$$B[k] = -\frac{3}{8}, -\frac{13}{32} \qquad \text{for } k = 0, 1;$$

$$F^*(N, 4, S) = F^*(2N, 4, S) - C[T \bmod 6] - \frac{5}{12}T - \frac{23}{72}T^2 - \frac{1}{12}T^3 - \frac{1}{144}T^4, \quad \text{if } N + 7 \leq S \leq 2(N + 1),$$

with

$$T = S - N - 6, \qquad C[k] = 0, \frac{25}{144}, \frac{1}{9}, \frac{1}{16}, \frac{1}{9}, \frac{25}{144} \quad \text{for } k = 0, 1, \ldots, 5;$$

$$F^*(N, 4, S) = \binom{N}{4} - F^*[N, 4, 4(N + 1) - S - 1] \qquad \text{if } 2(N + 1) < S < 4N - 6$$

$$F^*(N, 4, S) = \binom{N}{4} \qquad \text{if } S \geq 4N - 6.$$

Note that "$a \bmod b$" is the remainder on division of a by b.

REFERENCES

Almarza, C., J. A. López, and C. Flores, 1996: Homogeneidad y variabilidad de los registros históricos de precipitación de España (Homogeneity and variability of historical precipitation records in Spain). Pub. A-143, Instituto Nacional de Metorología, Spain, 318 pp.

Biggs, N. L., 1993: *Discrete Mathematics.* Oxford Science Publications, 480 pp.

Conover, W. J., 1999: *Practical Nonparametric Statistics.* 3d. ed. Series in Probability and Statistics, Wiley, 584 pp.

Cox, D. R., and P. A. W. Lewis, 1966: *The Statistical Analysis of Series of Events.* John Wiley and Sons, 285 pp.

——, and D. V. Hinkley, 1990: *Theoretical Statistics.* Chapman and Hall, 511 pp.

Feller, W., 1971: *An Introduction to Probability Theory and Its Applications.* Vol. 2. Wiley, 669 pp.

Folland, C. K., and Coauthors, 1999: Workshop on indices and indicators for climate extremes: Asheville, NC, USA, 3–6 June 1997 breakout group C: Temperature indices for climate extremes. *Climatic Change, 42,* 31–43.

Frei, C., and C. Schär, 2001: Detection probability of trends in rare events: Theory and application to heavy precipitation in the Alpine region. *J. Climate, 14,* 1568–1584.

Hirsch, R. M., D. R. Helsel, T. A. Cohn, and E. J. Gilroy, 1993: Statistical analysis of hydrological data. *Handbook of Hydrology,* D. R. Maidment, Ed., McGraw-Hill, Inc., 17.1–17.55.

Karl, T. R., and D. R. Easterling, 1999: Climate extremes: Selected review and future research directions. *Climatic Change, 42,* 309–325.

Keim, B. D., and J. F. Cruise, 1998: A technique to measure trends in the frequency of discrete random events. *J. Climate, 11,* 848–855.

Kendall, M. G., 1955: *Rank Correlation Methods.* 2d ed. Charles Griffin, 196 pp.

Mann, H. B., 1945: Nonparametric tests against trend. *Econometrica, 13,* 245–259.

McCullagh, P., and J. A. Nelder, 1989: *Generalized Linear Models.* 2d ed. *Monographs on Statistics and Applied Probability,* No. 37, Chapman and Hall, 511 pp.

Nicholls, N., and W. Murray, 1999: Workshop on indices and indicators for climate extremes: Asheville, NC, USA, 3–6 June 1997 breakout group B: Precipitation. *Climatic Change, 42,* 23–29.

Sneyers, R., 1990: On the statistical analysis of series of observations. WMO Tech. Note 143, WMO, 192 pp.

Solow, A. R., 1999: On testing for change in extreme events. *Climatic Change, 42,* 341–349.

Stanley, R. P., 1997: *Enumerative Combinatorics.* Vol. I. Cambridge University Press, 325 pp.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.

Zwiers, F. W., and V. V. Kharin, 2000: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *J. Climate, 13,* 3760–3788.