

## Monte Carlo Experiments on the Detection of Trends in Extreme Values

XUEBIN ZHANG

*Climate Research Branch, Meteorological Service of Canada, Downsview, Ontario, Canada*

FRANCIS W. ZWIERS

*Canadian Centre for Climate Modelling and Analysis, Meteorological Service of Canada, Victoria, British Columbia, Canada*

GUILONG LI

*Climate Research Branch, Meteorological Service of Canada, Downsview, Ontario, Canada*

(Manuscript received 28 March 2003, in final form 17 October 2003)

### ABSTRACT

Using Monte Carlo simulations, several methods for detecting a trend in the magnitude of extreme values are compared. Ordinary least squares regression is found to be the least reliable method. A Kendall's tau-based method provides some improvement. The advantage of this method over that of least squares diminishes when the sample size is moderate to small. Explicit consideration of the extreme value distribution when computing trend always outperforms the above two methods. The use of the  $r$  largest values as extremes enhances the power of detection for moderate values of  $r$ ; the use of larger values of  $r$  may lead to bias in the magnitude of the estimated trend.

### 1. Introduction

Changes in climate variability and, in particular, in the intensity and frequency of climate extremes are likely to affect society more strongly than changes in the mean climate (Katz and Brown 1992). This observation has motivated many studies of extremes in the climate of the twentieth century (e.g., Groisman et al. 2001; Iwashima and Yamamoto 1993; Kunkel et al. 1999; Zhang et al. 2001a,b) and in the climate projected for the twenty-first century (e.g., Zwiers and Kharin 1998; Kharin and Zwiers 2000; Meehl et al. 2000).

Changes in extremes can be assessed by identifying changes in either frequency or intensity. Keim and Cruise (1998) describe a method for estimating a trend in the frequency of extreme events that is based on a simplified test introduced by Cox and Lewis (1966). Frei and Schär (2001) describe a closely related technique that uses a binomial distribution to model the counts of rare events and a logistic regression to estimate trend in the counts. Many other studies have investigated changes in the magnitude of the extremes. Some ascertain changes in return values by fitting separate ex-

treme value distributions in different periods (e.g., Zwiers and Kharin 1998; Kharin and Zwiers 2000); others attempt to compute linear trends in extreme values (e.g., Kunkel et al. 1999; Zhang et al. 2001a).

The objective of this paper is to compare the performance of several methods for detecting significant linear trends in the magnitude of extreme values. We consider methods that utilize only the block maxima from the available time series of observations and methods that utilize the available data more effectively by modeling trends in the  $r$  largest observations per block, where  $r$  is greater than 1. The trend estimation methods are explained in section 2. These methods are evaluated by means of a Monte Carlo experiment, the design of which is described in section 3. Results are given in section 4, followed by conclusions and discussion in section 5.

### 2. Methods

Under fairly general conditions, the distribution of the maximum of a sample of identically distributed variables converges to one of three types of extreme value (EV) distributions as the length of the sample goes to infinity. This result, which is known as the "three types theorem," was originally demonstrated by Fisher and Tippett (1928) and later was presented in a more general setting by Gnedenko (1943). The three types are the

---

*Corresponding author address:* Xuebin Zhang, Climate Research Branch, Meteorological Service of Canada, 4905 Dufferin St., Downsview, ON M3H 5T4, Canada.  
E-mail: xuebin.zhang@ec.gc.ca

Gumbel or EV-I distribution (Gumbel 1958), the Fréchet or EV-II distribution, and the Weibull or EV-III distribution.

After parameterization, the three distributions can be generalized as the generalized extreme value (GEV) distribution, with the distribution function

$$G(y) = \begin{cases} \exp\{-\exp[-(y - \mu)/\sigma]\}, & \xi = 0 & \text{(EV-I)} \\ \exp\{-[1 - \xi(y - \mu)/\sigma]^{1/\xi}\}, & \xi > 0, \quad y \leq \mu + \sigma/\xi & \text{(EV-II)} \\ \exp\{-[1 - \xi(y - \mu)/\sigma]^{1/\xi}\}, & \xi < 0, \quad y \geq \mu + \sigma/\xi & \text{(EV-III).} \end{cases} \quad (1)$$

The parameter  $\xi$  is referred to as the shape parameter;  $\mu$  and  $\sigma$  ( $>0$ ) are the location and scale parameters, respectively.

for estimating trends in extremes (OLS, nonparametric Kendall's tau, and maximum likelihood) in this paper.

The rate of convergence to these asymptotic distributions is affected by the shape of the upper tail of the distribution of the sampled random variable. For example, all distributions in the so-called exponential family belong to the "domain of attraction" of the Gumbel distribution, meaning that the distribution of the maximum of a large sample from a distribution in this family eventually converges to the Gumbel distribution as the sample becomes large. However, the rate of convergence is not the same for all members of the family, with much faster convergence for an exponentially distributed variable (Leadbetter et al. 1983) than for other members of the exponential family such as the Gaussian distribution.

