FREQUENCY ANALYSIS OF DAILY MAXIMUM TEMPERATURE IN SOUTHERN QUÉBEC WITH A VIEW TO INTERPRET HEATWAVES

Research Report R-747 **August 2004**

FREQUENCY ANALYSIS OF DAILY MAXIMUM TEMPERATURE IN SOUTHERN QUÉBEC WITH A VIEW TO INTERPRET HEATWAVES

ANALYSE FREQUENTIELLE DES TEMPÉRATURES MAXIMUM JOURNALIÈRES DU SUD DU QUÉBEC POUR DES FINS D'INTERPRÉTATION DES CANICULES

Rapport présenté au Groupe Impacts et Adaptations du Consortium Ouranos

Par

M. Naveed Khaliq

André St-hilaire

Taha B.M.J. Ouarda

Bernard Bobée

AOUT 2004

Référence
Khaliq, M.N., A. St-Hilaire, T.B.M.J. Ouarda, B. Bobée. 2004. Frequency Analysis of daily maximum temperature in southern Québec with a view to interpret heatwaves. INRS-ETE, Research Report R-747, 34 pages and 3 appendices.

ISBN 2-89146-521-0

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SOMMAIRE EXÉCUTIF

Les extrêmes estivaux de température de l'air peuvent avoir un impact majeur sur la population humaine et même sur l'économie. Ces événements peuvent être caractérisés par leur amplitude et leur durée. Les canicules sont non seulement définis par des températures maximales journalières jugées extrêmes, mais aussi par une durée importante (i.e. plusieurs jours).

Le Groupe Impacts et Adaptation du consortium Ouranos a suggéré que les extrêmes de certaines variables climatiques jugées importantes pour le sud du Québec, telles que la température de l'air, soient analysées statistiquement. Ces analyses permettent de mieux caractériser la variabilité et les extrêmes de ces variables (ou des données dérivées) et permettent aussi de valider certaines méthodologies qui pourront ensuite être appliquées aux données provenant des simulations sous scénario de changements climatiques.

Les travaux de la présente étude ont donc pour objectif de valider une approche de type Intensité-durée-fréquence pour la caractérisation statistique des canicules dans le sud du Québec. Cette approche a été appliquée aux données de température de l'air maximum journalières provenant de quatre stations de la région, soit Montéral-McGill (7025280), Sherbrooke (7028124), Lennoxville (7024280) et Québec A (7016294).

La méthode utilisée consiste, en premier lieu, à calculer des séries de durée D (jours) à partir des séries de maximums journaliers. Pour ce faire, on calcule des moyennes mobiles sur une fenêtre de D jours pour toutes les stations. On extrait ensuite les maximums annuels de ces nouvelles séries de moyennes mobiles (D = 1, 2, 3, 4, 5, 6, 7 et 10 jours). L'analyse fréquentielle est alors faite séparément sur chacune des 8 séries ainsi obtenues pour chaque station.

L'objectif de l'analyse fréquentielle consiste à trouver x_T, la variable ou quantile de retour T, de probabilité au non-dépassement p, où $T = 1/(1-p)$. On utilise des observations d'événements extrêmes passés afin d'estimer les probabilités futures d'occurrence. On doit, au préalable, vérifier l'hypothèse d'indépendance et de distribution identique (i.i.d.) des observations. On cherche ensuite à estimer les quantiles x_T de période de retour T ou x_p, la probabilité au non-dépassement tel que : $x_p = \text{prob}\{X \leq x_T\} = 1 - 1/T$. Pour ce faire, on sélectionne une loi statistique à partir de plusieurs modèles disponibles. La loi sélectionnée doit ensuite être ajustée aux séries de mesures, et l'estimation d'un quantile x_T par une estimation ponctuelle $\hat{x}_T$ est alors donnée à partir de la loi ajustée.

Dans la présente étude, plusieurs lois ont été ajustées et ces ajustements ont été comparés. Par la suite, les quantiles estimés pour différentes périodes de retour T ont aussi été comparés pour différentes durées et un modèle général a été proposé.

