ELSEVIER

Journal of Hydrology 240 (2000) 90–105

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Trends in floods and low flows in the United States: impact of spatial correlation

E.M. Douglas[a], R.M. Vogel[a,*], C.N. Kroll[b]

[a]Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155, USA
[b]College of Environmental Science and Forestry, 1 Forestry Drive, State University of New York, Syracuse, NY 13210, USA

Received 4 April 2000; revised 25 August 2000; accepted 5 September 2000

Abstract

Trends in flood and low flows in the US were evaluated using a regional average Kendall's *S* trend test at two spatial scales and over two timeframes. Field significance was assessed using a bootstrap methodology to account for the observed regional cross-correlation of streamflows. Using a 5% significance level, we found no evidence of trends in flood flows but did find evidence of upward trends in low flows at the larger scale in the Midwest and at the smaller scale in the Ohio, the north central and the upper Midwest regions. A dramatically different interpretation would have been achieved if regional cross-correlation had been ignored. In that case, statistically significant trends would have been found in all but two of the low flow analyses and in two-thirds of the flood flow analyses. We show that the cross-correlation of flow records dramatically reduces the effective number of samples available for trend assessment. We also found that low flow time series exhibit significant temporal persistence. Even when the serial correlation was removed from the time series, significant trends in low flow series were apparent, though the number of significant trends decreased. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Floods; Low flows; Trends; Spatial correlation; Field significance; Hydroclimatology; Streamflow

1. Introduction

Records of atmospheric concentrations of carbon dioxide indicate a dramatic increase since the beginning of the Industrial Revolution. It is generally believed that such an increase in CO_2, a "green house" gas, could result in increased global mean temperatures. Bloomfield (1992) reported statistically significant rates of mean global temperature increase between 0.4 and 0.6°C per century. Results of general circulation model studies indicate that increased global temperatures could lead to regional increases in the amount and intensity of rainfall. This prediction has been verified for the North American continent by Vinnikov et al. (1990), Guttman et al. (1992), Groisman and Easterling (1994), and Karl and Knight (1998), among others, who found increases in precipitation amount and intensity across the US and Canada in recent years. The sensitivity of streamflow to changes in precipitation, and other climate parameters, is well documented, hence it is informative to investigate whether streamflow records exhibit evidence of increasing trends which may be linked to climate change.

1.1. Review of literature

A number of recent studies in the US have

* Corresponding author. Fax: +1-617-627-3994.
E-mail addresses: edougl01@tufts.edu (E.M. Douglas), rvogel@tufts.edu (R.M. Vogel), cnkroll@esf.edu (C.N. Kroll).

PII: S0022-1694(00)00336-X

investigated the presence of trends in streamflow data. The results of those studies vary widely depending on the spatial scale and location of the study area, with most of the significant trends occurring in the Midwest. Lettenmaier et al. (1994) detected strong increases in monthly streamflow across the United States during November through April for the period 1948 to 1988, with the largest trend magnitudes occurring in the north central region (Michigan, Illinois, Wisconsin and Minnesota). Hubbard et al. (1997) reported increases in annual runoff in 16 of 20 major hydrologic regions across the United States. Smith and Richman (1993) found increases in mean annual streamflow in Illinois ranging from 20 to 80% during the period from 1950 to 1987. Changnon and Kunkel (1995) found significant upward trends in floods in the northern Midwest during the period 1921–1985, and found a link between these trends and higher precipitation. Olsen et al. (1999) found large and statistically significant upward trends in flood flows over the last 100 years in the Upper Mississippi and Missouri rivers. Pupacko (1993) found a slight (non-significant) trend of increasing and more variable winter streamflow in the northern Sierra Nevada since the mid-1960s. Most recently, Lins and Slack (1999) reported increasing trends across the US in lower magnitude streamflow quantiles (annual minimum through the 70th quantile) but not at higher quantiles (90th quantile and annual maximum).

1.2. The Importance of spatial correlation

Of all the previously cited studies, Lettenmaier et al. (1994) performed the only trend study that accounted for the spatial correlation of flow records. The fact that most studies ignored the role of spatial correlation in the interpretation of their results belies the importance of such correlation in statistical analysis. However, this oversight is not without good cause since the assumption of independent observations is paramount to many trend tests.

The effect of spatial and/or temporal correlation among datasets on hypothesis testing is twofold. First, cross-correlation creates an overlap in the information contained in each datapoint. For example, if flood flows are spatially correlated (cross-correlated) and a trend is found at a site, one is more likely to find trends at nearby sites as well. From a statistical perspective, correlation reduces the effective sample size of the dataset. This results in a more "liberal" hypothesis test, meaning that, if correlation is ignored, the null hypothesis (of independence) will tend to be rejected more frequently than it should be. Second, the presence of correlation makes the analytical derivation of an exact probability distribution for the test statistic difficult, in which case an approximate distribution must be developed.

1.3. Study goals

Olsen et al. (1998) argue that the major impact of non-stationary behavior of a random variable is manifested in the extremes. This impact has been observed in climate records by Karl and Knight (1998) who reported that the proportion of total precipitation within the US contributed by extreme events (upper 10% of daily precipitation amounts) has increased significantly since the early 1900s. Similar trends in streamflow extremes, if they exist, would directly impact the accuracy of hydrologic analysis and design. Many studies have investigated the existence of trends in flood flows but few have performed the same analysis at the opposite extreme, in low flows. Trends in one or both of these variables could be seen as potential evidence of climate change and its impact on the hydrologic cycle, which could eventually lead to shifts in the availability of water across the US. Such climate change impacts have reportedly been observed within the last few years (Trenberth, 1999). Infrastructural accommodation to such shifts would be both environmentally and economically costly. Therefore, proper statistical investigation of the existence of such trends is paramount. Violation of the assumption of spatial independence of datasets, as is commonplace, can result in misleading and erroneous interpretations of the climate and/or streamflow record. In light of this, the objectives of this study are:

- To investigate the existence of trends in flood and low flows in such a manner that the potential impact of climate change can be assessed and so that our results can be compared with previous studies.
- To evaluate the effect of spatial correlation of flow records on the interpretation of hypothesis test

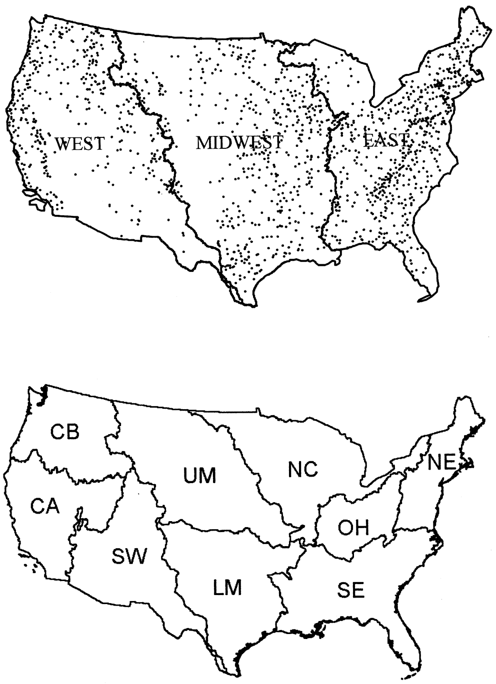

Fig. 1. Delineation of regions used for trend analyses: (top) three geographical regions with locations of HCDN stations; and (bottom) nine superregions, following Lettenmaier et al. (1994).

results by developing a hypothesis test that accounts for the spatial correlation of the data, thereby allowing comparisons with the results of analyses in which the spatial independence of flow records has been assumed.

