Detecting climate-related trends in streamflow data

Paul J. Pilon and Sheng Yue
MSC–Ontario, Environment Canada, CCIW, 867 Lakeshore Road, P.O. Box 5050, Burlington, ON, L7R 4A6, Canada

Abstract This paper reviews the results of a number of studies that have investigated streamflow data for the existence of trend. These studies provide evidence that trends in various, but not all, streamflow regimes are occurring at rates that are higher than one might attribute to chance alone. Results of different studies using different approaches were compared and were shown, at times, to have dramatic differences. These differences might potentially be due to pre-conditioning of data prior to trend detection in attempts to minimize the impacts of serial correlation on testing procedures. It was also evident that patterns of trend can vary over small spatial scales and that a relatively high-density network is required to effectively comprehend trend and how it might be altering across an area. A global network of streamflow sites representing pristine or stable conditions is needed to assess patterns of change. Selection criteria for sites within such a network are provided, and it is highlighted that local knowledge is required to perform this selection.

Keywords Hydromagnification; Mann-Kendall test; networks; prewhitening; trend detection; streamflow

Introduction
There is growing concern among many nations of the world regarding the future availability of freshwater resources and, in particular, about its potential scarcity and its quality. Between 1900 and 1995, water withdrawals have increased by an estimated factor of over six, more than double the rate of population growth (Shiklomanov, 1997). This rapid growth in water demand is due primarily to increased use of water for irrigation as well as the unquenchable thirst for water domestically and by industry. Of the various sources of freshwater, it is estimated that more than 90% of the world’s water supply emanates from streamflow and shallow groundwater reserves (Shiklomanov, 1997), although “no authoritative estimates exist of the percentage of world water use that depends on groundwater” (UNEP, 1996). Knowledge of the spatial and temporal characteristics of these resources and their use are paramount for effective decision making. Such knowledge has proven to be rather elusive as noted by Shiklomanov (1997) who found “extreme difficulty in preparing a global assessment because of the lack of sufficient and reliable information on water availability, quality and use in many areas of the world.” He focused much of his attention on streamflow data due to their availability and their inherent inter-linkages with shallow groundwater reserves. The availability and distribution of streamflow as a source of freshwater is of profound importance to humankind and our natural environment.

To further complicate these matters, there is also growing concern about the state of the current climate system and its implications for water availability, which is compounded by the paucity of suitable historical streamflow observations upon which to base analyses. These concerns have stemmed to a large degree from results of climate model simulations reflecting increased concentrations of CO₂. Gleick (1989) presented a review of a number of studies that indicated important shifts in various hydrological regimes such as decreasing soil moisture levels and runoff during the summer season, shifts in the source of waters and their magnitudes occurring in various months of the year, and large increases in winter streamflow. He viewed the results of these earlier works as “a first warning of possible important changes in regional water availability”. Such projection-oriented efforts using...
climate model simulations have commonly been interpreted to imply that changes to the climate system could lead to an “intensification” of the hydrological cycle with intensification possibly implying a worsening of extremal flows (Houghton et al., 1996). It has been suggested (Pilon et al., 1991; Burn, 1994; Chiew and McMahon, 1996) that certain responses of the hydrological cycle to climate induced change may be “hydro-magnified”.

Analyses of historical observational records provide information about possible climatological variability and change, and such analyses are also valuable in assessing the ability of climate circulation models to replicate current climate patterns. Of particular interest are the analyses of hydrological indices, given the potential for hydromagnification to occur and the importance of results on water resource management practices. Pilon and Kuylenstierna (2000) suggest a variety of candidate hydrological indices for climate change analyses and variability analyses that span discharge, spring season processes, natural groundwater, and glaciers and permafrost. Lawford (1992) provides the view that many such variables within the natural sciences could be of importance to such change studies. However, Burn (1994, p. 28) points out that many such variables may not have records available for a large number of locations nor have a sufficient length of record. Intuitively, there must be sufficient data and such data must also be of a suitable quality for them to be of benefit in studies of detection of change. He selects streamflow as the hydrological variable of choice for detecting the impacts of climate change as “it represents a basin integrated response to hydrological inputs and therefore affords good spatial coverage.”

The fundamental prerequisites for performing trend detection studies in hydrology are 1) the availability of reliable, systematic, and suitable data upon which to base the analyses and 2) the use of appropriate statistical methodologies upon which to provide evidence for consideration. Within this paper, a brief overview of previous work on detecting trends in historical streamflow data will be provided, due to the important socio-economic and environmental implications of such change. Prior to providing the overview of these previous efforts, criteria used for the selection of stations in change analyses will be reviewed, and an overview of a few commonly used statistical procedures will be given, along with some of their characteristics and capabilities. Following this, a discussion of results and recommendations for future work will be given.

**Selection of streamflow stations for detection studies**

A number of efforts in the creation of specific or “specialized” observational networks along with in-depth material on their selection criteria are provided by Pilon (2000). Various efforts have been made or are underway in establishing specialized networks for surface climate, hydrometric and water quality observational systems. Some of these networks are international in scope such as the Global Climate Observing System Surface Network (GSN) (Peterson et al., 1997) and the WMO Reference Climatological Stations (RCS) network (WMO, 1966, 1986), while others provide national oriented coverage (Slack and Landwehr, 1992; Alexander et al., 1998; Environment Canada, 1999; Harvey et al., 1999).

There is potential, as well, to debate whether data should represent stable or pristine conditions as compared with unstable and highly anthropogenically influenced flows. Refsgaard (1987) suggests, for example, that it is possible to remove the human influence on the records through hydrological modelling. This approach, which results in “naturalized” streamflows, would greatly expand the spatial and temporal coverage of the sites within the network. Some climatologists concur with this view and have included “homogenized” temperature records, corrected for the “heat-island” effect, in networks for the detection of climate change. This has been the case with the GSN (Peterson et al., 1997)
and the US Historical Climatology Network (USHCN) (WMO, 1993). As reported in Peterson et al., up to 36% of selected temperature sites were in urban communities having a population in excess of 10,000. Some may question if such a network could credibly detect climate change from greenhouse gas forcing, given the potential for misinformation contained within the data from corrections applied for the removal of the heat-island effect.