Finalement, les dates d'occurrence des canicules ont aussi été analysées afin de confirmer ou infirmer leur stationnarité (i.e. vérifier si la date d'occurrence de la canicule varie significativement au cours des années).

Les résultats de nos travaux peuvent se résumer ainsi:

1. L'amplitude des canicules est demeurée stationnaire au cours du siècle dernier, et ce, pour les quatre stations étudiées. Les valeurs moyennes des canicules de durée 10 jours ont varié entre 27,6 °C (Sherbrooke) et 29 °C (Montréal). Les canicules de durée 3 jours ont varié entre 30.19 °C (Sherbrooke) et 31.42 °C (Montréal).
2. L'évaluation de l'adéquation des distributions statistiques ajustées aux séries de canicules ont démontré que la loi log-normale est la meilleure loi à deux paramètres, tandis que la loi GEV semble être la loi à trois paramètres qui s'ajuste le mieux pour les échantillons provenant des quatre stations.
3. Les courbes Intensité-Durée-Fréquence obtenues ont un comportement mathématique similaire pour une station donnée et pour l'ensemble des séries étudiées. Il devient donc possible de modéliser les quantiles de canicules de différentes durées à partir d'un facteur d'échelle et d'une courbe de croissance. Le facteur d'échelle établit la relation entre une valeur adimensionnelle de canicule et la période de retour T. La courbe de croissance est une distribution statistique adéquate pour tous les échantillons. La loi GEV (Generalized Extreme Value) a été utilisée comme courbe de croissance dans cette étude. Six différentes formulations mathématiques du facteur d'échelle ont été testées. Les erreurs relatives associées à chacune de ces formulations sont toutes inférieures à 1,5%. De plus, les courbes de croissance peuvent être estimées à partir de la courbe des quantiles de durée 1 jour, ce qui permet d'obtenir un modèle régional multi-durée simple à développer.
4. Une analyse des dates d'occurrence des canicules (date à laquelle commence une canicule de durée D jours) a démontré que ces occurrences sont non-stationnaires dans plusieurs cas. Ainsi, pour la station 7024280 (Lennoxville), une pente négative significative existe pour les canicules de durée 1 à 4 jour, ce qui signifie que ces événements se produisent plus tôt dans l'année à la fin du siècle qu'au début de ce dernier. Cette conclusion s'applique aussi aux canicules (toutes durées confondues) de la station 7016294 (Québec A) et pour les événements de durée 1 à 3 jours à la station 7025280 (Montréal McGill). La station de Sherbrooke est aussi caractérisée par un avancement des dates d'occurrence de canicule de durée 1 à 3 jours au cours du siècle.
   Les canicules de durée supérieure à 5 jours ont plutôt tendance à se produire plus tard, comme en témoignent les pentes positives significatives des séries de date d'occurrence aux stations 7024280 (durée 6 et 10 jours), 7025280 (durée 4, 6 et 10 jours) et 7028124 (durée 6 à 10 jours).

Le rapport présente le détail des méthodologies utilisées et des résultats obtenus.

1.0 INTRODUCTION

Extreme air temperature is known to have major effects on human populations, especially in urban areas. It has been shown in many parts of the world that extreme maximum air temperatures of long duration, the so-called heatwaves, can have adverse effects on public health (Rainham and Smoyer-Tomic, 2003; Duenas et al., 2002; Huth et al., 2000) and on the economy (Subak et al., 2000). Recently, Europe has 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 1980-1996 (Rainham and Smoyer-Tomic, 2003). Although this study stated that the impact was minimal for the population of Toronto during this period, it is possible that the impact of heatwaves on urban population in Canada may become a more serious public health problem in the future.