2. Methodology

2.1. Data

Analyses were performed on data contained in the Hydro-Climatic Data Network (HCDN), a dataset compiled by Slack et al. (1993) which is comprised of average streamflow values recorded on a daily, monthly and annual basis at 1571 gaging stations across the continental US (see Fig. 1a for locations of gaging stations). The HCDN contains streamflow records collected between 1874 and 1988, with an average station record length of approximately 48 years. The basins represented in the HCDN are relatively free of anthropogenic impacts and therefore are ideal for investigating climate-induced changes to the hydrologic cycle. Only stations with records suitable at a daily timescale (Timescale equal to D) were used in this study, which reduced the total number of usable stations to 1474.

Flood flows for each station were recorded as the maximum average daily streamflow within each water year (beginning in October and ending in September of the following calendar year). Low flow series for each station were obtained by calculating a 7-day moving average from average daily discharge measurements. The smallest 7-day average was retained and recorded as that year's low flow. For the low flow data, the drought year (beginning in April and ending in March of the following calendar year) was used rather than the water year. Regional analysis was performed at two spatial scales including the three major geographic regions shown in Fig. 1a (East, Midwest and West) and the nine hydrologic "superregions" shown in Fig. 1b (following Lettenmaier et al., 1994). The presence of regional trends was evaluated over two timeframes: 30-year (1959–1988) and 50-year (1939–1988) periods. These timeframes most closely approximate those used by Lins and Slack (1999). The timeframes for the low flow data began and ended one year earlier than those for the flood flows, because the drought year begins six months ahead of the water year; therefore, 1988 did not have a full data set from which to calculate a low flow.

2.2. Hypothesis test

Most previous analyses of trends in hydrologic data have been performed using what has become known as the Mann–Kendall trend test. This hypothesis test is a non-parametric, rank-based method for evaluating the presence of trends in time-series data. The data are ranked according to time and then each data point is successively treated as a reference data point and is compared to all data points that follow in time. The test statistic, Kendall's S, (Kendall, 1962) is calculated as

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sign}(x_i - x_j) \tag{1}$$

where x is the data point (in this case, a flood or low flow value) at times i and j and sign( ) is equal to $+1$ if x_i is greater than x_j and -1 if x_i is less than x_j. For independent, identically distributed (iid) random variables with no tied data values

$$E(S) = 0 \tag{2}$$

$$\mathrm{Var}(S) = \frac{n(n-1)(2n+5)}{18} = \sigma^2 \tag{3}$$

When some data values are tied, the correction to Var(S) is

$$\mathrm{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{n} t_i(i)(i-1)(2i+5)}{18} \tag{4}$$

where t_i denotes the number of ties of extent i (i.e. a dataset with two tied values would have one tie of extent two or $i = 2$ and $t_2 = 1$). For n larger than 10, the test statistic

$$Z_s = \begin{cases} \dfrac{S-1}{\sigma} & \text{for } S > 0 \\ \dfrac{S+1}{\sigma} & \text{for } S < 0 \\ 0 & \text{for } S = 0 \end{cases} \tag{5}$$

follows a standard normal distribution (Kendall, 1962). Local (at-site) significance levels (p-values) for each trend test can be obtained from the fact that

$$p = 2[1 - \Phi(|Z_s|)] \tag{6}$$

where $\Phi(\,)$ denotes the cumulative distribution function (cdf) of a standard normal variate.

In order to evaluate trends at a regional scale rather than at individual sites, we employ a new test statistic, which we term the regional average Kendall's S ($\bar{S}_m$) computed as

$$\bar{S}_m = \frac{1}{m}\sum_{k=1}^{m} S_k \tag{7}$$

where S_k is Kendall's S for the kth station in a region with m stations. For (spatially) iid flow records of length n with no tied data values, the mean and variance of $\bar{S}_m$ are

$$E(\bar{S}_m) = 0 \tag{8}$$

$$\mathrm{Var}(\bar{S}_m) = \frac{n(n-1)(2n+5)}{18m} = \frac{\sigma^2}{m} \tag{9}$$

Var($\bar{S}_m$) may be corrected for the presence of tied data values as in Eq. (4). If $\bar{S}_m$ is calculated from iid data, then by the Central Limit Theorem, the distribution of $\bar{S}_m$ will be approximately normal for large m. Therefore, a normalized test statistic, Z_m, can be calculated as

$$Z_m = \frac{\bar{S}_m - E(\bar{S}_m)}{\sigma/\sqrt{m}} = \frac{\bar{S}_m}{\sigma/\sqrt{m}} \mathcal{N}(0,1) \tag{10}$$

and the significance of Z_m can be computed from the cdf of a standard normal variate. However, if the data from which $\bar{S}_m$ is calculated are cross-correlated, the variance becomes

$$\mathrm{Var}(\bar{S}_m) = \frac{1}{m^2}\left[\sum_{k=1}^{m} \mathrm{Var}(S_k) + 2\sum_{k=1}^{m-1}\sum_{l=1}^{m-k} \mathrm{Cov}(S_k, S_{k+l})\right] \tag{11}$$

Following Salas-La Cruz (1972), the covariance between stations k and $k+l$ is calculated as

$$\mathrm{Cov}(S_k, S_{k+l}) = \sigma^2 \rho_{k,k+l} \tag{12}$$

where $\rho_{k,\ k+l}$ is the cross-correlation coefficient between stations k and $k+l$. Therefore, the variance of $\bar{S}_m$ becomes

$$\mathrm{Var}(\bar{S}_m) = \frac{1}{m^2}\left[m\sigma^2 + 2\sum_{k=1}^{m-1}\sum_{l=1}^{m-k} \sigma^2 \rho_{k,k+l}\right] = \frac{\sigma^2}{m}[1 + (m-1)\bar{\rho}_{xx}) \tag{13}$$

where

$$\bar{\rho}_{xx} = \frac{2\sum_{k=1}^{m-1}\sum_{l=1}^{m-k} \rho_{k,k+l}}{m(m-1)},$$

the average cross-correlation coefficient for the region. For correlated flow sequences one could employ the test statistic

$$Z_m = \bar{S}_m / \sqrt{\mathrm{Var}(\bar{S}_m)} \tag{14}$$

with Var($\bar{S}_m$) given in Eq. (13). We compare this analytical approach for accounting for

