INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **25**: 485–504 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/joc.1141

FREQUENCY ANALYSIS AND TEMPORAL PATTERN OF OCCURRENCES OF SOUTHERN QUEBEC HEATWAVES

M. N. KHALIQ,[a,*] A. ST-HILAIRE,[b] T. B. M. J. OUARDA[b] and B. BOBÉE[b]
[a] *Ouranos, 550 Sherbrooke Street West, Montreal, Quebec H3A 1B9, Canada*
[b] *INRS-ETE, University of Quebec, 490 de la Couronne, Quebec City, Quebec G1K 9A9, Canada*

Received 11 August 2004
Revised 5 November 2004
Accepted 9 November 2004

ABSTRACT

Heatwaves can have adverse affects on public health and can considerably impact social and economic activities. Climate-change scenarios have shown that the temperature regime will likely be modified significantly over the course of the next 50 years and more. The frequency of occurrence and amplitude of heatwaves may be impacted by changes in the temperature regime. A heatwave can best be characterized by its magnitude and duration. Thus, both of these characteristics need to be studied together. This paper presents an approach based on the principle of parsimony by extending methodologies developed for the analysis of extreme hydrological events: the index-flood method and the regional flood frequency approach to perform at-site heatwave–duration–frequency (HDF) analysis. The approach is very similar to intensity–duration–frequency often used in the analysis of extreme precipitation events. The HDF analysis is performed using annual maximum series of heatwaves of 1–10 days duration observed at four selected sites in southern Quebec with long (i.e. >80 years) time series covering most of the 20th century. The two main tasks in this approach consist of modelling (1) $\mu(D)$, a function that relates mean heatwave to its duration, and (2) $g(T)$, a function describing the at-site dimensionless growth curve, where T is the return period. It is found that the $\mu(D)$ function can best be modelled using a relationship of the form $\mu(D) = aD^b$ (where a and b are parameters to be estimated). The dimensionless growth curve $g(T)$ was modelled using the generalized extreme value distribution. The HDF approach can model various quantiles of heatwaves in a fairly acceptable manner when assessed on the basis of relative root-mean-square error and 95% bootstrap confidence intervals. An analysis of the pattern of occurrences of heatwaves indicates that heatwaves of short durations (1–5 days) has shifted over time and occur earlier in the summer than before. Median heatwaves occur during the second and third weeks of July and the majority of heatwaves are concentrated over the time interval from the last 10 days of June to the first 10 days of August. Copyright © 2005 Royal Meteorological Society.

KEY WORDS: annual maximum series; frequency analysis; generalized extreme value distribution; heatwaves; southern Quebec; trend analysis

1. INTRODUCTION

It has been shown in many parts of the world that extreme air temperatures of long duration, so-called heatwaves, can have adverse effects on public health (Huth *et al*., 2000; Duenas *et al*., 2002; Rainham and Smoyer-Tomic, 2003) and on the economy (Subak *et al*., 2000). Recently, Europe experienced an intense heatwave that created vast health hazards and claimed thousands of lives (UK Guardian, 2003). In Canada, a study performed in the nation's largest city (Toronto) investigated the link between non-accidental human mortality and heat stress. A statistically significant relation was established between the two variables measured for the period 1980–96 (Rainham and Smoyer-Tomic, 2003). Although that study stated that the impact was minimal for the population of Toronto during this period, it is possible that the impact of heatwaves on the urban population in Canada may become a more serious public health problem in the future. Bonsal

* Correspondence to: M. N. Khaliq, Ouranos, 550 Sherbrooke Street West, Montreal, Quebec H3A 1B9, Canada;
e-mail: khaliq.naveed@ouranos.ca

et al. (2001) studied the characteristics of extreme temperatures over Canada. They found no significant trend over the course of the 20th century (1900–98) for the higher percentiles of daily summer maxima. They concluded that the number of extreme hot days showed little change during this period. However, mean annual temperature over Canada has increased by 0.9 °C between 1900 and 1998 (Zhang *et al.*, 2000), and a recent climate-change modelling study by Kharin and Zwiers (2000) has shown that daily maximum temperature extremes with a return period of 20 years may increase by as much as 2.4 °C over the course of the next 50 years. In light of such findings and scenarios, it is of the utmost importance to develop statistical tools that may assist in providing a sound basis of comparison for future changes in extreme air temperature events, such as summer heatwaves.

Frequency analysis is a widely used tool for describing the behaviour of hydro-meteorological variables, like flood peaks, low flows, flood volumes, drought duration, temperature and wind speed. To carry out at-site and regional frequency analyses, generally two different approaches are used to define the time series of the extreme variable of interest, i.e. annual maximum series and peaks-over-threshold series. In the annual maximum approach, only one value per year is used, whereas in the peaks-over-threshold approach there can be more than one value per year (i.e. all values above a certain threshold are included). In order to extract either annual maximum or peaks-over-threshold series from the observed data, instantaneous observations (e.g. flood peaks), time-averaged observations (e.g. hourly rainfall intensity) or time-accumulated values (e.g. flood volumes, 24 h rainfall depth, degree-days) are used. Depending upon the type of analysis and application, other variables can also be defined and analysed. For methodological developments and application procedures of frequency analysis techniques in the areas of hydro-meteorology, readers are referred to the reviews and comparison of techniques compiled by Cunnane (1989), GREHYS (1996), Madsen *et al.* (1997a, b), Lang *et al.* (1999), Smakhtin (2001) and Katz *et al.* (2002).

Frequency analysis of rainfall of various durations has long been in use to develop rainfall intensity–duration–frequency (IDF) curves for sizing hydraulic structures (Viessman *et al.*, 1977). IDF curves can be developed for a particular site, for a geographical region or for a homogeneous region containing more than one site. Systematic methods for grouping a number of sites to form a homogeneous region have been developed by Hosking and Wallis (1997) and Ouarda *et al.* (2001) in the context of regional flood frequency analysis. Sveinsson *et al.* (2002) and Ferro and Porto (1999) have used the approach of regional flood frequency analysis of Hosking and Wallis (1997) to develop regional IDF relationships for hydrological homogeneous regions. Inspired by the IDF analysis and regional flood frequency analysis, Javelle *et al.* (2000, 2002, 2003) have presented flood–duration–frequency (QDF) analyses. In their studies, a 2-day flood was defined as the average of two consecutive days of maximum daily flows. Analogously, the present study is related to frequency analysis of daily maximum temperatures to characterize heatwaves. Merely performing a frequency analysis of daily maximum temperatures is not sufficient to characterize heatwaves, since these extreme events can only be fully described by associating duration with temperature magnitude. Therefore, a heatwave is defined as a run of consecutive days (say from 1 to 10 days) for which the observed average daily maximum temperature (DMT) is the highest of the year and, similar to the QDF approach, a 2-day heatwave is defined as the largest average of two consecutive days of daily maximum temperature. This approach is referred to as heatwave–duration–frequency (HDF) analysis henceforth. The HDF concept is akin to the IDF and QDF approaches. There are two main objectives of this study. The first objective is to develop a parsimonious HDF approach by using annual maximum series, the index-flood method of Dalrymple (1960) to define index-heatwave and regional frequency analysis approach to perform at-site analysis. The second objective is to study the temporal pattern of occurrences of heatwaves. In most of the regional frequency analyses, the index-statistic, which is used to standardize at-site data, is usually taken as the at-site mean. In this work, the index-heatwave for any particular heatwave duration is taken as the corresponding mean value (i.e. the index-heatwave for 2-day heatwaves is the mean value of all observed 2-day heatwaves). The minimum duration of a heatwave is taken as 1 day in order to develop a continuous formulation of the HDF approach. However, the prescribed methodology is flexible to accommodate any other minimum duration longer than 1 day.

In southern Quebec, maximum daily summer temperatures can exceed 30 °C during the summer months. In urban areas such as Montreal, such high temperatures may have adverse effects on public health, depending