Bonsal et al. (2001) studied the characteristics of extreme temperatures over Canada. They found no significant trend over the course of the 20th century (1900-1998) 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 frequency analysis, 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. Peaks-over-threshold series is also called partial duration 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) or time-averaged observations (e.g. hourly rainfall intensity) or time-accumulated values (e.g. flood volumes, 24-hourly rainfall depth, degree-days) are used. Depending upon the type of analysis and application, other variables can also be defined and analysed. Haan (1977) and Cunnane (1989) document some hydro-meteorological variables which are often considered for frequency analysis. For methodological developments and application procedures of frequency analysis techniques in the areas of hydro-meteorology, the readers are referred to the reviews and comparison of techniques compiled by Cunnane (1989), GREHYS (1996), Madsen et al. (1997a, 1997b), 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/relationships for sizing various 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) in the context of regional flood frequency analysis. Sveinsson et al. (2002) and Ferro and Porto (1999) have developed regional IDF relationships for hydrological homogeneous regions following the approach of regional flood frequency analysis of Hosking and Wallis (1997). Inspired by the IDF analysis and regional flood frequency analysis, Javelle et al. (2000, 2002, 2003) have carried out flood-duration-frequency (QDF) analyses. In their studies, flood volumes were taken as accumulated flows over fixed durations and average flood flows were obtained by dividing the flood volume by the corresponding duration. Thus 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 interpret heatwaves, since these extreme events can only be fully described by associating duration with temperature magnitude. Therefore, in this work, a heatwave is defined as a run of consecutive days (say from 1-10 days) when the average daily maximum temperature is observed at the highest. This approach is referred to as heatwave-duration-frequency (HDF henceforth) analysis in the remainder of this report. The HDF concept is akin to the IDF and QDF approaches. Similarly to the QDF approach, a 2-day heatwave is defined as the largest average of two consecutive days of daily maximum temperature.

In Southern Québec, 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 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 did not take into account the duration of events.

The dataset used in this work consists of relatively long series of daily maximum temperatures of four different locations in Southern Québec for the five months period from May to September. The data length ranges from 81 to 101 years. While July and August are the peak summer months, individual hot days or a group of two or three hot days have been observed as early as mid May and as late as during 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. The heatwave durations considered range from 1-10 days. At-site HDF analysis is performed using annual maximum series, index-flood method of Dalrymple (1960) and regional flood frequency approach developed by Hosking and Wallis (1997). In QDF analysis and 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 duration is taken as the corresponding at-site mean value. Inspired from the regional flood frequency analysis and with some reasonably verifiable assumptions, the possibility of pooling standardized data across durations or adopting Hosking and Wallis (1997)'s regional frequency

analysis approach for at-site analysis, similarly as in Javelle et al. (2002, 2003) for QDF analysis, is explored and studied in order to develop a parsimonious HDF approach.

This report is organized into various sections. In section 2, objectives of the research project are presented. A methodology is presented in section 3. The third section also contains a subsection on data screening to ascertain the quality of data used in this work. Detailed results of HDF analysis and pattern of occurrences of heatwaves are presented in section 4, followed by discussion and conclusions in section 5. Appendix A contains tables and Appendix B contains figures associated with this report. A summary of the non-parametric and parametric tests used for checking quality of data and Hosking and Wallis (1993)'s homogeneity test are presented in Appendix C.

2.0 PROJECT SPECIFIC OBJECTIVES

The main objectives of this project are summarized below:

- To perform exploratory data analysis of heatwaves to explore behaviours of basic statistics (i.e. mean, variance and coefficients of variation and skewness) in order to formulate appropriate modeling hypotheses.
- To set up a basis for performing frequency analysis of heatwaves in a parsimonious manner
    - by modifying and utilizing the concepts and techniques developed for frequency analysis of other hydro-meteorological variables, and
    - by selecting a suitable statistical model, on the basis of some model selection criteria, to describe the behaviour of D-day heatwave, where D is an integer number of days.
- To perform model validation.
- To study time-trends both in the magnitude and pattern of occurrences of heatwaves.
- To present and discuss results obtained and to identify future research directions.