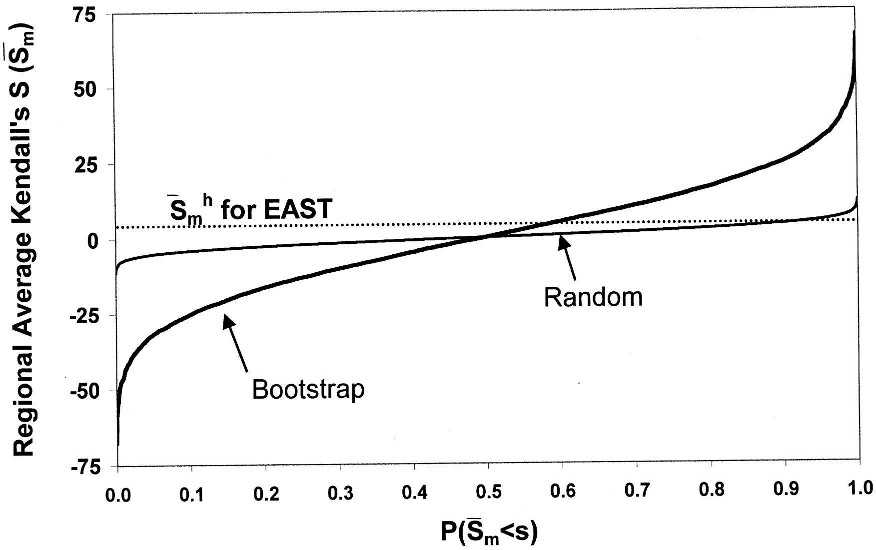

Fig. 2. Example of cdf of $\bar{S}_m$ for flood flows in East region at 30-year time frame (1959–1988). Bootstrap cdf represents spatially correlated data and random cdf represents spatially independent data.

cross-correlation described above with our empirical bootstrap methodology in Section 3.4.

2.3. Field significance

Field significance (denoted as α) is the collective significance of a group of hypothesis tests and may be defined in a number of different ways. Suppose we have a collection of N hypothesis tests each with computed "local" significance level p_i, $i = 1, \ldots, N$. If each test is independent, the local significance levels of the individual tests follow a uniform distribution so that the uniform cdf may be used to assess the overall "field" significance associated with the collection of N independent tests. This approach was suggested by Vogel and Kroll (1989). Another approach is to define X equal to the number of times we accept the null hypothesis using N individual $p\%$ level hypothesis tests. Then X follows a binomial probability distribution with parameters N and p. Livezy and Chen (1983) showed that, if $p = 5\%$, X must be *greater* than 5% of the total number of hypothesis tests in order to make inferences with 95% statistical confidence, even when N is large.

When cross-correlation between flow series exists, the hypothesis tests are no longer independent and the binomial distribution no longer describes the probability distribution of X, therefore, a simulation technique must be performed to assess overall field significance (α). Studies of trends in climatological data in which field significance has been assessed in this manner include Lettenmaier et al. (1994), Wilks (1996), Shabbar et al. (1997), Chu and Wang (1997) and Suppiah and Hennessy (1998). Of these studies, Lettenmaier et al. (1994) dealt with streamflow series.

2.4. The bootstrap

Bootstrapping entails the resampling of a data set B times, with replacement, to generate B bootstrap samples. The bootstrap samples are then used to approximate the statistical properties of a parameter of interest. The bootstrap method is outlined in Efron (1979). The null hypothesis for our trend tests was that the flow data exhibit no trends, are spatially correlated and serially independent. We used the bootstrap method to develop an empirical cdf for $\bar{S}_m$ in each region which conformed to our null hypothesis that no trends exist in the streamflow series. This empirical cdf was then used to determine the field significance associated with $\bar{S}_m$ computed from the historical data for that region. Our bootstrap method required that the

Table 1
Average regional lag-1 autocorrelation and number of stations with statistically significant autocorrelation coefficients for flood flow and low flow data

Region	Number of stations	Regional average r_1	95% Field significance[a]	Number of significant r_k[b]		
				Lag-1	Lag-2	Lag-3
Flood flow data						
East						
1959–1988	315	0.023	**22**	10	17	17
1939–1988	189	0.041	**15**	9	13	13
Midwest						
1959–1988	268	−0.022	**20**	5	9	6
1939–1988	120	0.049	**11**	7	3	9
West						
1959–1988	219	−0.056	**17**	9	2	5
1939–1988	107	0.008	**10**	8	6	8
Low flow data						
East						
1958–1987	310	0.199	**22**	68	28	19
1938–1987	162	0.185	**13**	56	24	11
Midwest						
1958–1987	263	0.244	**20**	72	31	23
1938–1987	105	0.319	**10**	60	22	13
West						
1958–1987	211	0.223	**16**	54	26	15
1938–1987	107	0.269	**10**	45	21	20

[a] Estimated number of independent 95% significance tests passed that will be equaled or exceeded by chance 5% of the time (following Livezy and Chen, 1983).

[b] Significance test for each autocorrelation, r_k, coefficient based on 95% confidence limits $= \pm 1.96(1/\sqrt{(n)})$ where $n = 30$ for 1958–1987 and $n = 50$ for 1938–1987. Therefore, r_k is significant if >0.358 for 30-year and if > 0.277 for 50-year timeframes.

flow series at each station be continuous over the timeframe of each test, therefore, the number of stations suitable for this investigation was further reduced. Roughly 45–75% of stations had at least 30 years of continuous data but only 21–35% of the stations had 50 years of continuous data. This was also a limitation of the trend analyses performed by Lins and Slack (1999).

The experiment proceeded as follows. The value of Kendall's S was calculated for each bootstrap sample and from this, the regional average test statistic, $\bar{S}_m$, was computed. This procedure was repeated 10,000 times. The cdf of $\bar{S}_m$ for the region was obtained by ranking the 10,000 values of $\bar{S}_m$ in ascending order and assigning a non-exceedence probability using the Weibull plotting position formula

$$P(\bar{S}_m \leq s) = \frac{r}{B+1} \quad (15)$$

where r is the rank and $B = 10,000$. The historical mean, $\bar{S}_m^{\mathrm{h}}$, was calculated for the same region using the historical data, rather than the bootstrap samples, and its field significance was assessed by comparing it with the empirical cdf of $\bar{S}_m$. This method is similar to that followed by Vogel and Kroll (1989) except they used synthetic datasets. Fig. 2 shows an example of the bootstrap cdf and the historical value of $\bar{S}_m^{\mathrm{h}}$ for flood flows in the East over the 30-year timeframe (1959–1988). Also plotted is the cdf for $\bar{S}_m$ assuming spatial independence between stations, developed by generating iid random flows for each station within the region. The difference in the values of field significance, α, between the spatially dependent ($\alpha = 0.41$) and spatially independent ($\alpha = 0.09$) cases is quite striking and illustrates how liberal the hypothesis tests become when spatial correlation is ignored.