*a. Kendall's tau*

Without loss of generality, we assume that  $t_1 \leq t_2 \leq \dots \leq t_n$  are the sampling times. Let

A time series  $(y_1, y_2, \dots, y_n)$  with a linear trend can be represented by

$$y_t = a + bt + \epsilon_t, \quad (2)$$

where  $\epsilon_t$  denotes a stationary noise process, and  $a$  and  $b$  are regression parameters. The magnitude of trend is computed by estimating coefficient  $b$ . A linear trend is said to have been detected if the estimate  $\hat{b}$  is significantly different from zero. When  $\epsilon_t$  follows a symmetric distribution with finite variance,  $b$  can be conveniently estimated by the method of least squares (OLS). If  $\epsilon_t$  is a Gaussian white-noise process, the statistical significance of the trend can be assessed using the Student's  $t$  test. Appropriate modifications can be made to the test if  $\epsilon_t$  is not white (e.g., von Storch and Zwiers 1999). Extreme values do not fall into a symmetric distribution, and thus it follows that least squares is not likely to be the most reliable method for identifying trends in the series of extreme values. A distribution-free, Kendall's tau-based estimator (Sen 1968) has frequently been used to estimate trend (Kunkel et al. 1999; Lins and Slack 1999; Zhang et al. 2001a,b), and the Mann-Kendall test is used to assess its statistical significance. A more rigorous method would be to model the extreme values with the method of maximum likelihood, with time being a covariate (e.g., Smith 1989; Coles 2001; Katz et al. 2002). We will consider all three methods

$$N = \sum_{1 \leq i < j \leq n} \delta(t_j - t_i), \quad (3)$$

where

$$\delta(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0. \end{cases} \quad (4)$$

Then, among all values of  $(t_j - t_i)$ ,  $1 \leq i < j \leq n$ , only  $N$  values are nonzero. Consider the set  $\mathcal{R}$  of the  $N$  distinct pairs  $(i, j)$  for which  $t_j > t_i$  and define

$$W_{ij} = (y_j - y_i)/(t_j - t_i), \quad (i, j) \in \mathcal{R}. \quad (5)$$

We arrange the  $N$  values in (5) in ascending order of magnitude and denote the  $k$ th smallest value by  $W_k$  ( $k = 1, 2, \dots, N$ ). The Kendall's tau estimator of  $b$  is the median of  $W_k$ , or

$$\hat{b} = \begin{cases} W_{(N-1)/2+1} & \text{if } N \text{ is odd,} \\ (W_{N/2} + W_{N/2+1})/2 & \text{if } N \text{ is even.} \end{cases} \quad (6)$$

That is, the trend is the median of all possible trend estimates obtained from pairs of observations obtained at distinctly different observing times. This estimator is statistically robust and is unbiased. When the time steps are equally spaced and random variables  $y_i$  are independent, this estimator has an asymptotic relative efficiency that is never less than 0.864 with respect to the least squares estimator (Sen 1968). The asymptotic relative efficiency is defined as the ratio between the asymptotic variance of the OLS estimator and that of the Kendall's tau estimator. The statistical significance of the Kendall's tau trend estimate can be assessed using the Mann-Kendall test (Mann 1945; Kendall 1955). Yue et al. (2002) showed that the power of the Mann-Kendall test in detecting significant trends depends on the underlying distribution.

### b. Generalized linear regression

The likelihood function for the GEV distribution (1) is given by

$$L = \prod_{i=1}^n \sigma^{-1} \exp \left[ - \left( 1 - \xi \frac{y_i - \mu}{\sigma} \right)^{1/\xi} \right] \times \left( 1 - \xi \frac{y_i - \mu}{\sigma} \right)^{(1/\xi)+1}, \quad (7)$$

where  $y_1, y_2, \dots, y_n$  are  $n$  extreme values. These are generally block maxima taken as the largest values from blocks of samples of the parent distribution (e.g., annual maxima of daily precipitation). The method of maximum likelihood can be used to estimate the three parameters of the GEV distribution by minimizing  $-\log(L)$ . Covariates can be introduced into the GEV model by, for example, expressing the location parameter as

$$\mu(t) = \alpha + \beta t. \quad (8)$$

This allows for the estimation of linear trend in the location parameter (see, e.g., Smith 1989). Trend can also be considered for the scale parameter, typically with an equation of the form

$$\log[\sigma(t)] = \theta + \varphi t. \quad (9)$$

Note that this formulation ensures that the scale parameter remains positive. Trend in the shape parameter  $\xi$  is not considered in this study because we decided to avoid the complications that arise from allowing all three GEV parameters to vary in time. We assume that it is not likely for there to be significant change in the shape of the tails of the kinds of variables that are typically considered in climate studies over the period of record (less than 100 yr) that is ordinarily available for analysis. Situations in which the tail does lengthen, or shorten, modestly relative to the main body of the distribution can be dealt with approximately by varying the scale parameter.