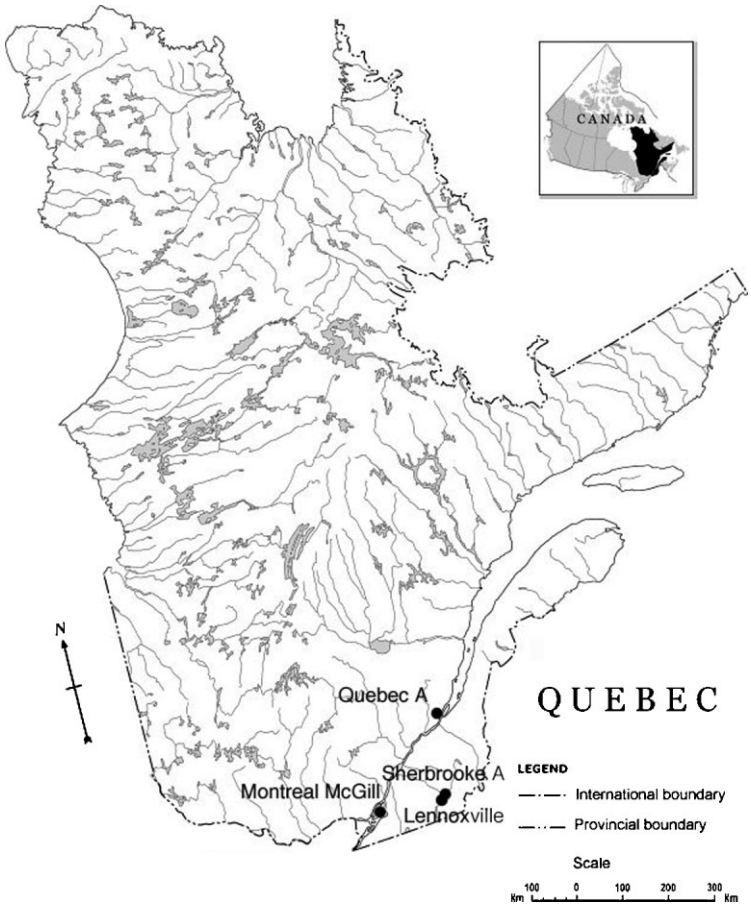

Figure 1. Study area map showing location of individual stations

on their duration. The results of the HDF analysis would be useful to determine heatwaves of various return periods and of various durations. It is hoped that this case study will complement the work done by Kharin and Zwiers (2000), who have used frequency analysis in their study of air temperature extremes but who did not take into account the duration of events. The HDF approach is demonstrated using homogenized daily maximum temperatures from four stations located in southern Quebec, Canada (Figure 1) for the 5 months period from May to September. The homogenized daily maximum temperature data have been compiled by Vincent *et al*. (2000) for climate-change studies in Canada. The use of homogenized time series reduces the risk of observing trends that may be caused by logistic problems (i.e. change in equipment, station location, sampling frequency) or equipment drift. The four stations studied (station codes) are Lennoxville (7 024 280), Quebec A (7 016 294), Montreal McGill (7 025 280) and Sherbrooke (7 028 124) with respective record lengths of 81 (1915–95), 101 (1895–1995), 98 (1895–1992) and 91 years (1904–94). Although July and August are the peak summer months, individual hot days or groups of two or three hot days have been observed as early as mid May and as late as the second last week of September. Therefore, the analysis of daily maximum temperature series including the months of May to September guaranteed the inclusion of all peak values.

This paper is organized into five sections. The methodology of HDF analysis is presented in Section 2, along with the data screening procedures. Detailed results of HDF analysis and the pattern of occurrences of heatwaves are presented in Section 3, followed by discussion in Section 4. Conclusions are presented in Section 5. Detailed analyses and results are presented for one station (Montreal McGill) in particular and for all four stations in general. A full set of analyses and results can be seen in Khaliq *et al*. (2004).

2. METHODOLOGY

2.1. Data extraction

The time series of heatwave of duration D-day is denoted by H_D. Thus, for year i, H_{D_i}, is the largest value of average maximum temperature observed consistently for any D-day duration during the May–September period (e.g. $H_{3_i} = 30\,^{\circ}\mathrm{C}$ if the observed daily maxima has been $30\,^{\circ}\mathrm{C}$ during three consecutive days in year i). Similarly, μ_D and σ_D are the mean and standard deviation of the H_D series. To extract heatwaves of longer duration than 1 day, temperature data for various durations need to be computed from DMT series. Temperature time series for various durations are extracted from DMT series using a moving average with a window of width $D = 2$, 3, 4, 5, 6, 7 and 10 days. Occasionally, reference has been made to H_D series for $D = 15$ and 30 days to verify some of the assumptions used in the analysis. However, these two series were not subjected to detailed analysis. The peak summer period (i.e. July–August) was given more importance when deciding the length of record to be considered for the analysis. For example, if for any year the DMT data for both July and August were missing then that year was not included in the analysis. Furthermore, any moving-window estimate containing a missing value was also not considered. The missing observations during the summer season for the studied stations are $\leq 1\%$.

2.2. Quality of heatwave data

All heatwave time series were checked visually and using statistical tests of significance to detect non-stationarities (i.e. presence of trends and jumps). Two non-parametric tests, the Spearman rank correlation test (Dahmen and Hall, 1990) and the Mann–Kendall (Kendall, 1975) test, were selected to verify the absence of trend and the independence of a time series. To strengthen the results of these two tests further, Sen's (1968) non-parametric method was used to obtain an estimate of a monotonic trend. In addition to these non-parametric tests, ordinary linear regression slope and correlograms (Salas *et al.*, 1980) were used to verify the absence of trend and independence of observations. Also, the parametric t-test for stability of mean and the F-test for stability of variance were selected to study stationarity of heatwave observations. Detailed descriptions of the t- and F-tests can be found in Snedecor and Cochran (1989) and the methods of application to ascertain stationarity are given in Dahmen and Hall (1990). The results of the Spearman rank correlation and Mann–Kendall tests and estimates of Sen slope and linear slope revealed that none of the H_D series for all the four stations exhibited evidence of trend at the 5% significance level. Thus, it can be concluded that H_D data for all four stations are free of time trends. These results of trend analysis are consistent with the previous findings of Bonsal *et al.* (2001), who found no significant trend for the higher percentiles of daily summer maxima. For studying persistence and temporal cycles, autocorrelations up to lag 20 along with 95% confidence intervals were obtained and the numbers of autocorrelations, falling outside the confidence intervals, were noted. For a 95% confidence interval, one out of 20 autocorrelations is expected to fall outside the confidence interval. Based on the study of autocorrelations, the only heatwave series that could be non-independent is H_1 at station Quebec A, because two autocorrelations fall outside the confidence interval. For applying the F-test for equality/stability of variance and the t-test for equality/stability of the mean, heatwave time series were divided into two and three non-overlapping subsamples of approximately equal length. Based on the combined results of the F- and t-tests at the 5% significance level, the H_1 series at Sherbrooke is found to be non-stationary in variance but stationary in mean. It was observed that the number of non-stationary cases increased when a more conservative level of significance, e.g. 10%, was used. It is conjectured that the inclusion of one non-independent and one non-stationary series into frequency analysis will have a negligible effect on the overall results of the remaining large number of series. However, the results of frequency analysis of these two series should be interpreted with caution.

2.3. Basic statistics of heatwaves

Basic statistics (i.e. mean, standard deviation and coefficients of variation and skewness) of all H_D series, as a function of duration, were studied to interpret their mutual variability and to formulate appropriate

hypotheses. It was found that the mean μ_D is a decreasing function of duration and it decreases in a parabolic fashion. However, unlike the mean, the standard deviation σ_D does not depict a regular increasing or decreasing pattern as a function of duration; perhaps, rather, a random pattern is exhibited, except for Quebec A, where a regular decreasing trend after duration $D = 2$ days was realized. The range (coefficient of variation) of σ_D for Lennoxville, Quebec A, Montreal McGill and Sherbrooke is 0.12 (0.008), 0.20 (0.042), 0.09 (0.022) and 0.12 (0.026) respectively. Small values for the range of σ_D suggest that it could be possible to assume a duration-independent value of σ_D for each station. In addition, it was observed that H_D series are just moderately skewed, i.e. the coefficient of skewness ranges between 0.39 and 0.84 for Lennoxville, between 0.12 and 0.37 for Quebec A, between 0.09 and 0.39 for Montreal McGill, and between 0.23 and 0.53 for Sherbrooke. A small range of coefficient of skewness indicates that it could be possible to assume a duration-independent value of skewness at each station. The behaviours of the basic statistics presented above, particularly the dominance of scale, are exploited in the domain of L-moments (Hosking, 1990) to formulate appropriate hypotheses in order to develop a parsimonious at-site HDF approach.

2.4. Selection of a statistical distribution for heatwaves

In order to model H_D series, some of the two-parameter and three-parameter statistical distributions are analysed and no attempt has been made to perform an exhaustive analysis. The two-parameter distributions include the normal (NOR), lognormal (LN2), and extreme value type I (EV1). The three-parameter distributions include the generalized extreme value (GEV), lognormal (LN3), Pearson Type III (P3) and log Pearson type III (LP3). For two-parameter distributions, only maximum likelihood (ML) estimators are used. For the LN3 and P3, both ML and ordinary moment estimators are used. For the GEV, both ML and probability weighted moments (PWM) estimators are used. The LP3 distribution is fitted using the procedure described by Bobée (1975). All of the parameter estimations are performed using the HYFRAN (2003) software. The choice of a best-fitting model is performed on the basis of three model selection criteria, i.e. the Akaike information criterion (AIC; Akaike, 1974), the chi-squared test statistic (Snedecor and Cochran, 1989) and the root-mean-square error (RMSE). According to each of these criteria, the best fitting model is the one with the smallest value of the criterion. Although the AIC goodness-of-fit criterion is constrained to ML estimation, it has been used regardless of the estimation method (Mutua, 1994; Strupczewski *et al*., 2001). The use of RMSE for model selection is demonstrated by Cunnane (1989) and a variant of it is demonstrated by Yu *et al*. (1993).