Wallis et al. (1991) describe the construction of a combined climate and hydrometric data set for the continental United States. They noted that the effects on streamflow by human activities presented somewhat of a different problem than for climate stations. They argued that most gauged streams within the United States “are affected to some extent” by human activity. The most significant effects on natural flows would be upstream containment structures, diversions, and water loss through consumptive uses. They focused their efforts on identifying a subset of US hydrometric sites that would be free of water management effects and presumably would represent stable land-use conditions. In further efforts to construct a specialized streamflow network in the US, Slack and Landwehr (1992) report on a review of all US Geological Survey streamflow records for the entire country and its protectorates. They established a set of criteria for the identification of candidate stations and worked closely with the District offices, who had local knowledge of basin conditions and the quality of the data, to review each potential candidate. This resulted in the identification of 1,659 sites, which comprise the Hydro-Climatic Data Network (HCDN).

In the development of their selection criteria, Slack and Landwehr (1992, p. 2) argued that “the pattern of past climate variation to be discerned in the streamflow record would be confounded by changes induced by anthropogenic activity.” They adopted a criterion that the streamflow characteristics must be representative of the natural or stable conditions, and the flow of the site must be representative of natural conditions at least on a monthly basis. They indicated that upstream controls or diversions must not affect monthly averaged flows. They developed and applied the following six selection criteria.
1. Data must be available in electronic format.
2. Data for both active and discontinued sites must span an entire water year (i.e. no seasonal sites).
3. There must be at least twenty years of suitable data, except for data sparse areas.
4. The accuracy of the data had to have been assigned a value of at least “good”.
5. There must be unimpaired basin conditions affecting the average monthly discharge.
6. Data must be obtained by use of national standard procedures inferring use of only measured discharge values, thereby not allowing estimated or reconstructed records.

Environment Canada established what they refer to as a Reference Hydrometric Basin Network (RHBN) for Canada (Environment Canada, 1999; Harvey et al., 1999). This network is similar in concept to the WMO RCS network, but the RHBN is intended to provide hydrometric time series data for use in detection, monitoring, and assessment studies. Selection criteria for the establishment of the RHBN were similar to that of Slack and Landwehr (1992) with some minor variants. In the Canadian approach, only active hydrometric stations were considered as the data were intended, in part, for on-going monitoring of potential change. This need also necessitated a criterion on potential longevity of the site, both from financial as well as basin stability perspectives. Longevity in this instance refers to a high possibility that the site would continue to exist and be useable for analytical purposes for the foreseeable future. Seasonal and lake level sites were also considered. In the prairie region of Canada, several stations are operated on a seasonal basis due to climatological and geographical considerations. These sites are typically operated from late winter through to late fall and were felt valuable for potential use in analyses reflecting
change in various flow regimes during this period of the year, otherwise there would be poor spatial representation for certain indices from this portion of the country. Special attention was also placed on development of an in-depth accuracy index of the data and a criterion reflecting the degree of basin development. In essence, it was felt that data should reflect basins having less than 10% of the surface area whose land-use was modified in some fashion. In addition, sites reflecting in-stream control structures were permitted provided they controlled less than 5% of an area of a basin. The Canadian effort resulted in the identification of 7 lake level stations, 37 seasonal streamflow and 211 continuous streamflow stations.

From the review of the selection criteria for hydrometric networks, it is evident that previous efforts in hydrometry tended to choose similar criteria. There is, however, a distinction that must be made between hydrometry and efforts being made in selecting climatological stations for analyses. It is evident in the work of Vose et al. (1992), WMO (1993), Jones (1994), and GCOS (1998) that climate stations reflecting direct human contamination of data, such as the heat-island effect, are commonly included in climate detection networks. There are some national exceptions to this such as the Historical Canadian Climate Database (Gullett, 1991). Hydrologists, on the other hand, have paid special attention in ensuring that the data from hydrometric sites were as free as practicably possible from direct human interference on the landscape and streamflows. There have been some exceptions due to the difficulty in assessing land-use and the stability of basin conditions (Chiew and McMahon, 1996), which are typically assessed based on local expert knowledge, or where it was felt that intervention would not have been significant (Robson et al., 1998; García and Vargas, 1998; and Genta et al., 1998) for the variable of concern.

It should be noted, however, that in the literature reviewed here, all streamflow-oriented trend detection studies have attempted to not include sites where the timing and quantity of flow are directly modified such as to jeopardize the overall goal of the analysis, although at times this might be difficult to ascertain. Most groups have preferred to target streamflows of natural or stable basins, in an attempt to minimize the direct impacts of changing land-use, water withdrawals and management practices on data and to facilitate interpretation of results. The presumption is that it would be difficult to differentiate between hydrological effects due to direct human intervention in the basin and those due to anthropogenically induced climate change.