The trend in the extreme values can be estimated by choosing  $\alpha$ ,  $\beta$ , and  $\sigma$ , or  $\theta$  and  $\varphi$ , to minimize  $-\log(L)$ . The statistical significance of the trend is evaluated by using a likelihood ratio test to compare a model with trend with one that does not include trend. Let  $M_0$  be a model for the extreme values with no trend (i.e.,  $\beta = \varphi = 0$ ) and let  $M_1$  be a model with trend. Also, let  $l^1$  and  $l^0$  be the log likelihoods under the models  $M_1$  and  $M_0$ . When trends are not present, the log likelihood ratio statistic (LRS)

$$T = 2(l^1 - l^0) \quad (10)$$

is asymptotically  $\chi_q^2$  distributed (Cox and Hinkley 1974), with  $q$  being the difference in the number of free parameters in the two models. We reject hypothesis  $M_0$  (there is no trend) at significance-level  $\alpha$  if  $T$  is bigger than the upper- $\alpha$  point of the  $\chi_q^2$  distribution. Further details can be found in Coles (2001). In the case of two

parameter models, we are most interested in the existence of trends in the parameters that lie in the same direction. However, we note that it is possible for there to be opposing trends in the location and scale parameters that result in a very small trend in the extremes.

Note that the results of the likelihood ratio test are not always directly comparable with those of the OLS and Kendall–Mann tests. The latter are designed to detect change in the mean value of the extremes, which is given by  $\mu(t) + \sigma(t)[1 - \Gamma(1 + \xi)]/\xi$  when the extremes have a GEV distribution with location and scale parameters that may contain trend. The LRS, on the other hand, is designed to detect change in one or more of the parameters of the GEV distribution, either individually or in combination. These two types of tests do ask equivalent questions when it can be assumed that the scale and shape parameters are constant in time.

### c. The $r$ -largest method

As noted above, it may be possible to improve on the trend estimates made from block maxima by using multiple extremes per block, thereby utilizing the available record more completely. For example, annual maximum daily precipitation may be extracted from daily or even hourly observations, and, thus, fitting an extreme-value distribution with annual maximum daily precipitation may discard some potentially useful information about precipitation extremes. One possible solution is to use the  $r$ -largest values in a block for small values of  $r$ . This is usually called the  $r$ -largest method.

By proper scaling, the  $r$ -largest-order statistics model gives a likelihood function with parameters that correspond to those of the GEV model for block maxima but that incorporates more of the observed extreme data. The likelihood function (7) for the  $r$  largest values taken for every block becomes

$$L = \prod_{i=1}^n \left\langle \exp \left\{ - \left[ 1 - \xi \frac{y_i^{(r)} - \mu}{\sigma} \right]^{1/\xi} \right\} \times \prod_{j=1}^r \sigma^{-1} \left[ 1 - \xi \frac{y_i^{(j)} - \mu}{\sigma} \right]^{(1/\xi)+1} \right\rangle, \quad (11)$$

where  $y_i^{(j)}$  is the  $j$ th-order statistic in block  $i$ . Trend in the extreme values can then be estimated by replacing  $\mu$  and  $\sigma$  with (8) and (9), respectively. Relative to a standard block-maxima analysis, the interpretation of the parameters is unaltered, but precision should be improved by the inclusion of extra information. The power of detection of trend would also improve, it is hoped.

The choice of  $r$  is a classic bias–variance trade-off: a small value of  $r$  results in a smaller sample of extremes and, thus, more uncertain parameter estimates and possibly lower power of detection trends. On the other hand, the GEV model is an approximation that, strictly speaking, holds only in the asymptotic limit. Errors in the approximation are likely to be more apparent for large

values of  $r$ . These errors will be especially apparent in the far tails of the distribution that lie beyond the range of observations contained in the sample of extremes (e.g., in applications for which it is necessary to estimate long-period return values from short samples). In principle,  $r$  should be chosen to be as large as possible without inducing large biases in the properties, such as return values, that are to be estimated from the fitted distribution. This is, to a large degree, a matter of judgment. The accuracy of the fit for a particular  $r$  within the available sample can, and should, be examined using standard diagnostics such as probability and quantile plots.

In this study, we compare estimated return values at 20, 100, 1000 yr obtained by fitting  $r$ -largest models with  $r$  ranging from 1 to 30 to simulated samples containing 50 or 100 yr of data as described below. Note that the GEV model is a special case of the  $r$ -largest model ( $r = 1$ ). We also test performance by fitting these models to observed precipitation at 200 stations located in a  $10^\circ$  latitude  $\times$   $10^\circ$  longitude box ( $40^\circ$ – $50^\circ$ N and  $80^\circ$ – $90^\circ$ W).