Detailed values of selection criteria are omitted, and summarized results are presented in Table I for the two-parameter distributions. It is obvious from this table that the order of best-fitting distribution is station dependent. The choice of best-fitting distribution is made as follows: for each record, each distribution is assigned a rank between 1 and 3, with rank 1 for the best fitting distribution (e.g. the distribution having

Table I. Summarized results of fitted two-parameter distributions. For each criterion, the order (rank) of best fitting is indicated

Criterion	Distribution	Rank			
		Lennoxville	Quebec A	Montreal McGill	Sherbrooke
AIC	NOR	3	2	2	2
	LN2	2	1	1	1
	EV1	1	3	3	2
Chi-squared statistic	NOR	2	1	3	2
	LN2	1	2	2	1
	EV1	1	2	1	3
RMSE	NOR	3	2	2	2
	LN2	1	1	1	1
	EV1	2	3	2	3

Table II. Summarized results of first four best fitting (based on AIC) three-parameter distributions. For each criterion, the order (rank) of best fitting is indicated

Criterion	Distribution	Rank			
		Lennoxville	Quebec A	Montreal McGill	Sherbrooke
AIC	GEV-ML	2	1	1	1
	LN3-ML	2	4	4	4
	P3-ML	1	2	3	2
	GEV-PWM	3	3	2	3
Chi-squared statistic	GEV-ML	4	4	2	3
	LN3-ML	3	3	2	1
	P3-ML	1	2	1	2
	GEV-PWM	2	1	2	2
RMSE	GEV-ML	2	2	2	1
	LN3-ML	3	3	3	2
	P3-ML	3	3	4	3
	GEV-PWM	1	1	1	1

lowest value of AIC). For each distribution, the ranks are summed over all eight durations and these totals of ranks are used to choose a best-fitting distribution. The appearance of the same rank in Table I means two (or more) distributions being at the same rank. Thus, all three criteria suggest that the LN2 is the overall best-fitting distribution. The NOR distribution, which has been used by Wettstein and Mearns (2002) to model daily maximum temperature series, ranks the second (or the third) best-fitting distribution for the current dataset. For the case of three-parameter distributions, summarized results of distribution fitting analysis are presented in Table II. According to AIC, the first four best-fitting distributions are the GEV-ML, LN3-ML, P3-ML and GEV-PWM. Hence, the results for the other two selection criteria are presented for these four distributions and parameter estimation methods only. Thus, based on the AIC, the overall best-fitting distribution is the GEV-ML, followed by LP3-ML and GEV-PWM. Based on the chi-squared statistic, the overall best-fitting distribution is the LP3-ML, followed by GEV-PWM and LN3-ML. Based on RMSE, the best-fitting distribution is the GEV-PWM, followed by GEV-ML and LN3-ML. In addition to the AIC, chi-square statistic and RMSE model selection criteria, the L-moment ratio diagram can also be used to select the most appropriate distribution for the H_D series. The definitions and methods of computing L-moments of various commonly used distributions are given in Hosking (1990). The L-moment ratio diagram represents the L-kurtosis of a sample as a function of its L-skewness. Figure 2 is such a representation of H_D series. The plotted points of the H_D series of Quebec A and Montreal McGill are close to LN3, P3 and GEV curves, indicating that any one of them could be a suitable choice for these two stations. However, this behaviour is less obvious for the H_D series of the remaining two stations, particularly for the H_D series of Lennoxville. Based on the results and discussion presented above, GEV-PWM is selected for modelling H_D series for all four stations.

2.5. The HDF approach

It is assumed that for any site the T-year heatwave, $H_D(T)$, can be expressed as the product of two terms. These two terms are the scale factor μ_D and the growth factor $g(T)$. The scale factor is called the index-heatwave and the growth factor describes the relationship between dimensionless heatwave and the recurrence interval T. The index heatwave is duration dependent, whereas the growth curve is assumed to be valid for the entire group of H_D distributions. This can be explained by the relationship $H(D, T) = \mu(D)g(T)$, where $H(D,T)$ is the modelled value of $H_D(T)$, $\mu(D)$ is the modelled value of μ_D and $g(T)$ is a dimensionless parent-fitted distribution, with a unit mean. This implies that all $H_D(T)$ distributions approach the same distribution when standardized by the respective mean value, and this assumption is in parallel to the

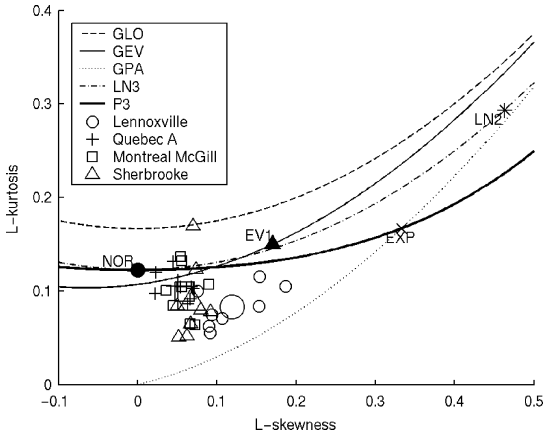

Figure 2. L-moment ratio diagram of heatwaves. Small symbols (open circle, cross, open square and open triangle) represent observed values and corresponding large symbols represent group averages. The GLO, GPA and EXP abbreviations stand for generalized logistic, generalized Pareto and exponential distributions

behaviours of the basic statistics presented in Section 2.3 and is further verified in Section 3.2. Based on the analysis presented above, $g(T)$ is taken to be the GEV distribution with the distribution function given by

$$F_Y(y; u, \alpha, k) = \exp\{-[1 - k(y-u)/\alpha]^{1/k}\} \tag{1}$$

where $-\infty < u, \alpha > 0$, and k are the location, scale and shape parameters respectively. The range of y is such that $1 - k(y-u)/\alpha > 0$. The parameters of the GEV distribution can be estimated either by pooling standardized observations across durations, and thereby estimating parameters from the pooled sample, or from the standardized PWMs averaged across durations. The latter approach, which is well documented by Hosking and Wallis (1997) for regional frequency analysis, is adopted here. The averaged standardized PWMs are defined below:

$$b(r) = \frac{1}{N_D} \sum_{\substack{D=1 \\ D \neq 8,9}}^{D=10} \frac{b_D(r)}{b_D(1)} \tag{2}$$

where $b(r)$ is the rth PWM of $g(T)$, $b_D(r)$ is the rth PWM of the H_D distribution, $b_D(1)$ is the first PWM of the H_D distribution, which is equivalent to μ_D, and N_D is the number of durations. The PWM parameter estimation for the GEV distribution is given in Hosking *et al.* (1985). The scale factor μ_D can be modelled as a function of D in a manner similar to the approach used in Javelle *et al.* (2002, 2003) for QDF analysis and in Ferro and Porto (1999) for regional IDF analysis. The general form of this relationship is given by $\mu(D) = f(D)$. Six different forms of $f(D)$ are tested. These different forms of $f(D)$ are given below (where a and b are parameters of the function):

$$f_1(D) = a_1 D^{b_1} \tag{3}$$

$$f_2(D) = a_2 D^{b_2}, \quad a_2 = \mu_1 \tag{4}$$

$$f_3(D) = a_3 D^{b_3} \tag{5}$$

$$f_4(D) = \frac{a_4}{1 + \ln(D)/b_4} \tag{6}$$

$$\begin{cases} f_5(1) = \mu_1 & \text{for } D = 1 \\ f_5(D) = a_5 D^{b_5} & \text{for } D = 2, 3, \ldots, 7, \text{ and } 10 \text{ days} \end{cases} \tag{7}$$

$$f_6(D) = \frac{a_6}{1 + D/b_6} \tag{8}$$

The forms $f_1(D)$, $f_2(D)$, $f_5(D)$ and $f_6(D)$ are fitted to μ_D values using a least-squares approach by linearizing the functions through log-transformation. The form $f_4(D)$ is fitted using a least-squares approach by linearizing the function through inverse transformation. The form $f_3(D)$ is fitted following the minimization approach introduced by Javelle *et al.* (2002), i.e. the parameter b_3 is estimated by minimizing the following error criterion:

$$\varepsilon = \frac{1}{n}\frac{1}{N_D}\sum_{j=1}^{n}\sum_{i=1}^{N_D}\left[\frac{x_{D_i}(j) - \overline{x}(j)}{\overline{x}(j)}\right]^2 \tag{9}$$

where n is the number of values within each sample of H_D. $\overline{x}(j)$ is the mean ordered observed duration-scaled value given by

$$\overline{x}(j) = \frac{1}{N_D}\sum_{i=1}^{N_D} x_{D_i}(j) \tag{10}$$

where $x_{D_i}(j)$ is the duration-scaled value obtained by dividing the $H_{D_i}(j)$ value by D^{b_3}. The parameter a_3 is taken as the average of the $\overline{x}(j)$ series. The rationale behind studying six different forms of $f(D)$ is explained below. The form $f_1(D)$ has been used by Ferro and Porto (1999) and Porras and Porras (2001) for IDF analysis and is included here to investigate whether the same form could also be used for HDF analysis. However both of these studies also noticed the duration-dependent form of the $f(D)$ relation, i.e. multiple forms govern the $\mu(D) = f(D)$ relationship. This possibility of multiple relations is also investigated in the case of HDF analysis by hypothesizing the $f_5(D)$ form for modelling μ_2 to μ_{10} and μ_1 independently. The form $f_2(D)$ is designed to study whether the μ_D values for $D = 2$ to 10 days can be scaled from μ_1. The form $f_3(D)$ is exactly the same as $f_1(D)$, but differs with respect to the parameter estimation procedure. The form $f_4(D)$ is designed as an alternative to $f_6(D)$, which has been used by Javelle *et al.* (2002, 2003) for QDF analysis. It is important to mention here that no physical meanings are associated with parameters of any of the $f(D)$ forms, and the focus has been on how well the μ_D values can be modelled using intuitively appealing functional forms of $f(D)$. Site-independent comparison of $f(D)$ functions can also be performed by defining the quantities $f^*(D) = f(D)/a$, $D^* = 1/D^b$ (for $f_1(D)$ to $f_3(D)$ and $f_5(D)$), $D^* = 1/(1 + \ln D/b)$ (for $f_4(D)$) and $D^* = 1/(1 + D/b)$ (for $f_6(D)$). Quantile estimates of heatwave of any desired return period T and of any desired duration D depend upon the choice of scaling function (i.e. $f_1(D)$ to $f_6(D)$) and upon the types of quantile being scaled, i.e. dimensionless quantiles given by $g(T)$ or $H(1, T)$ quantiles or $H(\overline{x}(j), T)$ quantiles of the averaged series $\overline{x}(j)$. In the following presentation, acronyms (M1 to M5) are used for presentation convenience.