It is also evident that efforts to monitor global indices of our climatological system in a concerted fashion far outstrip any efforts placed on hydrological indices of importance such as streamflow. To date, most efforts focusing on hydrological indices have been national in scope, possibly due to the effort required to select appropriate sites for analyses and the need for local knowledge in the application of selection criteria. Hence, it is rather difficult to extrapolate regional and local results to global ends. It is important that efforts be made to identify specialized streamflow sites on a global basis, as described by Pilon and Kuylenstierna (2000), to facilitate future detection-based and attribution analyses. In essence, international efforts are required to establish a global network of suitable hydrometric sites. Selection criteria similar to that proposed by Slack and Landwehr (1992), Environment Canada (1999), and Pilon (2000) should be considered to guide the development of such a network. These criteria would be used to identify pristine and stable hydrometric stations reflecting: breadth of coverage (seasonal, continuous, streamflow and lake level); degree of basin development; no significant regulation or diversions; length of suitable record; longevity; and data of reasonable quality. The results of such efforts form the foundation for studying the effects on our water resources of regional and global change within the climatological system.
Common statistical procedures

Cavadias (1992) and Robson et al. (2000) provide an overview of several testing procedures for analyzing potential change in hydrological or other related series. Of all approaches, the Mann-Kendall test (Mann, 1945; Kendall, 1975) has been widely used in much of the trend detection work in hydrology, and it will be referred to repeatedly in the next section that reviews some relevant detection studies. The Mann-Kendall test statistic, \( S \), is defined as follows:

\[
S = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{sgn}(x_j - x_i)
\]

where the \( x_j \) are the sequential data values, \( n \) is the length of the data set, and

\[
\text{sgn}(\cdot) = \begin{cases} 
1 & \text{if } > 0 \\
0 & \text{if } = 0 \\
1 & \text{if } < 0
\end{cases}
\]

The statistic \( S \) is approximately normally distributed when \( n \geq 8 \), with the mean and the variance as follows:

\[
E(S) = 0
\]

\[
V(S) = \frac{n(n-1)(2n+5)}{18}
\]

The standardized test statistic \( Z \) is computed by

\[
Z = \frac{S - 1}{\sqrt{V(S)}}
\]

The standardized Mann-Kendall statistic \( Z \) follows the standard normal distribution with mean of zero and variance of one. There are, however, some concerns with the application and interpretation of results when the variable under study happens to be serially correlated (von Storch and Navarra, 1995), a characteristic that frequently exists in hydrology. In such instances “pre-whitening” of the series to remove serial correlation is advocated. The work of Yue et al. (submitted and in press) illustrate that this choice of analytical approach may lead at times to differing conclusions about potential trends in series. Some of the implications of such choices and the impacts of serial correlation and prewhitening procedures are dealt with in the next section of this paper.

Having computed the significance of each individual site, the question arises as to whether the entire group of sites for an area demonstrate statistical evidence of trend, more so than one would expect to see from chance for a similar number of sites. The assessment of group response is termed “field significance”. If there is no cross-correlation among sites, then the binomial distribution can be used to assess the field significance of trend. Livezey and Chen (1983) used the binomial distribution to compute the probability related to the number of stations with significant trend in a study area. The binomial distribution is

\[
P(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}
\]
where $P(k)$ is the probability of $k$ occurrences in $n$ trials, and $p$ is a certain probability associated with each occurrence. The probability of $k$ or more occurrences being an anomaly or accidental result is

$$P(\chi \geq k) = \sum_{i=0}^{k} P(k)$$  \hspace{1cm} (7)$$

Should the sites demonstrate significant serial correlation, Douglas et al. (2000) have put forth a bootstrapping approach for determining the field significance based on an average value of the Mann-Kendall statistic, computed over the entire area of study. The regional Mann-Kendall statistic, $S_m$, is computed as the mathematical average of all individual site $S_i$ values for a particular network. The bootstrapping procedure is adopted to resample the sites to obtain the empirical null distribution of $S_m$ that is employed to assess the field significance of trend. This approach preserves the cross-correlation of the network. Yue et al. (in press) provide an extension of the bootstrap procedure to evaluate the tails of the test statistic’s sampling distribution. This extension allows an evaluation of the significance of upward and downward trend similar to the binomial approach, but is based on the bootstrapping which also preserves cross-correlation.

Trend detection studies

Trend detection studies are of importance in ascertaining what might be happening within the hydrological system and in furthering our understanding of what future changes might befall us. Although much attention is focused on tracking and establishing the global mean temperature and various estimates under future atmospheric states and time periods, it is evident that there are patterns to the change that are much more regional in nature and extent. Several studies have been performed on past observations to investigate the possibility of change being evident within a more local or regional climate system, for example Lettenmaier et al. (1994), Karl et al. (1995 and 1996), Karl and Knight (1998), Zhai et al. (1999), Zhang et al. (2000) and Bonsal et al. (2001). Various time scales of precipitation and temperature data have now been analyzed showing that annual precipitation may well be increasing in the middle to high latitudes of the Northern Hemisphere. However patterns are not always consistent and can vary. This was evident, for example, in the work of Lettenmaier et al. (1994) wherein annual indices of mean temperature are showing increasing trends to the west of the continental US and decreasing trend to the eastern portion. Results also point to an increasing annual daily temperature range.

Higher temporal resolutions of important climatological indices have also been analyzed. Lettenmaier et al. (1994) show graphically the trends for each month of the year for precipitation, temperature and streamflow. From their results, it is evident that the scale of change is local to regional in nature, which is also supported by the results of several other groups. For example, for the months of January, the west coast of the US is showing a positive trend in temperature, while the eastern area just south of the Great Lakes is showing a decrease. To contrast this, the March temperature sites are showing a warming for the west through to a line from the Great Lakes to the north of Texas. The remaining portion of the continental US for this month is showing no change. Their graphical presentation of monthly precipitation also shows signs of clustering or patterns. Certain parts of the country may be experiencing more monthly precipitation in certain months and less precipitation in others. For example, the area south of the Great Lakes seems to be experiencing a decrease in January precipitation, yet the same area appears to be receiving an increase in precipitation in various other months (May, October and November).

The purpose of presenting a brief overview of some of the results of Lettenmaier et al. (1994) was to illustrate the highly variable nature of patterns in trend in climate indices. In
other words, change is not simple, and it does not occur the same everywhere. It also does not occur the same throughout the various seasons, even for one index. It is also evident that there are important interactions and feedbacks amongst the various indices, aspects that have long been recognized by hydrologists. For example, it is reported (Zhang et al., 2000) that changes in regional temperature trends in the fall and spring of the year seem to coincide with changes in the duration of river ice. Quantitative physical attribution of change within the hydrological cycle is an area requiring further attention by the scientific community.