Unlike flood flows, the low flow data were found to

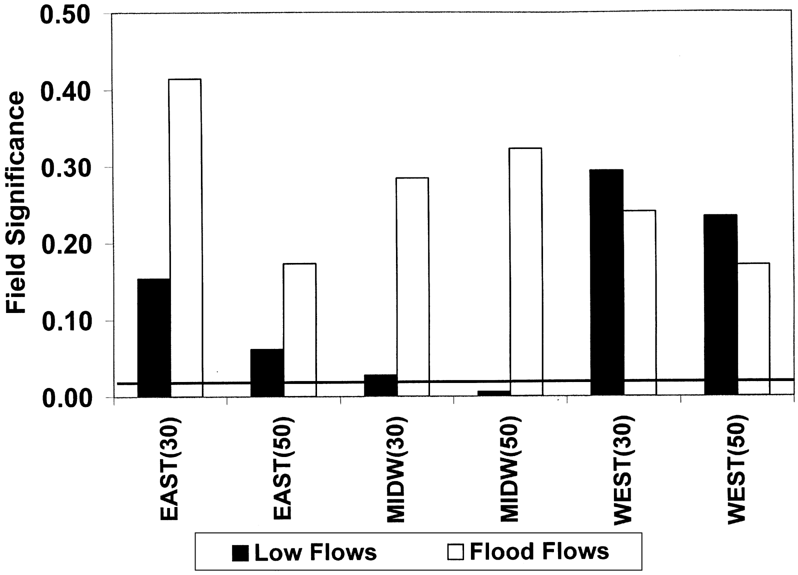

Fig. 3. Bootstrap field significance of trend tests for three geographical regions. Horizontal line represents a field significance of 0.025.

possess weak serial correlation, which implies a violation of our null hypothesis of serial independence. Table 1 shows the number of stations with statistically significant at-site serial correlation (autocorrelation) coefficients for the first three lags. The lag-1 autocorrelation coefficients (r_1) were the most significant relative to the number that would be expected due to chance (assuming independent data). The average lag-1 autocorrelation coefficients for each region are also presented in Table 1. Following the recommendation in von Storch (1995), the lag-1 autocorrelation was removed by pre-whitening the data. Pre-whitening was accomplished by assuming that the low flow data were generated by an AR(1)

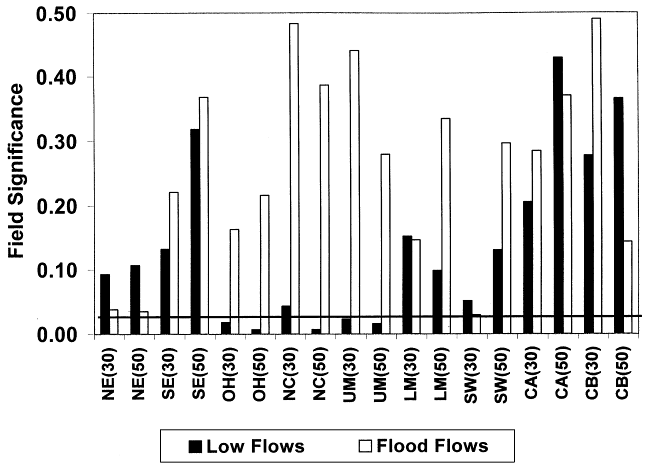

Fig. 4. Bootstrap field significance of trend tests for nine superregions. Horizontal line represents field significance of 0.025.

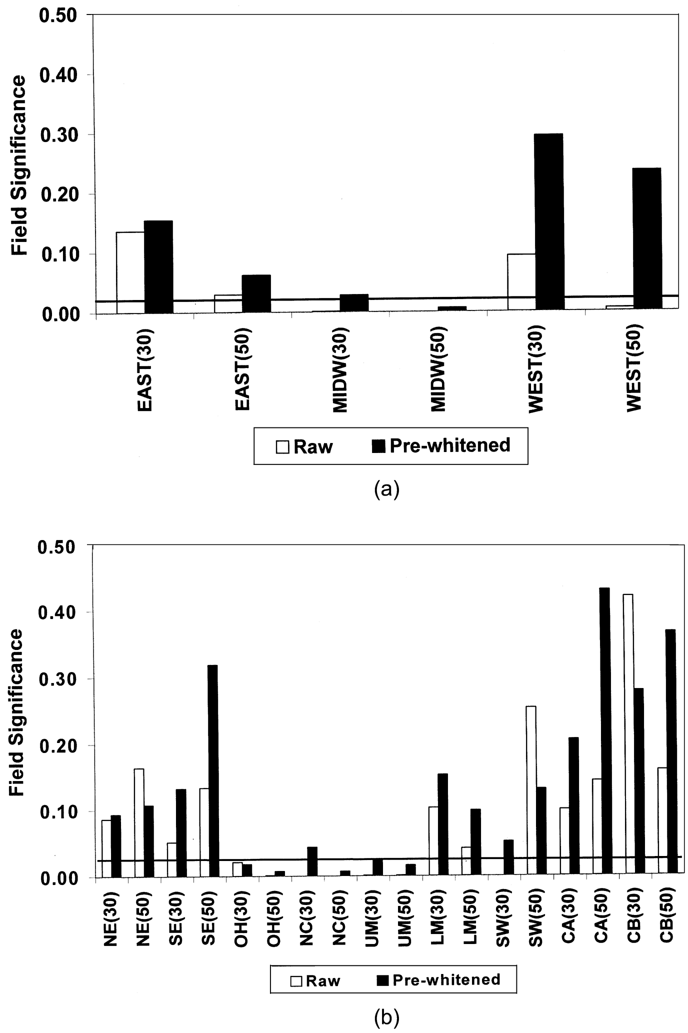

Fig. 5. Comparison of bootstrap field significance of trend tests for raw and pre-whitened low flow data for: (a) three geographic regions, and (b) nine superregions. Horizontal line represents a field significance level of 0.025.

process and then correcting the data as follows

$$Y_t = X_t - r_1 X_{t-1} \tag{16}$$

where X_t is the raw low flow time series. Pre-whitening reduced r_1 to near zero. The trend analyses were then performed on the original (raw) data and on the pre-whitened low flow data (Y_t) as described above. All interpretations of trend results are based on 95% statistical significance level tests.