Model 1 (M1): $H(D, T) = f_1(D)g(T) = a_1 D^{b_1}(u + (\alpha/k)\{1 - [-\ln(1 - 1/T)]^k\})$

Model 2 (M2): $H(D, T) = f_2(D)H(1, T)$

Model 3 (M3): $H(D, T) = f_3(D)H(\overline{x}(j), T)$

Model 4 (M4): $H(D, T) = f_4(D)g(T)$

Model 5 (M5): $H(D, T) = f_5(D)g(T)$

Model 6 (M6): $H(D, T) = f_6(D)g(T)$

Having obtained $H(D,T)$ quantiles for a set of durations and return periods, heatwave magnitude–duration–frequency relationships can be established.

2.6. Temporal pattern of occurrences of heatwaves

The pattern of occurrences of heatwaves was studied by extracting the time series of the starting date (denoted by SD_D) of the heatwave and converting it into Julian days. For any year i, the value of SD_{1_i}

is the day of the year when the highest DMT occurred and the value of SD_{3_i} is the day of the year when highest 3-day average DMT started occurring. The time trends in the pattern of occurrences are analysed using Spearman rank correlation and Mann–Kendall non-parametric tests, along with Sen slope and linear regression slope.

3. RESULTS

3.1. Functional forms of $\mu(D)$ function

Plots of observed and modelled values of μ_D are provided in Figure 3 for one station only. Plots of the other three stations reveal a similar behaviour. In Figure 3, $f(D)$ forms are also compared with respect to extrapolation, i.e. how well they perform beyond the range used for estimating their parameters. The $\mu(D) = f(D)$ relationships are extrapolated up to $D = 30$ days. It is observed that the estimated parameters of the $f_1(D)$ and $f_3(D)$ forms are almost similar and, hence, there is an almost negligible effect of the estimation method on the estimated parameters. The $f_2(D)$ form preserves μ_1, but it generally underestimates μ_2 to μ_5 values and overestimates in the extrapolated region. The $f_4(D)$ form performs similar to $f_1(D)$ and $f_3(D)$. The $f_5(D)$ form preserves μ_1, performing very well both in the parameter estimation region and in the extrapolated region. The well-behaved nature of $f_5(D)$ supports the findings of Ferro and Porto (1999) and Porras and Porras (2001) for noticing more than one functional form for some datasets. It is obvious from Figure 3 that the $f_6(D)$ form does not follow the trend of the μ_D curve, and a similar behavior is observed for the other three stations. This form, while being suitable for QDF analysis, is not an appropriate choice to model $\mu(D) = f(D)$ relationship for heatwaves. Hence, the results of analysis using the $f_6(D)$ form are not included in the remainder of this study. For all $f(D)$ forms, coefficient of determination R^2 values are quite high (> 0.95), indicating that any of the forms studied can be assumed to model the $\mu(D) = f(D)$ relationship. R^2 is mostly used for linear relationships, and its use for non-linear relations may not be as good a tool.

Therefore, measures of relative errors and relative RMSE (rRMSE) are also used to assess fitting of $f(D)$ forms (Table III). Together for all the stations, the relative error between μ_D and the corresponding modelled value for $f_1(D)$ and $f_3(D)$ varies between -1.0% and $+1.2\%$, for $f_2(D)$ between -1.4% and $+1.2\%$, for $f_4(D)$ between -1.1% and $+1.4\%$ and for $f_5(D)$ between -0.4% and $+0.5\%$. Similarly, the rRMSE ranges between 0.59% and 0.72% for $f_1(D)$ and $f_3(D)$, between 0.73% and 0.90% for $f_2(D)$, between 0.66% and 0.81% for $f_4(D)$ and between 0.24% and 0.26% for $f_5(D)$. Based on these criteria, it is concluded that the functional form $f_5(D)$ is most appropriate for modelling the $\mu(D) = f(D)$ relationship for the

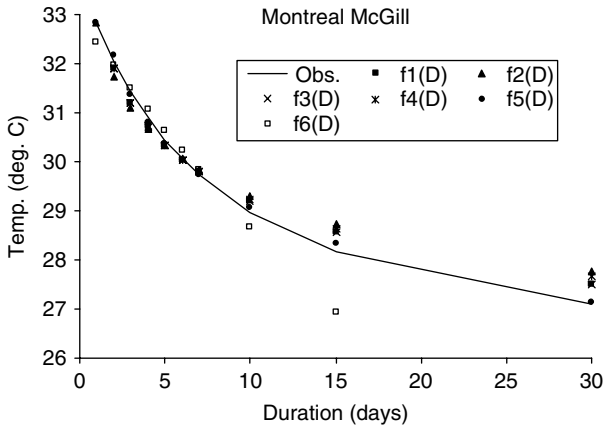

Figure 3. At-site observed mean heatwaves and fitted $\mu(D)$ functions. Mean heatwaves of 1 to 10 days' duration are used for fitting of $\mu(D)$ functions

Table III. Goodness-of-fit analysis of $\mu(D)$ functional forms

Criterion	$f_i(D)$	Lennoxville	Quebec A	Montreal McGill	Sherbrooke
Range (min. to max.) of relative error (%)	$f_1(D)$	−0.8 to 1.1	−1.0 to 1.2	−0.7 to 1.0	−0.7 to 1.0
	$f_2(D)$	−1.2 to 1.2	−1.4 to 1.2	−1.0 to 1.2	−1.0 to 1.2
	$f_3(D)$	−0.8 to 1.1	−1.0 to 1.2	−0.7 to 1.0	−0.7 to 1.0
	$f_4(D)$	−0.9 to 1.3	−1.1 to 1.4	−0.7 to 1.1	−0.7 to 1.2
	$f_5(D)$	−0.3 to 0.4	−0.4 to 0.5	−0.3 to 0.4	−0.4 to 0.4
rRMSE (%)	$f_1(D)$	0.66	0.72	0.59	0.60
	$f_2(D)$	0.82	0.90	0.73	0.74
	$f_3(D)$	0.66	0.72	0.59	0.60
	$f_4(D)$	0.74	0.81	0.66	0.69
	$f_5(D)$	0.24	0.26	0.24	0.25

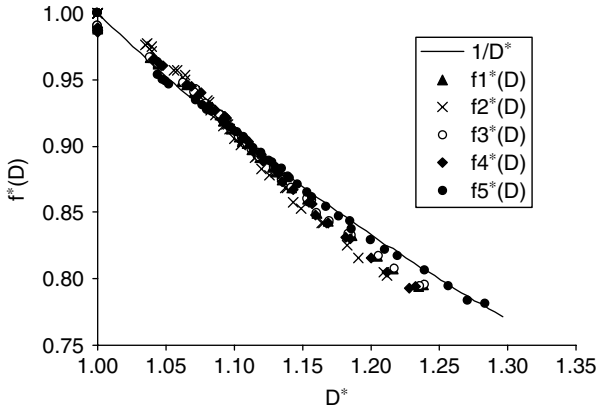

Figure 4. Site-independent comparison of fitted $\mu(D)$ functions

current dataset, but the other forms, $f_1(D)$, $f_3(D)$ and $f_4(D)$, also provide reasonable estimates of μ_D. Site-independent comparison of $f(D)$ forms, for all four stations, is provided in Figure 4, where $f^*(D)$ and D^* values are plotted. This is an alternate look at the goodness-of-fit of the $f(D)$ forms to observations. The closer the plotted points are to the $1/D^*$ line the better are the estimated values.

3.2. Development of at-site growth curve

The fitting of the GEV distribution to a selected set of H_D series is provided in Figure 5. It is obvious that the GEV distribution provides a good fit to observed heatwaves. The effect of scale is demonstrated in Figure 6, where standardized plots along with the fitted GEV distribution are provided. All standardized empirical distributions overlap each other, and the GEV distribution provides a good fit to the standardized data. Similar behaviour is observed for the other three stations. Thus, it is safe to assume that the various $H_D(T)$ distributions are linked through a scale factor. The validity of this assumption is investigated using the Hosking and Wallis (1993) homogeneity test, which was specifically devised for defining homogeneous groups of stations, by extending its definition from a group of stations to a group of hydro-meteorological series (H_D series in the present context) at a single site. A summary of this test is provided in Appendix A. It is found that Hosking and Wallis's homogeneity statistics are much less than unity at all four stations studied. This indicates that at-site standardized H_D series can be pooled together to form a homogeneous group and, hence, to derive a common at-site growth curve. The differences between at-site growth curves were examined by plotting all of them together (plot is not shown). It was observed that at-site growth curves

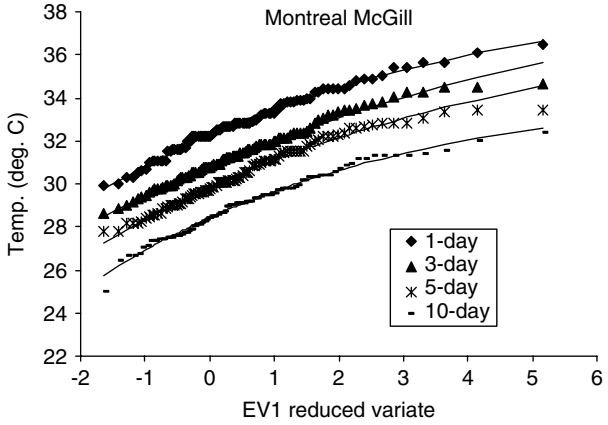

Figure 5. Heatwaves of selected durations along with fitted GEV distributions

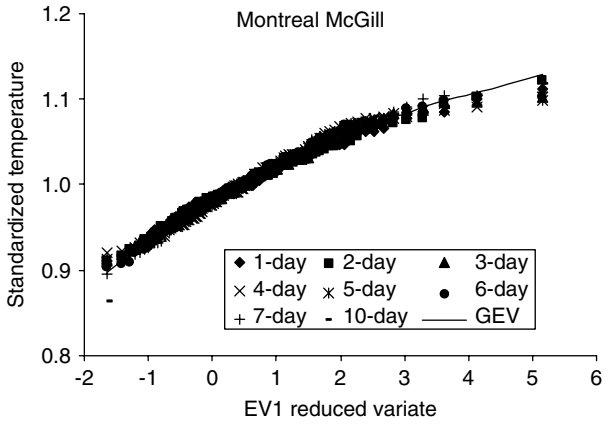

Figure 6. Plots of standardized heatwaves along with fitted GEV distribution

differ from each other in extreme upper and lower tails only. Just based on this observation, it is not possible to assume a single growth curve for these four stations without providing any physical justification with respect to weather-generating mechanisms that prevail over these stations during the summer season. This aspect is beyond the scope of the present study. It is worth mentioning that almost negligible differences were found between growth curve $g(T)$ parameters and quantiles using average standardized PWMs and PWMs of a standardized pooled sample. Thus, either approach is equally applicable. This observation indicates that non-parametric density estimation methods can also be used to obtain $g(T)$ growth curves using a standardized pooled sample.