Probably the most important aspect of a changing climate will be its impacts on water availability. Some studies, although fewer than on climatological series, have also been made to ascertain if changes to streamflow regimes have possibly been occurring (Lettenmaier et al., 1994; Lins and Michaels, 1994; Chiew and McMahon, 1996; García and Vargas, 1998; Genta et al., 1998; Robson et al., 1998; Lins and Slack, 1999; Douglas et al., 2000; Groisman et al., 2001; Zhang et al., 2001; and Yue et al., submitted and in press).

Lins and Michaels (1994), using a subset of 559 gauges from the HCDN of Slack and Landwehr (1992), analyzed monthly streamflow records from 1941 to 1988. They found that unimpaired streamflows within the United States have increased for nearly all of the 11 major hydrological regions of the country for the fall and winter period. They noted that there was one exception with flows increasing in the month of August in the New England area. These results were obtained using the nonparametric Mann-Kendall test. Lettenmaier et al. (1994) also performed an analysis on monthly streamflow for the continental US for the 1948 to 1988 period using the hydrometric network of 1009 sites that had been identified by Wallis et al. (1991). Lettenmaier et al. (1994) observed patterns of increasing trends starting in October through to April, with the largest number occurring in the winter months of January and February (see Table 1). They also provided maps displaying the results of significant trends for each month, where significance was established through use of the Mann-Kendall test. From these maps it was evident that from October through June there was a growing number of positive test results peaking in the two winter months of January and February, and that for this period, the upward trends were concentrated in the area to the south and east of the Great Lakes. Their analyses of monthly temperature and precipitation indicated increases in the temperature for March at 480 of 1,009 sites (47.6%) with increases in monthly precipitation for September through December of 199 (19.7%), 224 (22.2%), 280 (27.8%) and 114 (11.3%), respectively. Relating these results through knowledge of hydrology proved difficult, as they felt that “the observed trends in streamflow [were] not entirely consistent with the changes in the climatic variables and may [have been] due to a combination of climatic and water management effects.”

Lettenmaier et al. (1994) also analyzed a number of other indices including monthly precipitation, monthly temperature, and monthly temperature ranges, annual temperature,

Table 1  Number of significant \((p \leq 0.02)\) streamflow trends by month for 1009 hydrometric stations in continental US for the 1944 to 1988 period (adapted from Lettenmaier et al. (1994))

<table>
<thead>
<tr>
<th>Month</th>
<th>Significant Trend</th>
<th>Month</th>
<th>Significant Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upward (%)</td>
<td>Downward (%)</td>
<td>Upward (%)</td>
</tr>
<tr>
<td>January</td>
<td>440 (43.6)</td>
<td>86 (8.5)</td>
<td>July</td>
</tr>
<tr>
<td>February</td>
<td>422 (41.8)</td>
<td>33 (3.3)</td>
<td>August</td>
</tr>
<tr>
<td>March</td>
<td>349 (34.6)</td>
<td>19 (1.9)</td>
<td>September</td>
</tr>
<tr>
<td>April</td>
<td>197 (19.5)</td>
<td>68 (6.7)</td>
<td>October</td>
</tr>
<tr>
<td>May</td>
<td>151 (15.0)</td>
<td>64 (6.3)</td>
<td>November</td>
</tr>
<tr>
<td>June</td>
<td>159 (15.8)</td>
<td>45 (4.5)</td>
<td>December</td>
</tr>
</tbody>
</table>
annual precipitation, and annual streamflow. For the latter index they observed 303 increasing and 55 decreasing trends for the 1,009 sites. Overall they found that “strong trends in all variables exist at a large number of stations, many more than would be expected due to chance.” They also point out that aggregation of data into indices such as seasonal or annual can obscure patterns and significant trends in the data.

In an effort to produce a more detailed analysis of how change might be impacting the streamflow regime, Lins and Slack (1999) tested for trend in seven quantiles of the annual streamflow distribution. They looked at the annual daily minima (0th quantile) and maxima (100th quantile) as well as 10th, 30th, 50th, 70th, and 90th quantiles using the Mann-Kendall test. They based their analysis on a subset of 395 stations from the HCDN (Slack and Landwehr, 1992) that had continuous daily records for the period 1944 through to 1993. They also examined the impact of inter-decadal variability on their results by looking at the last 30 through 80 years of record, every ten years. They observed a strong pattern for the annual daily minimum streamflow for all periods analyzed. Nationally, there were more increasing trends observed than downward, yet both were greater than one might expect from chance. Similar results of increasing trends were found through the lower to mid-range of streamflow regimes, with the number of decreasing trends themselves decreasing as one progresses from the lower to mid-range. At the higher quantiles (90th and 100th), increasing and decreasing detection levels are approximately equal with a slight decrease observed for the annual maximum daily streamflow (see Table 2). They found that the pattern of the results for the inter-decadal analysis were similar with some variation in the overall number of detections, which one would expect. For example, for the period 1954 through to 1993, 31 stations exhibited an upward trend for annual maximum while 22 were decreasing.

Lins and Slack (1999) provided maps of the significant trend results for the annual minimum, annual median and maximum daily streamflow analyses. They observed that trends in the annual minimum were similar to the median and that there was evidence of spatial patterns or clustering. Their results indicated that much of the eastern portion of the US except for possibly the extreme southeast may have experienced increases in streamflow quantiles below the annual maximum. In contrast, the extreme northwestern portion of the US through to mid-California seem to be experiencing decreased streamflow quantiles.