3. Results and discussion

3.1. Trends in streamflow

Figs. 3 and 4 graphically illustrate the results of the bootstrap trend tests for flood and low flows in the three geographical regions and in the nine superregions, respectively. No significant trends were found in the flood flow data at either spatial scale, however, trends in the low flow data were observed at both scales, even after accounting for the spatial and serial

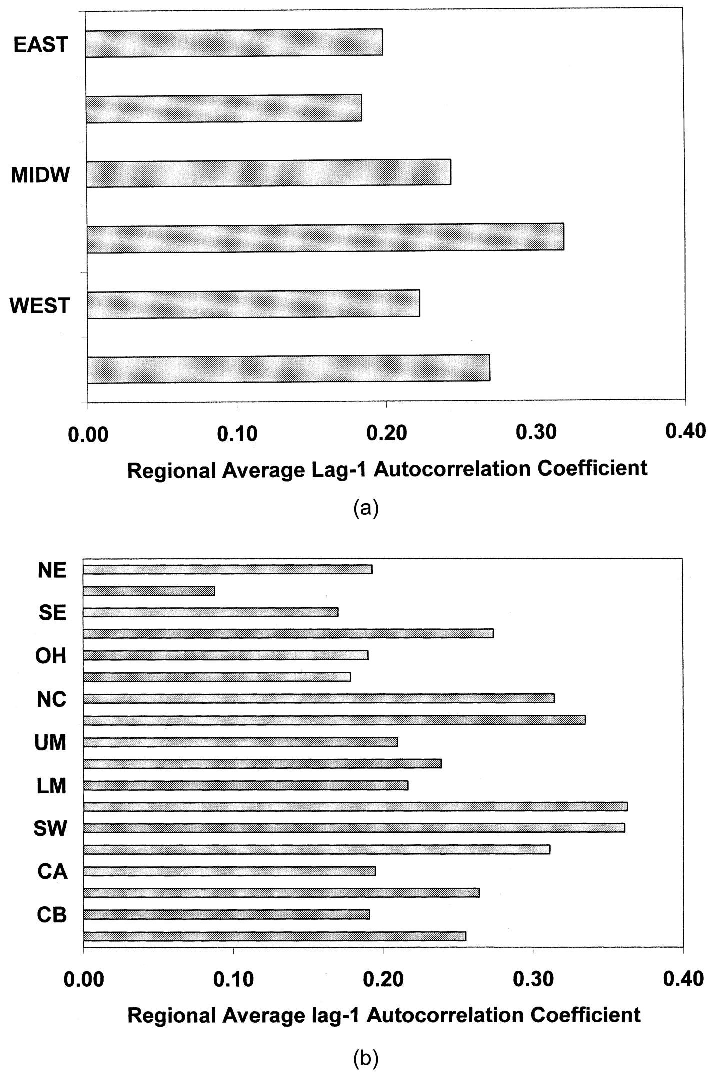

Fig. 6. Lag-1 serial correlation coefficients for the: (a) three geographical regions, and (b) the nine superregions. For each region, the upper bar is for the 30-year timeframe and lower bar is for the 50-year timeframe.

correlation of the flow series. At the larger spatial scale, significant upward trends were observed in the low flow data in the Midwest region over the 50-year timeframe. At the smaller spatial scale, significant trends in low flows were found in the Ohio (OH), and Upper Mississippi (UM) regions at both timescales and in the North Central (NC) region at the 50-year timescale. These results are similar to those of Lins and Slack (1999) who found a general lack of trends in annual maximum streamflow across the US, but showed upward trends in annual minimum streamflow in the UM and Great Lakes regions of the US. Interestingly, the greatest magnitude trends in monthly streamflow found by Lettenmaier et al. (1994) were also in the North-Central (NC) region.

Fig. 5 compares the field significance for the raw low flow data and the pre-whitened data. Without pre-whitening (white), more of the regions would have

Table 2
Comparison of field significance for spatially correlated (bootstrap) and independent cases: three geographical regions and nine superregions (shaded cells denote statistically significant trend at 95% level)

Region	Flood flows		Field significance		Low flows		Field significance	
	Stations	$\bar{S}_m$ [a]	Bootstrap[b]	Independent[c]	Stations	$\bar{S}_m$ [a]	Bootstrap[b]	Independent[c]
Three geographical regions								
East								
30-year	315	4.3	0.414	0.088	310	27.5	0.154	0.000
50-year	189	38.8	0.173	0.000	162	89.5	0.062	0.000
Midwest								
30-year	268	10.7	0.284	0.001	264	46.7	0.028	0.000
50-year	120	19.8	0.322	0.035	105	158.3	0.006	0.000
West								
30-year	219	20.5	0.240	0.000	211	14.8	0.293	0.000
50-year	114	56.2	0.170	0.000	107	42.0	0.233	0.000
Nine superregions								
30-year								
NE	113	52.1	0.038	0.0000	113	47.1	0.093	0.0000
SE	118	−19.2	0.221	0.0001	113	−34.2	0.132	0.0000
OH	69	−23.7	0.163	0.0002	69	74.6	0.018	0.0000
NC	124	1.2	0.483	0.4063	124	57.5	0.043	0.0000
UM	77	3.9	0.440	0.2716	74	48.7	0.023	0.0000
LM	82	21.8	0.146	0.0002	81	32.0	0.152	0.0000
SW	39	63.1	0.029	0.0000	38	52.1	0.051	0.0000
CA	93	22.9	0.284	0.0000	87	30.0	0.205	0.0000
CB	87	−1.1	0.490	0.4264	86	−17.0	0.277	0.0024
50-year								
NE	73	116.5	0.035	0.0000	64	87.5	0.107	0.0000
SE	77	18.8	0.368	0.0839	62	33.2	0.318	0.0016
OH	36	−46.2	0.216	0.0102	30	197.5	0.007	0.0000
NC	60	17.5	0.386	0.1280	56	190.6	0.007	0.0000
UM	29	35.6	0.279	0.0543	26	144.8	0.016	0.0000
LM	34	−19.6	0.334	0.1690	29	91.2	0.098	0.0000
SW	24	38.9	0.296	0.0556	22	78.2	0.130	0.0000
CA	35	30.0	0.370	0.0690	32	13.5	0.429	0.1467
CB	55	80.5	0.143	0.0000	53	23.0	0.365	0.0373

[a] Average Kendall's S calculated from historical data at m stations in each region.
[b] Field significance based on empirical cdf developed by bootstrap method.
[c] field significance based on significance level of Z_m (Eq. (10)) assuming iid data.

been interpreted as having statistically significant trends. For instance, at the larger spatial scale, significant upward trends were indicated in the West at the 50-year timeframe and in the Midwest at both timeframes. After pre-whitening (black), significant trends are indicated in the Midwest at the 50-year timeframe only. At the smaller scale, significant trends were indicated in two regions (SW and NC at the 30-year timeframe) when the raw data was analyzed but not after the data were pre-whitened. In most cases (but not all, as can be seen for the NE, SW and CB regions at the 50-year timeframe), serial correlation in the streamflow records tended to inflate the results of the trend test, making it appear as though there were more statistically significant trends than were actually present.