Note also that daily weather persists. For example, days with precipitation tend to occur in clusters (Katz et al. 2002). This can be a cause of concern when a relatively large  $r$  is used and if heavy rain concentrates in a particular season. Thus, it may be appropriate to “decluster” the data (Todorovic and Zelenhasic 1970). Under such a scheme one would select the  $r$  largest values in different clusters: for example, to select extreme daily precipitation separated by several days and to choose extreme high temperatures in different hot spells.

### 3. Monte Carlo simulation

Our Monte Carlo simulation experiment was designed to mimic a real situation that a climatologist might encounter in which the objective is to estimate trend in annual extremes. In many places of the world where the annual cycle is strong, annual extremes may only occur in a particular season. Thus, the annual extremes may actually be drawn from a seasonal sample that represents a period much shorter than 1 yr. For this reason, we also considered seasonal extremes. We produced 1000 simulations of daily precipitation (using an exponential-family distribution) with predetermined trends in the extremes. Each simulation represents 50 or 100 yr of daily precipitation data. A similar experiment was conducted with simulated Gaussian-distributed temperature data. Results were very similar to those reported below.

The daily precipitation time series are simulated in two steps. The first step is to simulate the frequency of precipitation days for each year (season). To produce a realistic level of interannual variability, we used a normal distribution, with its mean and standard deviation computed from observational data, to simulate the number of wet days for each year (season). The second step

is to simulate the precipitation amounts for the wet days. We use an exponential distribution for this purpose, because 1) this distribution has been used to simulate daily precipitation amounts in a popular weather generator (Richardson 1981) and 2) the true trend in the mean of simulated extremes is then easily prescribed, as shown below.

The cumulative distribution function of an exponential random variable  $X$  is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 - e^{-x/\lambda} & \text{if } x > 0. \end{cases} \quad (12)$$

Note that we follow the notational convention used in the statistical literature, where upper-case symbols designate random variables, and the corresponding lower-case variables indicate realizations of those variables. The mean and variance of  $X$  are

$$\mu_x = \lambda \quad \text{and} \quad \delta_x^2 = \lambda^2, \quad (13)$$

respectively. Let  $X_1, \dots, X_m$  represent  $m$  simulated daily precipitation amounts in a year. For convenience, they have been simulated so that they are independent, identically distributed, exponential random variables. Let  $Y$  be the annual maximum daily precipitation, that is, the maximum of  $\{X_1, \dots, X_m\}$ . Then,  $Y < y$  if and only if  $X_i < y$  for each  $i = 1, \dots, m$ . Using independence, we obtain that

$$\begin{aligned} P(Y < y) &= \prod_{i=1}^m P(X_i < y) = F^m(y; \lambda) \\ &= (1 - e^{-y/\lambda})^m \approx \exp(-me^{-y/\lambda}) \\ &= \exp[-e^{-y/\lambda + \lambda \log(m)/\lambda}] = \exp[-e^{-(y-\mu)/\lambda}]. \end{aligned} \quad (14)$$

This is the Gumbel distribution function with a location parameter  $\mu = \lambda \log(m)$  and a scale parameter  $\sigma = \lambda$ .

The mean of the Gumbel distribution described above is given by

$$\mu_y = \mu + \gamma\lambda = \lambda \log(m) + \gamma\lambda, \quad (15)$$

where  $\gamma$  is an Euler’s constant ( $\gamma \approx 0.577$ ). It is clear that the mean can be affected by trend in both the location and scale parameters. A trend in the number of wet days in a year ( $m$ ) would result in a trend in the location parameter whereas a trend in the intensity of precipitation ( $\lambda$ ) would cause a trend in both the location and scale parameters of the Gumbel distribution. If there is no change in the frequency of precipitation events (i.e., if  $n$  is fixed), the trend in the mean of the extremes is only a function of the trend in the intensity. For the sake of simplicity, we consider only such trend, but we note that in the real world there is observational evidence of changes in both frequency (e.g., Zhang et al. 2001b) and intensity (Karl and Knight 1998).

The baseline frequency and intensity parameters used in our simulation study were estimated from a Canadian