A further analysis on trends in US streamflow was performed by Douglas et al. (2000) wherein a regional average Mann-Kendall statistic and bootstrapping were used to evaluate field significance. Field significance, popularized by Livezey and Chen (1983), attempts to take into consideration the cross-correlation amongst stations in the overall assessment of the significance of the results. In essence, it has long been known that serial and cross correlation can reduce the effective sample size for an experiment as compared with independent data (Matalas and Langbein, 1962). When cross-correlation exists between sites, then the results of each site’s hypothesis tests are also not independent. The argument put forth by Douglas et al. (2000) is that some studies that have applied the Mann-Kendall test have not evaluated the effect of the spatial correlation structure on the interpretation of the results.

Table 2 Number of significant (p ≤ 0.05) streamflow trends for various streamflow quantiles for 395 hydrometric stations in the conterminous US for the 50-year period 1944–1993 (adapted from Lins and Slack (1999))

<table>
<thead>
<tr>
<th>Significant Trend</th>
<th>Streamflow Quantile</th>
<th>Minima</th>
<th>10th</th>
<th>30th</th>
<th>50th</th>
<th>70th</th>
<th>90th</th>
<th>Maxima</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upward</td>
<td></td>
<td>127</td>
<td>117</td>
<td>125</td>
<td>113</td>
<td>61</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td>Downward</td>
<td></td>
<td>36</td>
<td>26</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>17</td>
<td>21</td>
</tr>
</tbody>
</table>
Douglas et al. (2000) performed their experiment on two hydrological indices, namely the annual maximum daily streamflow and the annual 7-day low flow value, over two time periods comprising 1959 through to 1988 (30 years) and 1939 through to 1988 (50 years). The analysis was also performed over two spatial scales with the larger scale dividing the US into three areas described as the west, mid-west and east. The smaller scale comprised nine areas. These two scales were used to evaluate the impact of regional size on field significance. Hydrometric stations used in the experiment were a sub-set of the HCDC (Slack and Landwehr, 1992). Their results based on bootstrapping preserved the cross correlation structure of the data and reflected the serial correlation structure in the estimation of the significance of the results. Based on the bootstrapped evaluation of significance, they found no significant trends in flood flows for both scales, but trends in the 7-day low flow were observed for both scales. They found evidence of increasing trend in the large-scale mid-west area, and in the smaller scale areas denoted as Ohio, north-central and upper mid-west. Their overall results were generally consistent with the results of the previously referenced work on US streamflow data.

A few investigations have also been performed on Canadian data, and these provide a more complete coverage of possible trends in streamflow regimes for North America. Zhang et al. (2001) analyzed a number of indices including annual and monthly streamflow, and a number of quantiles from the minimum daily through to the maximum daily discharge including the 10th, 20th...90th. They also analyzed for trend in the freeze-up, break-up and days with ice cover records. Their analyses covered three periods from 1967–1996, 1957–1996 and 1947–1996, and all sites were selected from the Canadian RHBN (Environment Canada, 1999; Harvey et al., 1999). The simulation approach proposed by Livezey and Chen (1983) was used to assess overall field significance, with data being pre-whitened as described by von Storch and Navarra (1995) should the serial correlation coefficient exceed a value of 0.1. Zhang et al. (2001) concluded from their analysis that decreasing trends were occurring across most of Canada “throughout the entire daily streamflow regime” and that “Canada is not experiencing more extreme hydrological events.” Changes were evident in timing of streamflow, for example March and April streamflows were found to be increasing, river freeze-up and break-up dates were also changing. They found freeze-up to be occurring earlier in the fall and the spring break-up to be occurring earlier except in Atlantic Canada. Table 3 shows a sub-set of their results for the period 1957–96, wherein all results except for the 70th percentile were found to be field significant. Results were given as percentages, and the actual number of stations with significant trend was not reported. The results of the site trend statistics for the annual streamflow for this period were not found to be field significant and had 4.2% and 8.5% of the sites with upward and downward trends. The annual streamflow for 1967–1996 and 1947–1996 had shown significant downward trends, and both were found to be field significant.

Burn and Hag Elnur (submitted) have also analyzed several streamflow indices, including annual streamflow, monthly streamflow, annual maximum daily streamflow, the date of the occurrence of the annual maximum flow, and river freeze-up and break-up dates for

<table>
<thead>
<tr>
<th>Significant Trend</th>
<th>Streamflow Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minima</td>
</tr>
<tr>
<td>Upward</td>
<td>7.5</td>
</tr>
<tr>
<td>Downward</td>
<td>17.5</td>
</tr>
</tbody>
</table>

Table 3: Percentage of significant (p ≤ 0.10) streamflow trends for various streamflow quantiles for hydrometric stations in Canada for the period 1957–1996 (adapted from Zhang et al., 2001)
Canada. They employed the Mann-Kendall test and pre-whitened the series provided its serial correlation was significant at the 0.05 level. Their results were similar to those of Zhang et al. (2001) for coincidental indices of ice conditions, annual streamflow and annual maximum daily streamflow. Burn and Hag Elnur (submitted) found that there was no significant trend detected in annual streamflow for the period 1960–97 based on RHBN data. They analyzed 86 sites and detected five sites with increasing trend and six with decreasing trend. They also analyzed the annual maximum daily values and found for 104 sites that four showed evidence of increasing trend while 14 showed decreasing trend. Mapping of these results indicated a decreasing trend in the annual maximum flow in the south and an increasing trend in the north.

Table 4 provides a summary of the results of Burn and Hag Elnur (submitted) and Zhang et al. (2001) for monthly streamflow for two similar time periods. From, in part, these results, Zhang et al. (2001) concluded that monthly streamflow for most months has decreased, with the strongest occurring in the summer and autumn months. Their detection results, which are listed in Table 4, have possibly six downward trends and only two months with upward trends. Burn and Hag Elnur (submitted), on the other hand, reported five months with significant trend. Both results tend to show an increase in the spring months of March and April followed by lower flows in May and June. There is also some evidence of decreasing flows in September and October. These results are consistent with the occurrence of an earlier spring melt and an earlier fall freeze-up of the river system in the Canadian environment. Differences between the results of both studies may lie in the slightly different periods of analysis, a slight variation in the sites covered by the analyses, and in the different approaches taken for pre-whitening the data. It is also interesting that these results are somewhat different than those obtained by Lettenmaier et al. (1994), which are summarized in Table 1. Similarities or differences in spatial pattern of the trend prove more difficult to assess as Zhang et al. (2001) have only provided maps for the months of April and September. There appears to be little correspondence in patterns for these two months with the maps of Lettenmaier et al. (1994). It should also be noted that the periods of analysis for these two studies are not consistent, and the earlier work did not pre-whiten the series prior to the application of the Mann-Kendall test.