Von Storch (1995) warned that if r_1 is too large or the time series too short, pre-whitening the data in the manner described above affects any trend that may be present in the data. Values of r_1 used in pre-whitening

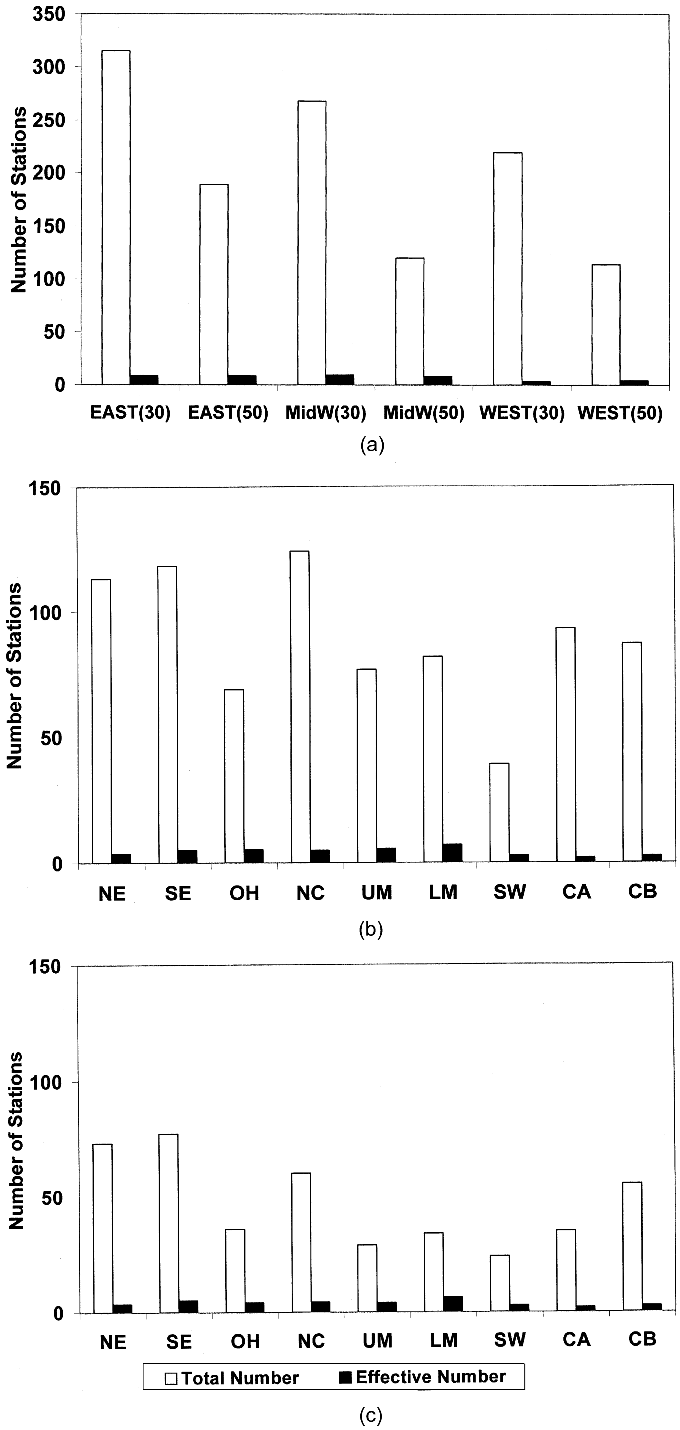

Fig. 7. Comparison of total number (white) and effective number (black) of stations for: (a) three geographical regions, (b) nine superregions over 30-year timeframe, and (c) nine superregions over 50-year timeframe.

the low flow data are presented in Fig. 6. The highest values are shown to be in the NC, LM and SW regions. Pre-whitening may have removed some of the trends as well as the persistence, so by presenting the trend results before and after pre-whitening, we have essentially bracketed the regions which have significant trends in the low flow data. The presence of upward trends in low flows in the midwestern US is supported by both sets of results.

3.2. The effect of cross-correlation on the interpretation of trend tests

Perhaps the most striking results of this study are illustrated in Table 2, which compares the field significance levels assessed under the case of spatial dependence between stations (the bootstrap analysis) and under the case of spatial independence (by calculating the significance levels of Z_m from Eq. (10)). It is clear that many more of the trend tests would have been considered statistically significant if cross-correlation was ignored. All trends in low flows and all but two trend in flood flows would have been significant at the larger spatial scale and all but two of the trends in low flows would have been considered significant at the smaller spatial scale. Also at the smaller spatial scale, half of the regions would have had evidence of statistically significant trends in flood flows. Correcting for tied values in the data increased the local significance (p-values) at some stations, but did not change the interpretation. These results illustrate quite dramatically how liberal the hypothesis tests become when one ignores the contribution of spatial correlation. Furthermore these results emphasize how misleading most previous trend detection analyses have been which ignored the important issue of spatial correlation.

It is generally understood that correlation, both spatial and serial, reduces the effective size of a sample used for hypothesis testing. Matalas and Langbein (1962) defined the relative information content of the mean as the ratio of the variance of the mean of a data set, assuming random (independent) observations, to the variance of the mean of the non-random data set. We computed the relative information content associated with $\bar{S}_m$ to quantify the reduction in the effective number of sites caused by the cross-correlation in the flood flow data. We calculate the relative information content, I, as

$$I = \frac{\mathrm{Var}(\bar{S}_m^{\mathrm{iid}})}{\mathrm{Var}(\bar{S}_m^{\mathrm{Bootstrap}})} \tag{17}$$

where $\mathrm{Var}(\bar{S}_m^{\mathrm{iid}})$ is the variance for the independent case (σ^2/m as given in Eq. (9)) and $\mathrm{Var}(\bar{S}_m^{\mathrm{Bootstrap}})$ is the variance of $\bar{S}_m$ accounting for the spatial correlation structure of the data in each region. The effective number of sites in each region, M', can be calculated as

$$M' = MI \tag{18}$$

where M is the total number of sites and I is given in Eq. (17). Fig. 7 compares the total number of sites used in the analyses of trends in flood flows and the effective number of sites using this approach. This exercise dramatically illustrates the reduction in information caused by cross-correlation between sites, so much so that the relative information content for the regional means is equivalent to only two to nine sites, depending on the assumed spatial scale. This explains the drastic difference in interpretation of trend results between the independent and the dependent cases illustrated earlier in Table 2. $\mathrm{Var}(\bar{S}_m^{\mathrm{iid}})$ was not corrected for the presence of tied data; this correction would have reduced the effective number even further. The fact that the effective number is less than 10 sites, regardless of the original number of sites, also indicates that our results would not likely be very different had we been able to use more sites in each region. Matalas and Langbein (1962) noted that when the data are cross-correlated, there are severe limits on substituting density of sites with length of record.