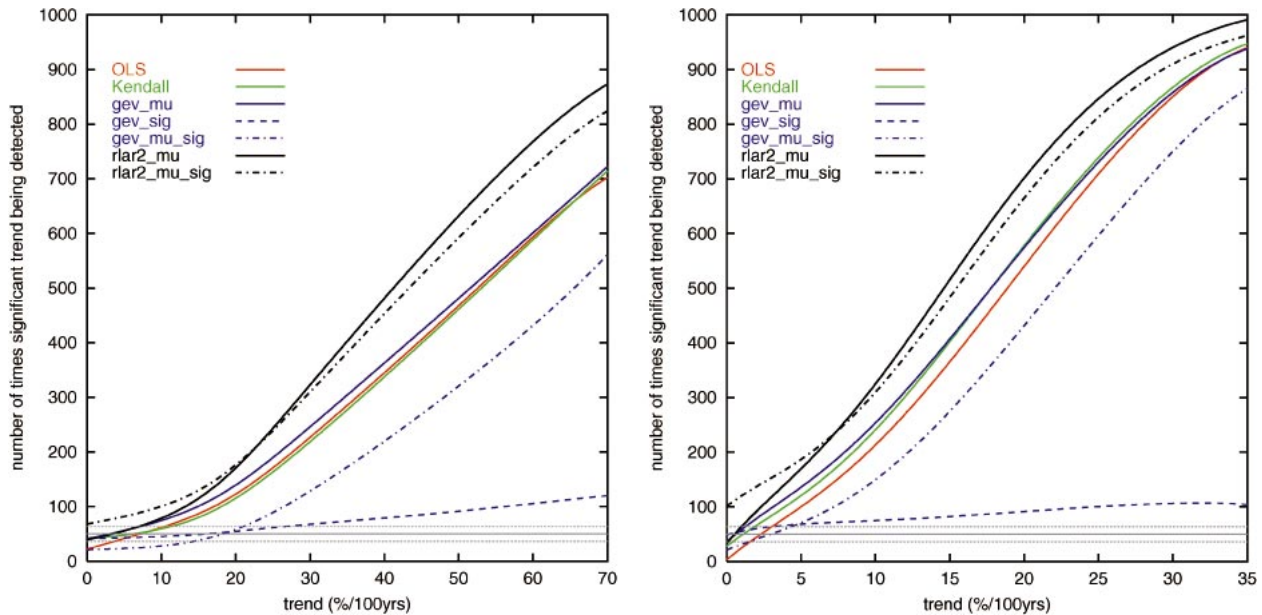


FIG. 1. The rate at which a statistically significant trend was detected at the 5% level in 1000 simulations. Results are shown for samples of size of (left) 50 and (right) 100 yr. The nominal 5% significance level is indicated by the gray line. The 95% acceptance region for testing the null hypothesis that the rejection rate is equal to the nominal level is given by the dashed gray lines. “OLS” and “Kendall” indicate linear least squares fit and Kendall’s tau-based estimator results; “gev\_mu,” “gev\_sig,” and “gev\_mu\_sig” represent linear trends being considered in the location parameter, in the logarithm of scale parameter, and in both location parameter and the logarithm of scale parameter of the GEV distribution, respectively. Only annual maximum daily values were used in these five cases. The labels “rlar2\_mu” and “rlar2\_mu\_sig” indicate results from the use of the  $r$ -largest method with  $r = 2$ , and trend in the location parameter, and location and log-scale parameters, respectively.

station (Kemano) located at  $54.05^{\circ}\text{N}$ ,  $128.63^{\circ}\text{W}$  in British Columbia. The annual number of wet days at this station has a mean of 123 days and a standard deviation of 27 days. The average precipitation intensity is  $12 \text{ mm day}^{-1}$ . The average annual maximum daily precipitation is about 65 mm when computed using (15). We use the summer data for the seasonal time series. The average number of wet days in this case is 31 with a standard deviation of 5 days. The precipitation intensity is  $7 \text{ mm day}^{-1}$ . This results in a long-term mean seasonal maximum daily precipitation of about 28 mm. Linear trends of various magnitudes were imposed on the intensity parameter  $\lambda$ . The resulting trends computed from (15) are considered to be the “true” trends in our evaluation of trend detection methods.

Three statistics have been used to aid comparison among the different methods. They are the bias, root-mean-square error (rmse), and variance. The bias is the difference between the true trend and various trend estimates described above. The rmse is the square root of average of bias squared. Variance is the standard deviation of trend estimate in a particular simulation.

#### 4. Results

For the annual series, the number of times that a statistically significant trend was detected at the 5% level in 1000 simulations is plotted in Figs. 1 and 2. Results

are displayed for  $n = 50$  and 100 and for the different methods discussed in section 2. Note that Fig. 1 includes results for the  $r$ -largest method when  $r = 1$  and  $r = 2$ . For clarity, comparable results for  $r = 5$  are displayed in Fig. 2. When using the  $r$ -largest method, we considered trend in the location parameter with a trend described by (8) and in the scale parameter as in (9). We also considered trend in both the location and scale parameters, because the simulated extreme values time series would have linear trends in both parameters.

When there is no trend in the time series, the number of times a significant trend is detected by the various methods is generally less than the nominal level of 5%. Exceptions occur for some of the  $r$ -largest methods when  $r > 1$  and when trends are allowed in both location and scale parameters. Those methods detect trend significantly more often than 5% of the time when none is present. We examined their results in detail and found that “significant” trends of opposite sign were often identified in the location and scale parameters, yielding almost zero trend in the mean. This finding suggests that caution is needed in interpreting results when a two-parameter-trend model has been used.