3.3. Heatwave magnitude–duration–frequency relations

Estimated quantiles for models M1 to M5 are assessed against base-model quantiles, which are obtained by fitting the GEV distribution to each H_D series. The significance of over- or under-estimation of base quantiles can be assessed by plotting all estimated quantiles within 90%, 95% or 99% confidence bands. A 95% confidence interval is selected here. All estimated and base quantiles, along with the 95% confidence band, for 2- and 50-year return periods are plotted in Figure 7 for one station only. The 95% confidence intervals are estimated using the vector bootstrap technique (Efron and Tibshirani, 1994). The quantiles estimated with model M5 lie just inside the 95% confidence band for all heatwave series at all stations. The estimated $H(1,T)$ quantiles by models M1, M3 and M4 lie at the boundary of the upper 95% confidence interval and those for

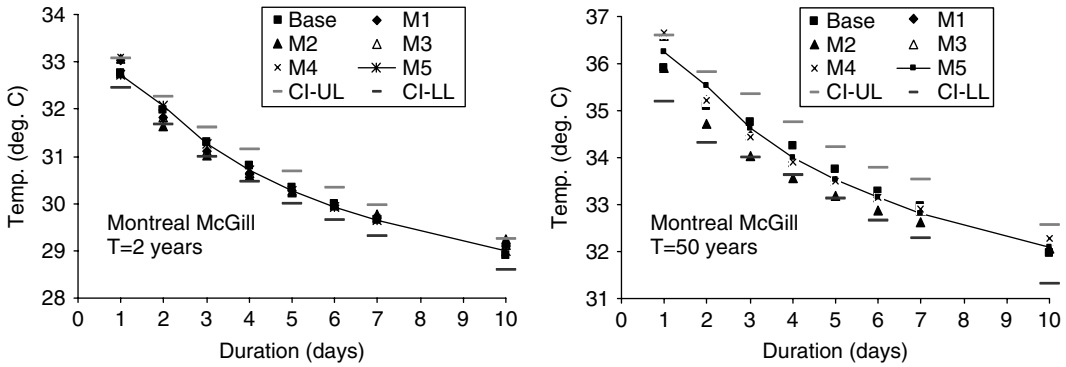

Figure 7. Plots of base and modelled quantiles along with 95% upper and lower confidence intervals. CI-LL: confidence interval lower limit; CI-UL: confidence interval upper limit

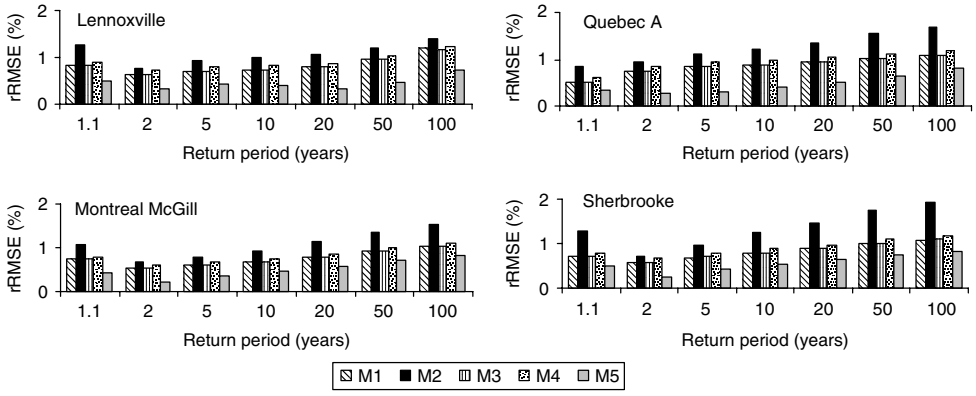

Figure 8. Plots of rRMSE for selected quantiles

other heatwave durations are inside the confidence band. The estimated $H(2,T)$ to $H(6,T)$ quantiles by model M2 lie at or near the lower boundary of the 95% confidence interval. In conclusion, the overall best estimates are provided by model M5, whereas models M1, M3 and M4 perform similarly but always overestimate $H(1,T)$ quantiles. The adequacy of the M1 to M5 models in preserving the base model quantiles is also assessed on the basis of the rRMSE criterion. Plots of rRMSE for various models are provided in Figure 8. The rRMSE is generally small for small return periods and large for large return periods. The rRMSE is the lowest for M5 at all four stations and very similar for M1 and M3. Although the rRMSE for M2 is the highest among all models at all stations, the overall performance can be considered adequate (rRMSE $<2\%$). This method is of particular interest because of the manner in which various quantiles are scaled. The useful feature of this approach is that it would permit one to derive quantiles of heatwaves of higher durations from quantiles of heatwaves of 1- day duration only.

3.4. Temporal pattern of occurrences of heatwaves

Starting dates of the first occurrence of a heatwave, along with dates of additional occurrences of the same magnitude, are extracted and studied. These dates (in terms of Julian days) are plotted in Figure 9 for one station only. It can be seen from Figure 9 that the frequency of additional occurrence is non-uniform across durations of heatwaves. In some years, first and second occurrences had occurred in association, meaning that a heatwave persisted for an additional day (e.g. if the starting dates of the first and second occurrences had been 14 August and 15 August respectively). Also, there are some years where a third occurrence occurred in association with first and second occurrences, meaning that a heatwave persisted for two additional days.

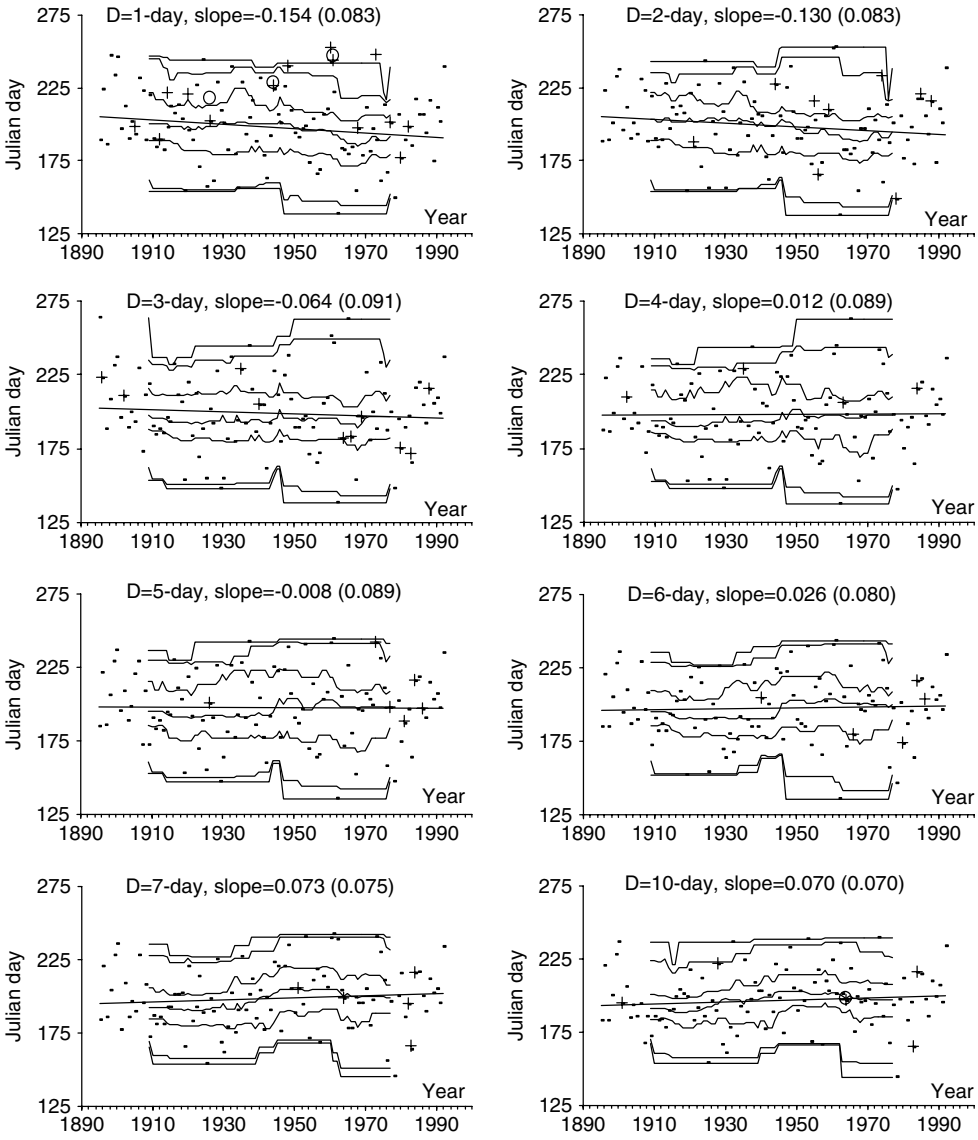

Figure 9. Dates of first (dots), second (crosses) and third occurrences (circles) of heatwaves of same magnitude in a year observed at Montreal McGill. Long-term trends (thick lines) are shown for dates of first occurrence, along with estimates of linear slope and standard error (in parentheses). Thin lines (from bottom to top) represent 30-year moving window estimates of minimum, 5P, 25P, 50P, 75P, 95P and maximum values

Because of lack of uniformity of additional occurrences, the starting dates of first occurrence (i.e. SD_D time series) are considered for further analysis. The analysis of the SD_D time series would be useful to answer questions like: Are we experiencing early summers?