3.3. Comparison of field significance estimates using empirical and analytical methods

For comparison with the bootstrap methodology which preserves the empirical cross-correlation structure of the data, this section evaluates the analytical approach described earlier in Eqs. (11)–(14). This method assumes that the data are identically distributed and that the test statistic, Z_m, is normally distributed (Eq. (10)). Vogel and Wilson (1996) demonstrate that time series of annual minimum and time series of annual maximum streamflow are approximately identically distributed throughout the US. It should be noted, however, that since the

Table 3
Comparison of variance and field significance estimated with average regional cross-correlation coefficient (analytical method) with results of bootstrap method for three geographical regions and nine superregions

Region	Stations	ρ_{xx}[a]	Variance		Field significance	
			Bootstrap[b]	Analytical[c]	Bootstrap[b]	Analytical[c]
Three geographical regions						
East						
30-year	315	0.212	673.9	369.9	0.414	0.434
50-year	189	0.212	3089.4	1682.4	0.173	0.243
Midwest						
30-year	268	0.177	565.7	341.8	0.284	0.326
50-year	120	0.177	2627.6	1754.5	0.322	0.350
West						
30-year	219	0.420	1327.8	891.5	0.240	0.287
50-year	114	0.420	6075.2	3459.7	0.170	0.235
Nine superregions						
30-year						
NE	113	0.264	849.9	875.2	0.038	0.037
SE	118	0.163	534.4	616.4	0.221	0.203
OH	69	0.198	658.6	588.6	0.163	0.178
NC	124	0.218	704.7	627.5	0.483	0.482
UM	77	0.130	443.9	561.8	0.440	0.427
LM	82	0.167	556.6	441.2	0.146	0.178
SW	39	0.418	1360.1	1095.9	0.029	0.044
CA	93	0.460	1463.4	1604.8	0.284	0.275
CB	87	0.380	1216.2	1172.8	0.490	0.487
50-year						
NE	73	0.264	3917.1	4079.6	0.035	0.031
SE	77	0.163	2484.9	2847.6	0.368	0.353
OH	36	0.198	3148.1	3523.4	0.216	0.205
NC	60	0.218	3301.9	3317.7	0.386	0.380
UM	29	0.130	2286.7	3562.0	0.279	0.228
LM	34	0.167	2736.9	2276.5	0.334	0.354
SW	24	0.418	6320.5	5237.0	0.296	0.312
CA	35	0.460	6794.7	7757.7	0.370	0.358
CB	55	0.380	5591.9	5531.9	0.143	0.141

[a] Average regional cross-correlation coefficient (from Walker, 1999).
[b] Variance and field significance calculated from the bootstrap cdf.
[c] Variance and field significance calculated using analytical method (see Eqs. (13) and (14)).

effective number of sites in all regions was less than 10, the normality assumption for the test statistic may not be valid. Table 3 compares the variance and field significance of flood flows estimated from the bootstrap analysis with the analytically derived Var($\bar{S}_m$) (Eq. (13)) and field significance (significance levels of Z_m in Eq. (14)). Var($\bar{S}_m$) was calculated using regional average cross-correlation coefficients ($\bar{\rho}_{xx}$) reported by Walker (1999). At the smaller scale (nine superregions), the estimates of variance and field significance from the two methods were quite close in most cases. The difference in estimates at the larger scale was likely due to the fact that $\bar{\rho}_{xx}$ for each of the larger regions was the average of smaller regional averages. Such averaging may have resulted in a loss of information content contained in the smaller regional averages.

The analytical method is much simpler to implement and, as long as the region is not too large, it appears to yield reasonably accurate results. This method holds promise as a useful hypothesis test for incorporating spatial correlation in trend analysis and may be more suitable for scales ranging from field site to drainage basin. Further investigation is needed to determine the scale limitations and the robustness of the analytical method.

3.4. Implications of trends in streamflow observed in this study

The upward trends in precipitation found by Lettenmaier et al. (1994) occurred in roughly the same vicinity as the upward trends in annual minimum streamflow found by Lins and Slack (1999). This area of concentrated upward trends is enclosed within the regions found in our study to have significant upward trends in low flows. Qualitatively, there appears to be some correlation between the locations of upward trends in low flows and upward trends in annual precipitation. The relationship between streamflow and precipitation is more quantitatively illustrated by the result of two recent studies. Sankarasubramanian et al. (2001) found that the regional sensitivity of streamflow to changes in precipitation is highest in the Midwestern regions, indicating that the observed trends in low flows in our study and in the Lins and Slack study could have been produced by the observed changes in precipitation.

One question that begs to be answered is: why are upward trends observed in low flows but not in flood flows? Perhaps trends in low flows (which represent ground water outflow to rivers and streams) reflect an increase in basin storage resulting from the observed increases in annual precipitation. Such changes in storage could be influenced by climate change for the following reason. Consider the water balance for a drainage basin

$$P + G_{\text{in}} - ET - G_{\text{out}} - Q = \Delta S/\Delta t \quad (19)$$

where P is precipitation, ET is evapotranspiration, G is groundwater runoff, Q is streamflow, and ΔS is the change in water storage within the basin over a time step, Δt. Under stationary conditions, the storage term ($\Delta S/\Delta t$) is assumed to equal zero at an annual timestep. However, if one or more of the random variables on the left-hand side of the equation is non-stationary, it would cause other terms in the water balance to also be non-stationary. For instance, if the precipitation term (P) is increasing over a timeframe greater than the annual time step, this could affect either the streamflow term (Q) causing it to also increase over time or the storage term (ΔS) making it greater than zero for an annual time step. Our study investigated whether the former has occurred. Our results indicate that it could also be the latter. Climate change may actually be impacting the storage term by increasing the volume of water held in ground water storage (which supplies baseflow to streams) or as soil moisture within a basin. An increase in soil moisture is a less likely storage mechanism due to the fact that increases in air temperature, which would tend to dampen such a change, have also been observed over the last half-century (Vinnikov et al., 1990; Bloomfield, 1992; Lettenmaier et al., 1994). However, increasing ground water elevations in Illinois between the late 1950s and the mid-1970s were observed by Smith and Richman (1993) as well as statistically significant increases in mean annual streamflow in some areas. They further note that little is known about the response of shallow water tables to long-term climate shifts. If indeed climate change is resulting in an increase in basin storage, it is possible that storage capacity will eventually be exceeded and at that point, trends in the upper streamflow quantiles may be observed. Other explanations for the lack of response in flood flows to observed increases in precipitation may be that: (1) the timing of increased flow does not correspond to timing of flood events, (2) the magnitude of the increases is not large enough to affect the magnitude of floods, or (3) the magnitude of the increases is still negligible compared to the natural variability of flood flows. Lettenmaier et al. (1994) reported that the number and location of upward trends in average monthly streamflow varied by season. Lins and Slack (1999) reported that the number of stations with significant trends was approximately equal for the annual minimum through the annual median streamflow quantiles. More investigation into the links between the timing and magnitude of trends in precipitation and streamflow needs to occur before these issues can be more definitively evaluated.