The linear least squares (OLS) trend estimates have the largest variance among the methods considered. The power of trend detection obtained with this method is also generally low. Both effects are the result of violating the distributional assumption that underlies the

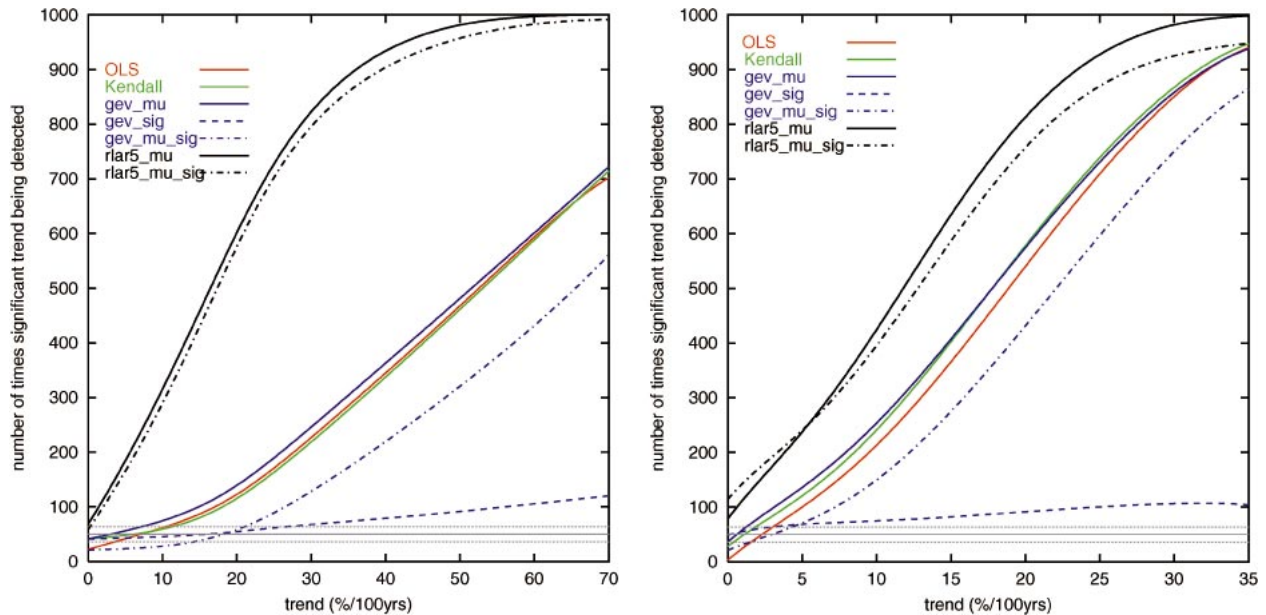


FIG. 2. As Fig. 1, except “rlar5\_mu” and “rlar5\_mu\_sig” indicate results from the use of the  $r$ -largest method with  $r = 5$ , and trend in the location parameter, and location and log-scale parameters, respectively.

OLS method. The Kendall’s tau–based method is found to be more powerful than the OLS method when the sample size is large (100). The  $r$ -largest methods with trend in the location parameter consistently outperform both the OLS and Kendall methods, suggesting the usefulness of doing proper modeling. The number of detections by the  $r$ -largest method increases with  $r$  as can be seen by comparing Fig. 1 and 2. This emphasizes the importance of making better use of the information available in the data. The importance of larger sample sizes is undeniable: the power of detection is always larger even when the magnitude of trend is halved if the sample size is doubled.

We expected that the  $r$ -largest method with trend in both parameters would perform the best because the trend imposed in the simulation data should result in trend in both the location and scale parameters. Our simulations indicated that these “better” methods were not as powerful as the methods that only consider trend in the location parameter, however. This suggests that the potential gain in the power of detection from including a trend in scale parameter is lost because of the need to estimate an extra parameter. Another possibility is that, in the case of the scale parameter, the signal-to-noise ratio may be too small to allow reliable detection. It is also possible that (9) may not approximate the prescribed linear trend in the scale parameter well enough. This may not be the main reason, however. We replaced (9) with  $\log[\sigma(t)] = \theta + \varphi t^{1/2}$  and  $\log[\sigma(t)] = \theta + \varphi t^{1/4}$  and did not gain improvement in the power of detection. The rate of detection should not be used as the sole criterion in selecting a method for trend analysis of extremes, however, because a more powerful

method may also have a larger bias (i.e., reject the null hypotheses of no trend more frequently than expected when there is, in fact, no trend).

The biases, variances, and root-mean-square errors of the trend estimates from the different methods for  $n = 100$  are presented in Table 1. In agreement with Fig. 1, the variance and rmse of the OLS estimates are larger than those of Kendall’s tau. The variances of the estimate made with the  $r$ -largest method with trend in the location parameter only are smaller than these of OLS and Kendall’s tau, and the bigger the  $r$  is, the smaller the variance is. However, the rmse are not always smaller, because the trend estimated by those methods only reflects changes in the scale parameter and, hence, is biased toward smaller values by design. The bigger the  $r$  value is, the bigger the bias is. The  $r$ -largest estimates with trend in both location and scale parameters have substantially less bias, but this comes at the cost of larger variance that may actually increase the rmse.