3.4.1. Trend analysis of SD_D values. Summarized results of trend analysis of dates of first occurrence of heatwaves are presented in Table IV. For Lennoxville, both Sen's slope and linear regression slope methods produce negative slopes for the SD_1 to SD_4 series, positive slopes for the SD_6 to SD_{10} series and opposite values for SD_5. Taking 1 May as the reference point, a negative slope means that the starting date of a heatwave is moving towards the beginning of May and a positive slope means that the starting date is moving away from the beginning of May. Both the Sen slope and the linear regression slope produce statistically

Table IV. Linear regression and Sen slope estimates for starting dates of first occurrence of heatwaves (SD_D)[a]

Station	Method	Slope							
		$D=1$	$D=2$	$D=3$	$D=4$	$D=5$	$D=6$	$D=7$	$D=10$
Lennoxville	Linear	−0.226	−0.207	*−0.326*	−0.159	−0.007	0.080	0.089	0.075
	Sen	−0.202	−0.200	*−0.315*	−0.147	0.000	0.084	0.077	0.068
Quebec A	Linear	−0.028	−0.069	−0.070	−0.116	*−0.154*	−0.032	−0.042	−0.025
	Sen	0.000	−0.054	−0.048	−0.096	*−0.125*	−0.016	0.000	0.000
Montreal McGill	Linear	−0.154	−0.130	−0.064	0.012	−0.008	0.026	0.073	0.070
	Sen	−0.135	−0.122	−0.044	0.036	0.015	0.054	0.105	0.106
Sherbrooke	Linear	−0.106	−0.118	−0.150	−0.089	−0.041	0.062	0.049	0.082
	Sen	−0.086	−0.103	−0.148	−0.073	−0.026	0.065	0.045	0.083

[a] Note: all slope estimates are non-significant at the 5% significance level unless indicated in italics

significant results, at the 5% significance level, for the SD_3 series, and the Spearman rank correlation and Mann–Kendall tests confirm these results. Although non-significant, the values of slope estimate for the SD_1 to SD_4 series are quite high compared with the remaining series. For Quebec A, the linear regression slope method produces negative slopes for all series and Sen slope produces either negative or zero slopes. Both methods indicate that the slope for the SD_5 series is significant at the 5% significance level. Also, the Spearman rank correlation and Mann–Kendall tests confirm this result. For Montreal McGill, both the Sen slope and the linear regression slope methods produce negative slopes for the SD_1 to SD_3 series, positive slopes for SD_4 and SD_6 to SD_{10} series and opposite slopes for SD_5, while all being non-significant at the 5% significance level. Linear slope estimates, along with standard errors, are provided in Figure 9 for Montreal McGill only. For Sherbrooke, both the Sen slope and the linear regression slope methods produce negative slopes for the SD_1 to SD_5 series and positive slopes for the remaining series, but none of the tests produces statistically significant results at the 5% significance level. In general, if the significance level is raised, then more significant cases emerge.

3.4.2. Percentile plots of SD_D values. Percentile plots of SD_D values were developed to study their behaviour as a function of duration within the summer season. One such plot is provided in Figure 10, where 5th (5P), 25th (25P), 50th (50P, i.e. median), 75th (75P) and 95th (95P) percentiles and minimum and maximum values are plotted. Median heatwaves occurred during the third week of July at Montreal McGill and during the second and third weeks of July at the remaining stations. The range (maximum–minimum) of dates of occurrence decreases as the duration of heatwaves increases. For long-duration heatwaves (e.g.

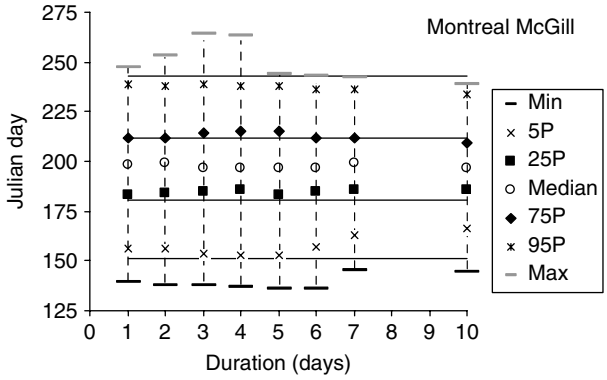

Figure 10. Plots of minimum, 5P, 25P, 50P, 75P, 95P and maximum values for dates of first occurrence of heatwaves. Horizontal lines (from bottom to top) represent end of May, June, July and August respectively

heatwaves of 7 and 10 days durations), the range of occurrence is generally concentrated over the interval from the third week of May to the last week of August at Quebec A and from the last week of May to the last week of August at the other three stations. For short-duration heatwaves (e.g. heatwaves of 1, 2 and 3 days' durations), the range of occurrence is generally concentrated between the third week of May and the fourth week of August at Quebec A and between the third week of May and the third week of September at the other three stations.

In order to extract and study time trends of various percentiles (i.e. 5P, 25P, 50P, 75P and 95P) and minimum and maximum values, a 30-year moving window technique is used and the results are displayed (at the centre of the moving window) in Figure 9 for Montreal McGill only. In general, 25P, 50P and 75P plots exhibit behaviour similar to the long-term trend, but with local fluctuations. The range is generally an increasing function of time, which is because of the overall positive slope of the maximum value and overall negative slope of the minimum value for the majority of the SD_D series. In general, this indicates that the time window during which heatwaves occur has expanded during the second half of the century.

3.4.3. Association of heatwave magnitudes and dates of occurrences. The relationship between heatwave magnitude and date of occurrence is analysed by plotting magnitude against date of occurrence (Figure 11) and estimating linear regression slopes. This relationship is significant (non-significant) at the 5% (1%) significance level for two heatwave series (i.e. H_1 and H_5) of Lennoxville, indicating that there is a tendency for intense heatwaves (i.e. heatwaves associated with extremely high temperatures) to occur later in the summer season. For the remaining heatwave series of Lennoxville, and for all heatwave series of other three stations, no significant relationship was found at the 5% significance level, indicating that there is no tendency for heatwaves of larger (smaller) magnitude to occur later (earlier) in the summer season. It is possible to determine the time window, where the majority of heatwaves are concentrated, through relative frequency estimates over non-overlapping time windows. For this purpose, each month is divided into three time windows (first window: 1 to 10 days; second window: 11 to 20 days; third window: 21 to 30 days or 21 to 31 days) and counts of occurrences are performed over these non-overlapping time intervals. The counts of occurrences are converted into relative frequencies and the results are shown in Figure 11. The majority of these distributions can be considered as having a symmetric pattern with the centre located at around the third week of July. The majority of heatwaves have occurred during the time interval from the last 10 days of June to the first 10 days of August. It is also interesting to analyse the relationship between the most extreme heatwaves and their corresponding dates of occurrences. For this purpose, the 10 most extreme heatwaves and their corresponding dates of occurrence are plotted in Figure 12. Additional occurrences of the 10th extreme heatwave are also plotted in Figure 12. It is clear from the scatter of plotted data that no relation exists between the 10 extreme heatwaves and the corresponding dates of occurrence, except that July is the most probable month for their occurrence.

4. DISCUSSION

The results obtained in the current study confirm that adopting the regional index-flood frequency analysis approach for at-site HDF analysis (similar to the approach used by Javelle *et al*. (2002, 2003) for QDF analysis) describes the behaviour of heatwaves of 1–10 days' duration in a valuable and parsimonious manner. Heatwave events are defined solely on the basis of extreme daily maximum temperatures without making reference to night-time daily minimum temperatures or to any other related meteorological variable. Alternate definitions based on combined occurrence of high daily maximum and minimum temperatures above high thresholds can also be adopted to extract heatwave time series for frequency analysis. This type of definition of heatwave has been used by Drouin and King (2004) to issue heatwave warnings in the health industry. The proposed HDF methodology can be adapted to this definition of a heatwave or to any other definition of a heatwave based on a suitably chosen combination of meteorological variables, e.g. the US National Weather Service (http://www.nws.noaa.gov/om/heat/index.shtml) define a heatwave as an extended period of high temperature above a certain threshold, which varies with location, associated with high humidity.