A comparison of results for streamflow quantiles may also be made based on the results obtained by Lins and Slack (1999), as listed in Table 2, with that obtained by Zhang et al. (2001), as listed in Table 3. A review of these two tables shows there are some marked

<table>
<thead>
<tr>
<th>Month</th>
<th>Burn and Hag Elnur (submitted)</th>
<th>Zhang et al. (2001)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of sites</td>
<td>Upward (%)</td>
</tr>
<tr>
<td>Jan.</td>
<td>92</td>
<td>10 (10.9)</td>
</tr>
<tr>
<td>Feb.</td>
<td>90</td>
<td>10 (11.1)</td>
</tr>
<tr>
<td>Mar.</td>
<td>102</td>
<td>28 (27.5)</td>
</tr>
<tr>
<td>April</td>
<td>106</td>
<td>31 (29.2)</td>
</tr>
<tr>
<td>May</td>
<td>107</td>
<td>14 (13.1)</td>
</tr>
<tr>
<td>June</td>
<td>111</td>
<td>3 (2.7)</td>
</tr>
<tr>
<td>July</td>
<td>113</td>
<td>7 (6.2)</td>
</tr>
<tr>
<td>Aug.</td>
<td>114</td>
<td>3 (2.6)</td>
</tr>
<tr>
<td>Sept.</td>
<td>115</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>Oct.</td>
<td>114</td>
<td>11 (9.6)</td>
</tr>
<tr>
<td>Nov.</td>
<td>101</td>
<td>9 (8.9)</td>
</tr>
<tr>
<td>Dec.</td>
<td>92</td>
<td>6 (6.5)</td>
</tr>
</tbody>
</table>

Table 4 Number and percentage of significant (p ≤ 0.10) monthly streamflow trends and the number of sites analyzed for the 1960–1997 period (adapted from Burn and Hag Elnur (submitted)) and the percentage of significant (p ≤ 0.10) monthly streamflow trends for the 1957–1996 period reported by Zhang et al. (2001)
differences in results, particularly for the low to mid-range of streamflow quantiles. Lins and Slack (1999) had found increasing trends through to the 70th percentile, with slightly elevated decreasing trends at the minimum and 10th percentile range. Possibly the most striking difference is obtained for the annual daily minimum flow. Zhang et al. (2001) reported 10.2% of sites with upward trends and 16.3% of sites exhibiting downward trends for the period 1947–1996. It is difficult to ascribe why results should be so incongruous. Spatial patterns of Zhang et al. (2001) appear to be generally consistent with the results of Lins and Slack (1999) along the Canada-USA border. Discrepancies in results might be due to the different analytical periods analyzed in both studies and the differences in analytical procedures of the two studies (Zhang et al. used prewhitening while Lins and Slack did not). Potential land-use and water management effects might have also impacted upon streamflow data for one or more sites within either or both study networks. There is also the possibility that change is simply not occurring consistently throughout the combined study area. Yue et al. (submitted) were concerned with the potential impacts of serial correlation on the performance of the Mann-Kendall test. They also looked at the effects of prewhitening as advocated by von Storch and Navarra (1995), which is used in various studies (e.g., Zhang et al., 2001; Burn and Hag Elnur, submitted). von Storch and Navarra (1995) have shown that if a series had serial correlation and no trend, application of the Mann-Kendall test would result in an overestimate of the number of detections well above the selected significance level of the test. Prewhitening was shown to have corrected this deficiency, thereby reverting the detection rate to that expected for the selected level of significance. Yue et al. (submitted) results supported the work of von Storch and Navarra, but what Yue et al. also illustrated was that should trend be evident in the series along with positive serial correlation, then the prewhitening procedure removes a portion of the trend, possibly making it undetectable. Douglas et al. (2000) also provided results that inferred prewhitening might decrease the detection rates. Yue et al. (submitted) suggested a correction to the prewhitening procedure where linear trend was first identified and removed using the Theil-Sen approach (Theil, 1950; Sen, 1968); the series was then prewhitened. Subsequently, the trend term is then added to the prewhitened residual series. Simulations indicated that removal of trend as a first step did not negatively impact on the estimate of the serial correlation, and that the significance of the identified trend could be suitably identified in the resultant series. Prewhitening within this procedure is only performed should the serial correlation be statistically significant ($p < 0.10$). They termed their modification to the classical prewhitening approach as Trend-Free Prewhitening (TFPW).

Using TFPW, Yue et al. (in press) investigated the trend in three Canadian streamflow indices, namely annual daily minimum, annual mean, and annual daily maximum streamflow. They also extended the field significance statistic of Douglas et al. (2000), whose significance was established using bootstrapping, to evaluate the tails of the distribution. They analyzed data from the Canadian RHBN (Environment Canada, 1999, Harvery et al., 1999) for the 1957–1997 period and for all possible sites. They provided a summary of the results for three different approaches, namely conventional Mann-Kendall application, prewhitened Mann-Kendall, and TFPW. Results were also given for the significance of the average Mann-Kendall statistic, $S_m$, and the significance from assuming a simple binomial model. The binomial model can be used to assess the field significance of trend if cross or serial correlation have not impacted on the analyses.