The results for the seasonal series are similar to those for annual series, except that the differences among the GEV models, Kendall’s tau, and linear least squares are smaller. The difference between seasonal and annual extreme results is likely due to the difference in the intensity of the precipitation extremes that were simulated on these two time scales. The somewhat less intense summer extremes, coupled with the shorter summer sample, result in trends that are somewhat less detectable. As a result, the simulations discriminate between the different methods less strongly.

The mean 1000-yr return value estimate obtained from 1000 simulations of the series annual extremes in which there is no trend is almost identical regardless of

TABLE 1. Comparison between biases (bias), variances (var), and root-mean-square errors (rmse) of linear trends, calculated by fitting different models to 1000 simulated samples of size 100. The models are identified with the same label as in Fig. 1. Unit for bias, var, and rmse are millimeters. Trend is represented by percentage change in the mean over a 100-yr period.

Trend		OLS	Kendall	gev_mu	gev_mu_sig	rlar2_mu	rlar2_mu_sig	rlar5_mu	rlar5_mu_sig
0	Bias	0.2	0.0	0.1	0.0	0.0	0.2	-0.2	0.1
	Var	5.1	4.5	4.0	5.0	3.0	4.5	2.2	4.8
	Rmse	5.1	4.5	4.0	5.0	3.0	4.6	2.2	4.8
10	Bias	-1.5	-1.4	-2.0	-1.0	-2.3	-1.4	-3.3	-1.6
	Var	5.5	5.0	4.9	5.7	3.4	5.0	2.3	4.5
	Rmse	5.7	5.2	5.2	5.9	4.1	5.3	4.0	4.8
20	Bias	0.0	0.2	-1.2	0.9	-2.6	1.1	-4.3	0.4
	Var	6.1	5.4	5.0	5.9	3.6	5.2	2.5	5.0
	Rmse	6.1	5.4	5.3	6.0	4.5	5.3	5.0	5.0
30	Bias	0.3	-0.2	-1.9	0.8	-3.8	1.9	-6.5	1.4
	Var	6.3	5.4	5.3	6.1	3.8	5.5	2.5	5.0
	Rmse	6.3	5.4	5.6	6.2	5.4	5.9	7.0	5.3
40	Bias	-0.1	-1.0	-3.8	0.7	-6.7	2.1	-8.8	1.5
	Var	6.3	5.4	5.7	6.2	4.0	5.8	2.9	5.2
	Rmse	6.3	5.5	6.9	6.2	7.9	6.2	9.2	5.5

whether  $r = 1$  (GEV) or  $r = 30$ . For the simulated seasonal data, the 100-yr return values drop quickly for  $r > 10$ . The mean 1000-yr return value is about 5% smaller than that obtained from the GEV model when  $r > 5$ . This suggests that choosing  $r = 5$  is probably appropriate for seasonal and annual extremes when characteristics of the extremes are similar to those of the simulated extremes, and a larger value of  $r$  may be appropriate for annual extremes.

In agreement with our findings from the simulated data, the mean 200-station 100-yr return value derived from the annual maxima of observed daily precipitation varies less than 5% when  $r$  ranges from 1 to 30. However, the mean 200-station 1000-yr return value is about 10% smaller than that obtained with the GEV model, even when  $r = 2$ . This suggests that the behavior of the observed extremes is somewhat different from the simulated extremes. Clearly,  $r$  should be selected conservatively if the objective is to estimate return values for periods that are substantially longer than the period of record.

## 5. Conclusions and discussion

We have compared the ability of several methods to detect trends in extreme values. Our comparison was made with the aid of a Monte Carlo simulation experiment. The extreme values we simulated contain linear trends in both the location and scale parameters.

The OLS method, which requires the residual time series to be normally distributed, has very poor power because of the violation of the distributional assumption. A Kendall rank correlation-based method, which does not require a distributional assumption, has been frequently used as an alternative to OLS in the literature. It outperforms OLS only when the sample size is large.

A generalized linear regression method, which explicitly incorporates trend into the parameters of the generalized extreme value distribution, has stronger

power of detection when compared with the OLS and Kendall methods. The GEV method that considers trend only in the location parameter does a better job than a method that considers trends in both location and scale parameters, suggesting the advantage of using a more parsimonious model.

The  $r$ -largest method that uses more than one extreme per (annual) block significantly improves the performance of the GEV method. The magnitude of trend is, however, underestimated when only including a trend term in the location parameter, and the larger the value of  $r$  is, the bigger the bias is. As always, one should try to reduce the number of parameters to be estimated when computing trend with  $r$  largest methods. A practical approach would be to estimate trend in one parameter at a time and to use a more complex model that includes trends in multiple parameters only when significant trends have been separately detected in the individual parameters.