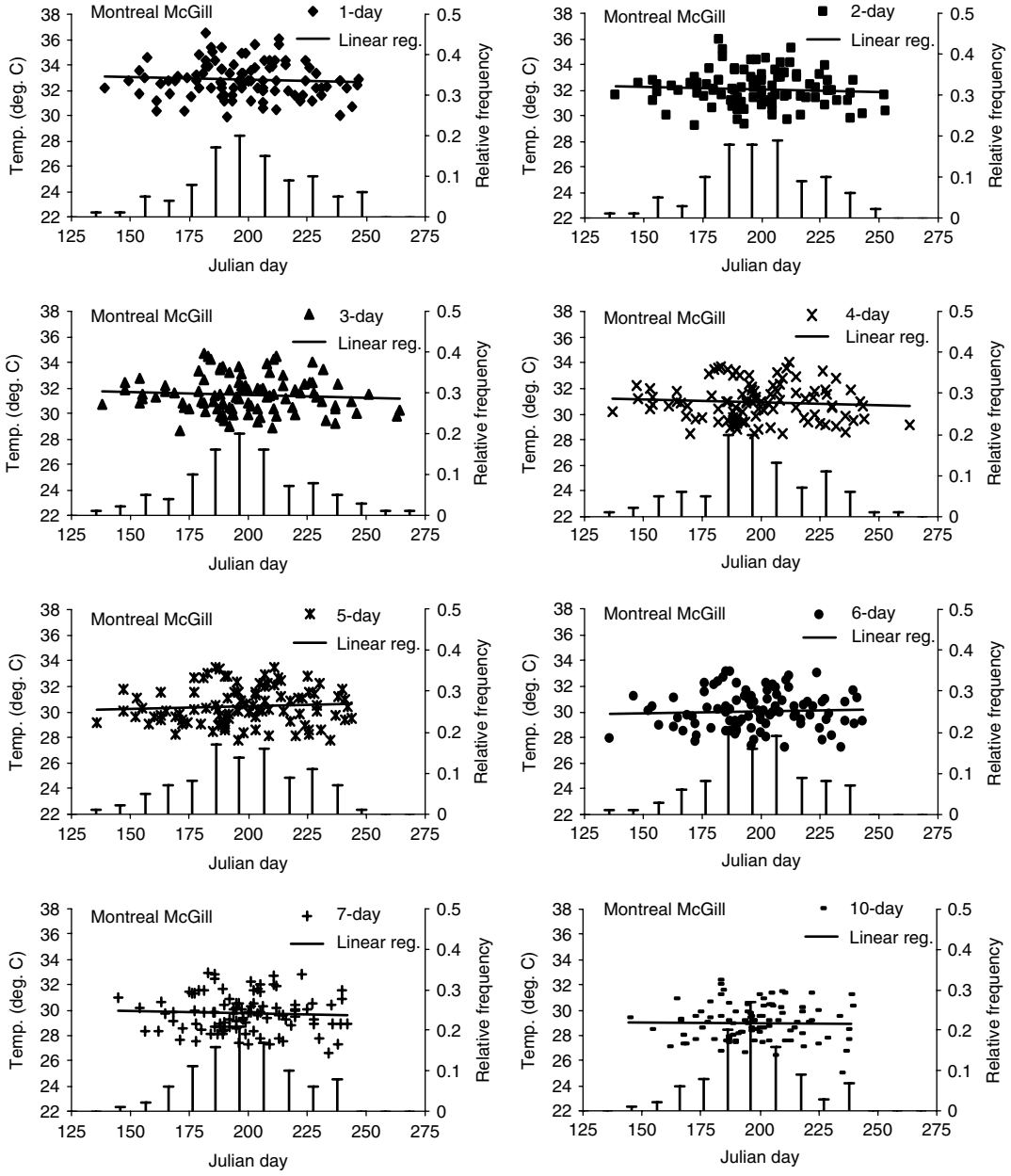

Figure 11. Relationship between heatwave magnitude and date of occurrence for heatwave durations of 1 to 10 days. Relative frequencies are shown as stem plots (vertical lines)

The choice of a distribution for frequency analysis of at-site and regional hydro-meteorological time series is a difficult task, since it can have a significant impact on low-frequency quantiles. It is difficult to determine, in an objective way, which distribution is the most appropriate. Often, the choice of a distribution is made on the basis of goodness-of-fit measures to observations or on the basis of the descriptive and predictive ability of an assumed distribution along with a parameter estimation procedure. So, whatever the choice of $g(T)$ is, the $\mu(D)$ function remains unaffected. Six different forms of $\mu(D)$ function are investigated, including the ones used for IDF and QDF analyses. It is possible to assume other forms of $\mu(D)$, e.g. a second-degree polynomial would have provided an equal fitting to the one observed with $f_5(D)$. However, it

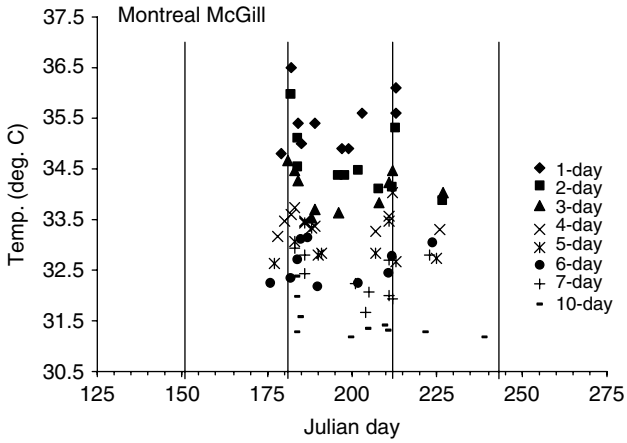

Figure 12. Ten largest heatwaves and their corresponding dates of occurrence for heatwave durations of 1 to 10 days. Vertical lines (from left to right) represent end of May, June, July and August respectively

is more interesting to investigate the suitability of $\mu(D)$ functional forms that have been used for modelling other hydro-meteorological variables in an endeavour to find a unique $\mu(D)$ functional form that is applicable across various analyses. Further, it is also interesting to adopt or formulate those $\mu(D)$ functional forms that allow site-independent comparison. It is found that the $\mu(D)$ functional form (i.e. $\mu(D) = aD^b$) that has been used for IDF analyses (Ferro and Porto, 1999; Porras and Porras, 2001) is the best form for modelling the $\mu(D)$ function for the current heatwave dataset. Therefore, it would be interesting to investigate its validity for modelling the $\mu(D)$ function in the case of QDF analysis in order to generalize the form of the $\mu(D)$ function. On the basis of limited analysis carried out in this work, the $g(T)$ function is modelled using the GEV distribution with the PWM method for parameter estimation. It is possible that a different distribution and parameter estimation procedure could emerge as the best-fitting model if an exhaustive analysis were undertaken. It is noted in Section 3 that negligible differences are found between dimensionless quantiles obtained by fitting the GEV distribution using averaged standardized PWMs and PWMs of a standardized pooled sample. Hence, the $g(T)$ function can also be modelled using a suitable non-parametric density kernel and standardized pooled sample. This would be beneficial if one wishes to bypass the procedure of model selection. It is also noted in Section 3 that heatwaves of the same magnitude have occurred more than once in some years. In order to include these additional events into HDF analysis to develop $\mu(D)$ and $g(T)$ functions, the peaks-over-threshold approach may be needed. However, because of the availability of relatively long series of observations for the current analysis, the annual maximum approach is assumed to be adequate.

Combined plots of dimensionless growth curves suggested that it is possible to assume a common growth curve for a subset of stations, i.e. the subset forms a homogeneous region. This is a promising conclusion, as it may lead the way to the definition of homogeneous regions based on hydro-meteorological and physiographic considerations. The methodology proposed in this study could then be expanded and allow for at-site and regional HDF analyses. The regional HDF analysis would be useful for determining heatwave quantiles at ungauged sites. This point remains outside the scope of the present study, and future work should address this point in the context of homogeneous regions.

Using moving samples of finite length, it has been observed that variance and low-frequency quantiles, particularly of short-duration heatwaves, are generally a decreasing function of time. Therefore, a non-stationary frequency analysis for the current heatwave dataset is highly recommended for future studies. In reality, a generalized non-stationary approach needs to be devised, and that should reduce to a stationary approach as a special case.

5. CONCLUSIONS

Motivated by the IDF and QDF analyses, this paper presents the HDF approach for modelling the at-site annual largest daily maximum temperatures of various durations to characterize heatwaves. It was found that the $\mu(D)$ function can best be modelled using a relationship of the form $\mu(D) = aD^b$ and $g(T)$, the dimensionless at-site growth curve, using the GEV distribution. The LN3 and P3 distributions could be alternate choices to model the $g(T)$ function. If one wishes to use a two-parameter distribution to model the current dataset, then the best choice is LN2. The models M1 and M3 provided almost similar results, but slightly better than M4. The approach of model M1 is simple and should be preferred over that of model M3. The scaling procedure of model M2, while preserving quantiles of 1 day heatwaves, was strictly applicable only for heatwaves of 7 and 10 days' duration. Model M5 provided the best performance, and the estimated quantiles were very similar to the base quantiles.

It was found that heatwaves of short durations (1–5 days) exhibited stronger negative slopes and those of long durations (> 5 days) exhibited moderately positive slopes. There were more significant cases at a more conservative significance level than the commonly used 5% level. It was observed that 95% of heatwave series of short duration showed a downward trend and 55% of them had a negative slope magnitude > 0.1 day year^{-1}. This implies that, over the study period, there had been a shift in the occurrence of heatwaves of short duration. The combined results of Spearman rank correlation and Mann–Kendall tests and Sen slope and linear regression slope indicated that the trends in some of the SD_D series were not just caused by sampling variations. It was also found that the median heatwave occurred during the third week of July at one station (Montreal McGill) and during the second and third weeks of July at the other three stations. Except for two cases, no association was found between heatwave magnitude and the corresponding date of occurrence, meaning that there is no general tendency for heatwaves of larger (or smaller) magnitude to occur during the early part, during the middle part, or during the later part of the summer season. However, frequency counts indicated that heatwave occurrences follow a distributional pattern with its centre located at about the third week of July. The pattern of occurrences of the 10 largest heatwaves did not exhibit any temporal behaviour within the summer season, except that July is the most probable month for their occurrence.

ACKNOWLEDGEMENTS

The financial support provided by OURANOS is gratefully acknowledged. We thank two anonymous referees for their very helpful comments.