4. Conclusions

Regional trends in flood flows and low flows across the US were evaluated using the regional average

Kendall's S ($\bar{S}_m$) at two spatial scales (three geographic regions and nine smaller regions) and over two timeframes (the most recent 30 years and the most recent 50 years). Empirical cumulative density functions (cdfs) for $\bar{S}_m$ were obtained using a bootstrap method which preserves the observed spatial covariance structure of the streamflows. Field significance was assessed by comparing the average regional S calculated from the historical data in each region and timeframe with the corresponding empirical cdf of $\bar{S}_m$. Trends were considered statistically significant if the absolute value of field significance was less than 0.025.

No evidence of statistically significant trends was found in the flood flows at either scale. Statistically significant upward trends in low flows were found at the larger scale in the Midwest and in three of the smaller regions: the Ohio (OH), the north-central (NC) and the upper Midwest (UM) regions. The presence of trends in low flows in the midwestern US may be attributed to observed increases in precipitation based on Sankarasubramanian et al. (2001) who found that of all regions of the US, streamflow was most sensitive to precipitation in the Midwest.

A dramatically different interpretation of these trend analyses would have been achieved if the spatial correlation of the flow series within the regions had been ignored. In that case, these analyses would have yielded statistically significant trends in all but two of the low flow analyses and in approximately half of the flood flow analyses. An evaluation of the effective number of sites within each region showed how the presence of cross-correlation severely limits the amount of information that can be gleaned from hypothesis tests performed at many sites within a region. We also found that low flow time series exhibit significant serial correlation. Even when the lag-1 autocorrelation was removed from the time series, significant trends in low flow series were apparent, though the number of significant trends decreased.

Acknowledgements

Although the research described in this article has been funded in part by the United States Environmental Protection Agency (EPA) through STAR grant number R824992-01-0 to Tufts University and EPA grant number R825888 to the SUNY College of Environmental Science and Forestry, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no endorsement should be inferred.

References

Bloomfield, P., 1992. Trends in global temperatures. Clim. Change 21, 1–16.

Changnon, S.A., Kunkel, K.E., 1995. Climate-related fluctuation in Midwestern floods during 1921–1985. J. Water Resour. Plann. Mgmt 121 (4), 326–334.

Chu, P.-S., Wang, J.-B., 1997. Recent climate change in the tropical western Pacific and Indian Ocean regions as detected by outgoing longwave radiation. J. Clim. 10, 636–646.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7 (1), 1–26.

Groisman, P.Y., Easterling, D.R., 1994. Variability and trends in total precipitation and snowfall over the United States and Canada. J. Clim. 7, 184–205.

Guttman, N.B., Wallis, J.R., Hosking, J.R.M., 1992. Regional temporal trends of precipitation quantiles in the US. Research Report RC 18453 (80774), IBM Research Division, Yorktown Heights, NY.

Hubbard, E.F., Landwehr, J.M., Barker, A.R., 1997. Temporal variability in the hydrologic regimes of the United States. In: Gustard, A. (Ed.), FRIEND'97 — Regional Hydrology: Concepts and Models for Sustainable water Resource Management. IAHS Publication, vol. 246. IAHS, Oxfordshire, UK.

Karl, T.R., Knight, R.W., 1998. Secular trends of precipitation amount, frequency, and intensity in the United States. Bull. Am. Meteorol. Soc. 79 (2), 231–241.

Kendall, M.G., 1962. Rank Correlation Methods. 3rd ed. Hafner Publishing Company, New York.

Lettenmaier, D.P., Wood, E.F., Wallis, J.R., 1994. Hydro-climatological trends in the continental United States, 1948–88. J. Clim. 7, 586–607.

Lins, H.F., Slack, J.R., 1999. Streamflow trends in the United States. Geophys. Res. Lett. 26, 227–230.

Livezy, R.E., Chen, W.Y., 1983. Statistical field significance and its determination by Monte Carlo techniques. Mon. Weather Rev. 111, 46–59.

Matalas, N.C., Langbein, W.B., 1962. Information content of the mean. J. Geophys. Res. 67 (9), 3441–3448.

Olsen, J.R., Lambert, J.H., Haimes, Y.Y., 1998. Risk of extreme event under nonstationary conditions. Risk Anal. 18 (4), 497–510.

Olsen, J.R., Stedinger, J.R., Matalas, N.C., Stakhiv, E.Z., 1999. Climate variability and flood frequency estimation for the upper Mississippi and lower Missouri rivers. J. Am. Water Resour. Assoc. 35 (6), 1509–1520.

Pupacko, A., 1993. Variations in northern Sierra Nevada streamflow: implications of climate change. Water Res. Bull. 29 (2), 283–290.

Salas-La Cruz, J.D., 1972. Information content of the regional mean. Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resource Systems, vol. 2, University of Arizona, Tucson, December 11–14.

Sankarasubramanian, A., Vogel, R.M., Limbrunner, J.F., 2001. The climate elasticity of streamflow in the United States. Water Resour. Res. (in press).

Shabbar, A., Bonsal, B., Khandekar, M., 1997. Canadian precipitation patterns associated with the southern oscillation. J. Clim. 10, 3016–3027.

Slack, J.R., Lumb, A.M., Landwehr, J.M., 1993. Hydro-Climatic Data Network (HCDN). , Water-Resource Investigations Report 93-4076. US Geological Survey, Washington, DC.

Smith, K., Richman, M.B., 1993. Recent hydroclimatic fluctuations and their effects on water resources in Illinois. Clim. Change 24, 249–269.

Suppiah, R., Hennessy, K., 1998. Trends in total rainfall, heavy rain events and number of dry days in Australia, 1910–1990. Int. J. Climatol. 10, 1141–1164.

Trenberth, K.E., 1999. The extreme weather events of 1997 and 1998. Consequences 5 (1), 3–15.

Vinnikov, K.Y., Groisman, P.Y., Lugina, K.M., 1990. Empirical data on contemporary global climate changes (temperature and precipitation). J. Clim. 3, 662–677.

Vogel, R.M., Kroll, C.N., 1989. Low-flow frequency analysis using probability plot correlation coefficients. J. Water Res. Plann. Mgmt 115 (3), 338–357.

Vogel, R.M., Wilson, I., 1996. Probability distribution of annual maximum, mean, and minimum streamflows in the United States. J. Hydrologic Engng 1 (2), 69–76.

Von Storch, H., 1995. Misuses of statistical analysis in climate research. In: von Storch, H., Navarra, A. (Eds.), Analysis of Climate Variability: Applications of Statistical Techniques. Springer, Berlin.

Walker Jr., F.R., 1999. Long term variability in flood arrival rates: the presence of annual flood event clustering. PhD dissertation, Cornell University, Ithaca, NY.

Wilks, D.S., 1996. Statistical significance of long-range Optimal Climate Normal temperature and precipitation forecasts. J. Clim. 9, 827–839.