Yue et al. (in press) reported that the annual daily maximum streamflows for Canada showed a significant downward trend based on the bootstrapped results for $S_m$ of the TFPW Mann-Kendall test scores. The mean Mann-Kendall field significance statistic was not significant for either the annual mean or annual daily minimum streamflows. TFPW
bootstrapping for annual daily minimum streamflow data did indicate significant upward trend (13 of 71 sites) and significant downward trend (11 of 71 sites) for $p \leq 0.05$. Annual mean streamflows also showed significant upward trend (8 of 63 sites). Interpretation of these results must be based, in part, on understanding the field significance statistic advocated by Douglas et al. (2000), which was used by Zhang et al. (2001) and Burn and Hag Elmur (submitted). This field statistic, $S_m$, represents the average of all Mann-Kendall results for the area of analysis and as such is a measure of the central tendency of the test statistic’s distribution. The extension of Yue et al. (in press) looks at the tails of the distribution as above or below a pre-set value. The results for the annual daily minimum streamflows using TFPW bootstrapped results show that although the central tendency of the results is not suspect, both tails of the distribution have more exceedances than one would expect from change alone. In other words, the distribution has been flattened, but its central tendency has not been significantly altered. One would expect that the average test statistic would identify change should it be occurring in one direction and would be less apt to identify change should it not be uniform in direction within the study area.

Yue et al. (in press) extended their analysis of the three Canadian streamflow regimes to all possible annual streamflow stations within the RHBN having as a minimum 20 years of observations. This was performed for two reasons. The first was to allow a better spatial analysis of the patterns of trends. They felt that the patterns for the regimes were consistent for the 1957–1997 period and the “all data” case. The second reason was to allow for a more accurate analysis of individual site significance. They also provided a summary of individual site significance for the all site case for the classical Mann-Kendall test, the prewhitened Mann-Kendall test, as advocated by von Storch and Navarra (1995), and TFPW. Table 5 shows the summary of their results.

Yue et al. (in press) have also indicated that the number of sites with positive significant serial correlation ($p \leq 0.10$) for the above three indices are 23, 16, and 5, for the period 1957–1997 and 77, 51, and 29 for all possible stations, for the order of the index as it appears in Table 5. Using the TFPW with the binomial distribution, they found that for the all station case, only the annual daily maximum flow showed no significant upward trend ($p \leq 0.10$). In other words, all indices were showing evidence of upward and downward trends except for the annual maximum, which showed a consistently downward trend. A visual assessment of the results was also provided and indicated that, for example, annual daily minimum flows were showing strong signs of clustering of results in sub-regional areas of the country, and at times rather broad bands of like-signed results reaching across the continent. Once again, trend was not occurring similarly in all areas at the same time. Patterns for the annual minimum also seemed to be consistent with that observed for the annual mean. Annual maximum showed a general downward trend with little upward pattern evident. These results also show that pre-whitening tended to provide more conservative estimates of the number of sites having significant trend.

A visual comparison of the maps of Lins and Slack (1999) and Yue et al. (in press) for sites having significant trend in annual daily minimum and maximum streamflow series

<table>
<thead>
<tr>
<th>Index</th>
<th>No. of Sites</th>
<th>Number of Upward Trends (%)</th>
<th>Number of Downward Trends</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MK Prewh. TFPW</td>
<td>MK Prewh. TFPW</td>
</tr>
<tr>
<td>Min.</td>
<td>209</td>
<td>46 (22.0) 29 (13.4) 42 (20.0)</td>
<td>38 (18.2) 29 (13.9) 41 (19.6)</td>
</tr>
<tr>
<td>Mean</td>
<td>213</td>
<td>19 (8.9) 10 (4.7) 20 (9.3)</td>
<td>25 (11.7) 25 (11.7) 30 (14.1)</td>
</tr>
<tr>
<td>Max.</td>
<td>213</td>
<td>8 (3.8) 7 (3.3) 8 (3.8)</td>
<td>34 (16.0) 27 (12.7) 36 (16.9)</td>
</tr>
</tbody>
</table>
show excellent agreement along the bordering areas. These patterns change as one progresses further northward. Looking at the annual minimum shows there is consistent downward trend in the western US. As one crosses into westernmost Canada, there is strong evidence of downward trends as well. These patterns shift to upward trend as one moves further northward yet again. A similar tendency is also seen for eastern Canada. Progressing from southern US to northern Canada, there is first a band of downward tendencies shifting to upward tendencies in the Great Lakes basin, which then shift to downward trends as one progresses further to the north-east.

Looking now to South America, there have been two studies dealing with five river systems. Genta et al. (1998) investigated the annual and monthly streamflows for four major rivers in the southeastern portion of the continent. These included the Uruguay (500,000 km²), Negro (39,700 km²), Paraná (975,000 km²), and Paraguay rivers (1,100,000 km²). Dams were built in the Negro and Uruguay basins in 1939 and 1980. The authors also reported that the Negro basin has not been affected by Amazonian deforestation as have the other basins. A plot of the annual streamflow for the four systems indicated a large similarity of patterns with monotonically increasing flows from the mid 1940s to 1960s onward. Genta et al. (1998) divided the records into two time periods, the first comprising the beginning of the record through to 1940, which spans approximately 30 years of record. The second period was from 1940 through to 1995, the last year of available data. It was reported that a Wilcoxon test was applied to the split sample for the four basins resulting in the conclusion that the median flows for the latter years were significantly larger than for the preceding years. A monthly analysis of streamflows showed that the monthly values tended to be larger for all basins most of the time, except for the Negro River.