Extreme values are scarce by definition, meaning that estimates are often required for levels of a process that are greater than have already been observed. This calls for a proper analysis of extremes. Our analysis demonstrates very clearly the advantage of using extreme-value theory in analyzing trends for extremes. There have been relatively few studies that have used extreme-value theory to model, detect, or project trends in extremes of weather and climate. In addition, making effective use of the available information is important to trend detection in extremes. Thus, the use of an  $r$ -largest method is encouraged. It should be cautioned, however, that a large value of  $r$  will amplify errors in the GEV approximation of the behavior of extremes from finite samples. These errors will, in turn, lead to bias in the estimation of trends and return values. Thus, moderation should be exercised in the choice of  $r$ . One must also make sure that the observed extreme values lie within the domain of attraction of the GEV distribution by testing the residual against a GEV distribution.

A final note is that our simulation experiments did not directly account for variations in the persistence (serial correlation) of real-world precipitation. Introducing serial correlation into the simulations would, in effect, reduce the size of the sample from which the annual or seasonal extreme has been drawn. Although we have focused primarily on extremes taken from annual samples of independent, identically distributed rainfall amounts, we draw the same conclusions from the much smaller seasonal samples. Therefore, the guidance that we provide on the relative performance of the different trend assessment techniques and the choice of  $r$  should also hold for moderately serially correlated data.

*Acknowledgments.* This study was supported by Canada's Climate Change Action Fund. We thank Gabi Hegerl, Jian Sheng, Val Swail, and two anonymous reviewers for their comments, which improved this paper. We also thank S. Coles for his S-plus routines (available online at <http://www.maths.lancs.ac.uk/~coles/>) used in this analysis and for helpful discussions.

#### REFERENCES

- Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag Heidelberg, 208 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, 1974: *Theoretical Statistics*. Chapman and Hall, 511 pp.
- Fisher, R. A., and L. H. C. Tippett, 1928: Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, **24**, 180–190.
- 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.
- Gnedenko, B. V., 1943: Sur la distribution limite du terme maximum d'une série aléatoire (Limiting distribution of maximum values of random series). *Ann. Math.*, **44**, 423–453.
- Groisman, P. Y., R. W. Knight, and T. R. Karl, 2001: Heavy precipitation and high streamflow in the contiguous United States: Trends in the twentieth century. *Bull. Amer. Meteor. Soc.*, **82**, 219–245.
- Gumbel, E. J., 1958: *Statistics of Extremes*. Columbia University Press, 375 pp.
- Iwashima, T., and R. Yamamoto, 1993: A statistical analysis of the extreme events: Long-term trend of heavy daily precipitation. *J. Meteor. Soc. Japan*, **71**, 637–640.
- Karl, T. R., and R. W. Knight, 1998: Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Amer. Meteor. Soc.*, **79**, 231–241.
- Katz, R. W., and B. G. Brown, 1992: Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, **21**, 289–302.
- , M. B. Parlange, and P. Naveau, 2002: Statistics of extremes in hydrology. *Adv. Water Resour.*, **25**, 1287–1304.
- 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 and Company, 196 pp.
- Kharin, V. V., and F. W. Zwiers, 2000: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *J. Climate*, **13**, 3760–3788.
- Kunkel, K. E., K. Andsager, and D. R. Easterling, 1999: Long-term trends in extreme precipitation events over coterminous United States and Canada. *J. Climate*, **12**, 2515–2527.
- Leadbetter, M. R., G. Lindgren, and H. Rootzén, 1983: *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, 336 pp.
- Lins, H. F., and J. R. Slack, 1999: Streamflow trends in the United States. *Geophys. Res. Lett.*, **26**, 227–230.
- Mann, H. B., 1945: Non-parametric tests against trend. *Econometrica*, **13**, 245–259.
- Meehl, G. A., F. W. Zwiers, J. Evans, T. Knutson, L. Mearns, and P. Whetton, 2000: Trends in extreme weather and climate events: Issues related to modeling extremes in projections of future climate change. *Bull. Amer. Meteor. Soc.*, **81**, 427–436.
- Richardson, C. W., 1981: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resour. Res.*, **17**, 182–190.
- Sen, P. K., 1968: Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Stat. Assoc.*, **63**, 1379–1389.
- Smith, R. L., 1989: Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Stat. Sci.*, **4**, 367–393.
- Todorovic, P., and E. Zelenhasic, 1970: A stochastic model for flood analysis. *Water Resour. Res.*, **6**, 1641–1648.
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.
- Yue, S., P. Pilon, and G. Cavadias, 2002: Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *J. Hydrol.*, **259**, 254–271.
- Zhang, X., K. D. Harvey, W. D. Hogg, and T. R. Yuzyk, 2001a: Trends in Canadian streamflow. *Water Resour. Res.*, **37**, 987–998.
- , W. D. Hogg, and E. Mekis, 2001b: Spatial and temporal characteristics of heavy precipitation events over Canada. *J. Climate*, **14**, 1923–1936.
- Zwiers, F. W., and V. V. Kharin, 1998: Changes in the extremes of climate simulated by CCC GCM2 under CO<sub>2</sub> doubling. *J. Climate*, **11**, 2200–2222.