APPENDIX A: THE HOSKING AND WALLIS (1993) HOMOGENEITY TEST

Hosking and Wallis (1993) proposed a homogeneity test to form a group of homogeneous sites. This test assesses the homogeneity of a group of sites at three different levels by focusing on three measures of dispersion for different orders of the sample L-moments ratios.

(1) A measure of the dispersion for the L-CV:

$$V_1 = \frac{\sum_{i=1}^{N} n_i (t_{2(i)} - \overline{t}_2)^2}{\sum_{i=1}^{N} n_i} \tag{A.1}$$

(2) A measure of dispersion for both the L-CV and the L-skewness coefficients in the L-CV–L-skewness space:

$$V_2 = \frac{\sum_{i=1}^{N} n_i[(t_{2(i)} - \overline{t}_2)^2 + (t_{3(i)} - \overline{t}_3)^2]^{1/2}}{\sum_{i=1}^{N} n_i} \tag{A.2}$$

(3) A measure of dispersion for both the L-skewness and the L-kurtosis coefficients in the L-skewness–L-kurtosis space:

$$V_3 = \frac{\sum_{i=1}^{N} n_i[(t_{3(i)} - \overline{t}_3)^2 + (t_{4(i)} - \overline{t}_4)^2]^{1/2}}{\sum_{i=1}^{N} n_i} \tag{A.3}$$

where $\overline{t}_2$, $\overline{t}_3$, and $\overline{t}_4$ are the group means of L-CV, L-skewness, and L-kurtosis respectively, $t_{2(i)}$, $t_{3(i)}$, $t_{4(i)}$, and n_i are the values of L-CV, L-skewness, L-kurtosis and the sample size for site i respectively, and N is the number of sites in the pooling group. The underlying concept of the test is to measure the sample variability of the L-moment ratios and compare it with the variation that would be expected in a homogeneous group. The heterogeneity measures are evaluated using

$$W_k = \frac{(V_k - \mu_{V_k})}{\sigma_{V_k}} \quad \text{for } k = 1, 2, 3 \tag{A.4}$$

where μ_{V_k} and σ_{V_k} are the expected mean and standard deviation of dispersion measures for a homogeneous group. These are estimated through repeated simulations, by generating homogeneous groups of sites having the same record lengths as those of the observed following the methodology proposed by Hosking and Wallis (1993). A group of sites may be regarded as 'acceptably homogeneous' if $W_k < 1$, 'possibly heterogeneous' if $1 \leq W_k < 2$, and 'definitely heterogeneous' if $W_k \geq 2$. Note that some of the original notation has been changed to suit the present study. This test can be applied to form a homogeneous group of hydro-meteorological data series by extending the definition of homogeneity from a group of homogeneous sites to a group of homogeneous hydro-meteorological data series, e.g. at-site heatwave series in the context of the present study.

REFERENCES

Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions of Automatic Control*. **AC-19**(16): 716–722.

Bobée B. 1975. The log-Pearson type 3 distribution and its application in hydrology. *Water Resources Research*. **11**(5): 681–689.

Bonsal BR, Zhang X, Vincent LA, Hogg WD. 2001. Characteristics of daily and extreme temperatures over Canada. *Journal of Climate*. **14**: 1959–1976.

Cunnane C. 1989. Statistical distributions for flood frequency analysis. WMO-No. 718, WMO, Geneva.

Dahmen ER, Hall MJ. 1990. *Screening of Hydrological Data. ILRI Publication* No. 49. International Institute for Land Reclamation and Improvement (ILRI): Netherlands.

Dalrymple T. 1960. Flood frequency methods. *United States Geological Survey Water-Supply Paper* **1543 A**: 11–51.

Drouin L, King N. 2004. Changements climatiques et santé publique. Poster presented at *1st Ouranos Symposium on Climate Changes*, 9–10 June, Montreal, QC, Canada.

Duenas C, Fernandez MC, Canete S, Carretero J, Liger E. 2002. Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean coast. *Science of the Total Environment* **299**: 97–113.

Efron B, Tibshirani RJ. 1994. *An Introduction to the Bootstrap*. Monographs on Applied Probability and Statistics. Chapman and Hall.

Ferro V, Porto P. 1999. Regional analysis of rainfall–depth–duration equation for south Italy. *Journal of Hydrologic Engineering* **4**(4): 326–336.

GREHYS. 1996. Presentation and review of some methods for regional flood frequency analysis. *Journal of Hydrology* **186**(1–4): 63–84.
Hosking JRM. 1990. L-moments: analysis and estimation of distributions using linear combinations of ordered statistics. *Journal of the Royal Statistical Society, Series B: Methodological* **52**(1): 105–124.
Hosking JRM, Wallis JR. 1993. Some statistics useful in regional frequency analysis. *Water Resources Research* **29**(2): 271–281.
Hosking JRM, Wallis JR. 1997. *Regional Frequency Analysis*. Cambridge University Press: Cambridge, UK.
Hosking JRM, Wallis JR, Wood EF. 1985. Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics* **27**(3): 251–261.
Huth R, Kysely J, Pokorna L. 2000. A GCM simulation of heatwaves, dry spells, and their relationships to circulation. *Climate Change* **46**: 29–60.
HYFRAN. 2003. A software package for statistical modeling. INRS-Eau, University of Quebec, Sainte-Foy, Quebec, Canada.
Javelle P, Galéa G, Grésillon JM. 2000. The flow–duration–frequency approach: former and new developments in French. *Revue des Sciences de l'Eau* **13**(3): 305–323.
Javelle P, Ouarda TBMJ, Lang M, Bobée B, Galéa G, Grésillon JM. 2002. Development of regional flood–duration–frequency curves based on the index flood method. *Journal of Hydrology* **258**(1–4): 249–259.
Javelle P, Ouarda TBMJ, Bobée B. 2003. Spring flood analysis using the flood–duration–frequency approach: application to the provinces of Quebec and Ontario, Canada. *Hydrological Processes* **17**: 3717–3736.
Katz RW, Parlang MB, Naveau P. 2002. Statistics of extremes in hydrology. *Advances in Water Resources* **25**: 1287–1304.
Kendall MG. 1975. *Rank Correlation Methods*. Charless Griffin: London.
Khaliq MN, St-Hilaire A, Ouarda TBMJ, Bobée B. 2004. Frequency analysis of daily maximum temperature in southern Quebec with a view to interpret heatwaves. INRS-ETE Research Report R-747.
Kharin V, Zwiers FW. 2000. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *Journal of Climate* **13**: 3760–3788.
Lang M, Ouarda TBMJ, Bobée B. 1999. Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* **225**: 103–117.
Madsen H, Rasmussen PF, Rosbjerg D. 1997a. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. At-site modeling. *Water Resources Research* **33**(4): 747–757.
Madsen H, Pearson CP, Rosbjerg D. 1997b. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 2. Regional modeling. *Water Resources Research* **33**(4): 759–769–.
Mutua FM. 1994. The use of Akaike information criterion in the identification of an optimum flood frequency model. *Hydrological Sciences Journal* **39**(3): 235–244.
Ouarda TBMJ, Girard C, Cavadias G, Bobée B. 2001. Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology* **254**(1–4): 157–173.
Porras Sr PJ, Porras Jr PJ. 2001. New perspective on rainfall frequency curves. *Journal of Hydrologic Engineering* **6**(1): 82–85.
Rainham DGC, Smoyer-Tomic KE. 2003. The role of air pollution in the relationship between a heat stress index and human mortality in Toronto. *Environmental Research* **93**(1): 9–19.
Salas JD, Delleur JW, Yevjevich VM, Lane WL. 1980. *Applied Modeling of Hydrologic Time Series*. Water Resources Publications: Littleton, CO.
Sen PK. 1968. Estimates of the regression coefficient based on Kendall's tau. *Journal of the Americal Statistical Association* **63**: 1379–1389.
Snedecor GW, Cochran WG. 1989. *Statistical Methods*. Macmillan: New York.
Smakhtin VU. 2001. Low flow hydrology: a review. *Journal of Hydrology* **240**: 147–186.
Strupczewski WG, Singh VP, Mitosek HT. 2001. Non-stationary approach to at-site flood frequency modeling. III. Flood analysis of Polish rivers. *Journal of Hydrology* **248**: 152–167.
Subak S, Palutikof JP, Agnew MD, Watson SJ, Bentham CG, Hulme M, McNally S, Thornes JE, Waughray D, Woods JC. 2000. The impact of anomalous weather of 1995 on the U.K. economy. *Climatic Change* **44**(1–2): 1–26.
Sveinsson OGB, Salas JD, Boes DC. 2002. Regional frequency analysis of extreme precipitation in northeastern Colorado and Fort Collins flood of 1997. *Journal of Hydrologic Engineering* **7**(1): 49–63.
UK Guardian. 2003. Global warming may be speeding up, fears scientist. *UK Guardian* 12 August.
Viessman W, Knapp JW, Lewis GL, Harbaugh TE. 1977. *Introduction to Hydrology*, second edition. Harper and Row: New York.
Vincent LA, Zhang X, Bonsal BR, Hogg WD. 2000. Homogenized daily temperatures for trend analyses in extremes over Canada. In *Proceedings of the 12th Conference on Applied Climatology*; 86–89.
Wettstein JJ, Mearns LO. 2002. The influence of North Atlantic oscillation on mean, variance and extremes of temperature in the northeastern United States and Canada. *Journal of Climate* **15**: 3586–3600.
Yu FX, Li WQ, Singh VP, Naghavi B. 1993. Estimating LP3 parameters using a combination of the method of moments and least squares. *Journal of Environmental Hydrology* **1**: 9–19.
Zhang X, Vincent LA, Hogg WD, Niitsoo A. 2000. Temperature and precipitation trends in Canada during the 20th century. *Atmosphere–Ocean* **38**: 395–429.