In a second study of South American streamflows, García and Vargas (1998) have analyzed the annual and monthly flows for the Río de la Plata basin (3,100,00 km²) for potential shifting of the mean. They analyzed data from several gauges within the major tributaries of the basin, including the following rivers: Paraná, Uruguay, Paraguay, Iguazú (64,400 km²), and Paranapanema (100,000 km²). They also found a significant shift in the mean annual streamflow between 1970 to 1972, along with some minor shifts occurring in the 1917–1918 and 1943–1944 periods. Graphs were provided that showed the monthly streamflows for the major basins for various periods including 1971–1993, the period where annual streamflow has been shown to have increased. The monthly streamflows for the Paraguay River were shown to be all higher for the more recent period as was also the case for the Paranapanema River. The Iguazú River showed less dramatic departures than its counterparts, yet it showed increased mean flows for the latter period for 10 of 12 months. The Uruguay River also showed a tendency to increased monthly streamflow except for April and May. The Paraná River showed a tendency to have increased monthly streamflow except for the month of March, which happens to be typically a high flow period. The authors noted that the river system has over 280 km³ of storage volume in over 60 dams. They also noted that they were not certain how anthropogenic activities might have impacted on the results of the analysis, due to a general lack of available data upon which to base additional analyses.

In one European study, Robson et al. (1998) investigated the possibility of trends in flood series within the United Kingdom. Peaks over threshold data and annual daily maximum streamflows from 890 gauging stations were pooled to form two annual series spanning three time spans of 1941–1980, 1941–1990 and 1870–1995. Although the authors noted that urbanization has increased dramatically over the time horizon of the study, efforts were not made to select pristine or stable basins. Linear regression, normal scores regression, and Spearman’s correlation test with a permutation bootstrap to assess significance were applied and resulted in no significant trends being identified. An analysis of
annual rainfall indicated no trend in the 40-year series of 1941-1980, but there was a “suggestion of trend in the 50-year rainfall data” series. The winter precipitation series for the 50-year period showed strong evidence of trend. Unfortunately, no seasonal flood data were readily available to allow a further analysis of the association between seasonal rainfall and streamflow.

Chiew and McMahon (1996) reported on a “global” investigation into detection of changes in the mean and trends via five statistical tests for annual streamflow volumes and “peak discharge”. Reporting of results is based on the number of tests that showed significant results per site tested. The tests included the Mann-Kendall, the cumulative deviation, the Worsley Likelihood ratio, distribution-free CUSUM, and the Kruskal-Wallis. These tests were performed on 142 rivers throughout the world, with the greatest concentration of sites in the US, Europe, and eastern Australia. The authors took measures to ensure that the basins represented unregulated flow conditions, yet they indicated it was not possible for them to ensure that the land-use and consumptive practices had been stable, due to lack of local knowledge regarding most of the basins. Selected sites all had 50 or more years of record and drain an area greater than 1,000 km². The largest basin analyzed exceeds 8,000,000 km², while the median basin size is reported as 16,000 km².

They reported some evidence of significant results for trends of the two indices in North American data by four or five of the tests in 20% of the stations analyzed. Negative trend was uncovered at several northeastern sites, while upward trends were found in annual streamflow volumes, which should be analogous to annual mean streamflow, in the southeast. In contrast, the results of one to three tests were significant in less than 10% of the cases for volumes and 25% for the peak flow sites. Trends detected in European data were generally downward. In Australia, only about 15% of the 21 sites analyzed showed signs of trend by more than one test. Very limited data were available for the other continents and will not be reported herein. The authors indicate that “if flow conditions in a geographical region have altered as a result of climate change, most of the statistical tests should detect changes in the streamflow records of all stations in the region.” This is a hypothesis that does not seem supported by the evidence obtained by the other studies performed on the higher density North American data sets, depending of course on the definition of a “region”.

Discussion and recommendations
This paper has focused on the statistical identification of trend in streamflow series. Various studies cited in this paper have provided evidence of a large number of sites that show the existence of trend for certain streamflow regimes, more than one would expect due to chance alone. Mapping of trend results, when network density is sufficient, have shown visible patterns or tendencies. This paper has not focused on the attribution of trend within the hydrological cycle and what might be causing the streamflow regimes to alter. This is an area that requires further attention.

Primarily from the analyses performed on North American data, there is evidence of significant trend in many streamflow indices and indications that the patterns of trend might be shifting over relatively small geographical areas. This was evidenced when mapped results had definite alternating patterns across the North American continent such as those observed for the annual minimum daily streamflow series. It is certain from these emerging patterns that change is not occurring the same in all areas, nor in the same general direction for the same index (e.g., annual daily minimum streamflow).

Even though various studies have been performed using similar statistical tests, differences in results can occur stemming from the impacts of how serial and cross-correlations are taken into account. Some studies have not prewhitened their data series, while others
have prewhitened in various ways. Selection of field significance statistics can also impact on interpretation of results. It is evident that the selection of prewhitening procedures (or not) and limited use of statistical approaches to assess field statistics can, at times, result in dramatically different conclusions. More effort is required to assess the impacts of violating assumptions of common statistical procedures and of approaches that attempt to overcome them. Additional efforts are also required to assess the impacts of using different methodologies and in an assessment of which approaches might be most accurate.

Of particular importance is the recognition that results as obtained over most of North America were made possible by the careful selection of hydrometric sites to represent pristine or stable conditions. The relatively high density of sites across the US and Canada made the identification of potential geographical patterns possible. A very sparse network might have resulted in either utter confusion of results or the incorrect identification of local trend.

In order to advance the investigation of potential change in streamflow regimes, there is need for the establishment of a global network of carefully chosen hydrometric basins. The selection criteria must be applied uniformly and include local knowledge in the evaluation of potential candidate sites. It is recommended that the sites reflect pristine or stable conditions to avoid the confusion of anthropogenic signal from direct human intervention on the flow regime (changing land-use patterns, river regulation, etc.). The density of the global network must be sufficient to allow for recognition of regional evidence of patterns of change, otherwise statistical results might prove meaningless or misleading. It is recognized that certain areas of the world might not be able to contribute to such a network due to the degree of human intervention on their streamflow regimes that would adversely impact on the trend analysis.

References