3.0 METHODOLOGY

3.1 DATA EXTRACTION

As described earlier in the introduction section, a heatwave is a run of consecutive days with highest average daily maximum temperature (DMT). 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 May-September period (e.g. $H_{3_i} = 30\,°C$ if the observed daily maxima has been 30 °C during 3 consecutive days in year i). Similarly, μ_D and σ_D are the mean and standard deviation of H_D series. For performing HDF analysis, DMT data observed during May-September at four stations in Southern Québec, Canada is used. Station detail is given in Table A1.The length of data series ranges between 81 and 101 years. The peak summer period (i.e. July-August) is given more importance when deciding the length of record to be considered for 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. As par definition of

heatwave, the temperature data for various durations needs to be computed from the DMT series. Temperature time series for various durations are extracted from DMT series using a moving average with a window of width D-day (i.e. D = 2-, 3-, 4-, 5-, 6-, 7- and 10-day). The procedure is explained in Figure B1. H_D extreme values can thus be extracted from the original DMT series to form new time series using either annual maximum or peaks-over-threshold approach. In this work, the annual maximum approach is used to extract H_D series. Thus for every station, 8 samples (one for each duration) of H_D series are obtained which constitute the basic set of data for the present analysis. Occasionally, reference has been made to H_D series for D = 15- and 30-day to verify some of the assumptions used in the analysis. However, these two H_D series are not subjected to detailed analysis.

3.2 QUALITY OF H_D SERIES

For a successful frequency analysis, it is imperative to check data quality through graphical plots and statistical tests of significance. In graphical plots of sequential data, one can visually observe the presence of trends and jumps and the statistical tests help in explaining whether the trends and jumps, if any, are real or just because of sampling variability. There are two general approaches for testing quality of data: parametric and non-parametric. Parametric approaches make certain assumptions about the nature of data and non-parametric approaches do not make any assumptions about the underlying statistical distribution of the data. By making no assumptions about the distribution of the data, non-parametric tests are more widely applicable than parametric tests, which often require normality in the data. While more widely applicable, the trade off is that non-parametric tests are less powerful than parametric tests.

Two non-parametric tests, Spearman's rank correlation and Mann-Kendall tests are selected to verify the absence of trend and the independence of a time series (i.e. there is no correlation between the order in which the data have been collected and the increase or decrease in the magnitude of those data). Nevertheless, the independence of a time series depends on both the level of aggregation and separation in time of the data points. These tests are also recommended by WMO (WMO, 1966) for checking quality of hydro-meteorological data. To further strengthen the results of these two tests, Sen's (1968) non-parametric method is used to obtain an estimate of the magnitude of a monotonic trend. In addition to these non-parametric tests, ordinary linear regression slope and correlograms are used to verify absence of trend and independence of observations. Also, the t-test for stability of mean and the F-test for stability of variance are used to study stationarity of heatwave observations. A short description of all the above indicated tests is given in Appendix C.

Results of Spearman's rank correlation and Mann-Kendall tests are provided in Table A2. It is interesting to note that none of the H_D series for all the four stations exhibit any evidence of trend at 5% significance level. Results of Sen's slope are provided in Table A3. Again, the 95% confidence interval around the slope estimate does not support the presence of a monotonic trend in any of the H_D series. Estimates of linear slope, along with 95% confidence intervals are provided in Table A3. For obtaining linear slopes, linear regression was performed between values of H_D series and time points. Plots of H_D time series, along with linear slope estimates, are provided in Figure B2. Also, the regression slope does not support the presence of a linear trend

at 5% significance level. Based on the results of all these tests, it can be concluded that H_D data for all four stations are free of trends and jumps.

Computed values of lag-1 autocorrelation coefficient, r_1, and upper and lower limits of confidence interval, at 5% significance level, are provided in Table A4. None of the r_1 values is significant at 5% significance level. The correlograms for H_1 to H_{10} series are provided in Figure B3. In this figure, autocorrelations up to lag 20 along with 90%, 95% and 99% confidence intervals are plotted. For a quick inference, the numbers of r_k values, which fall outside the confidence intervals, are given in Table A5. None of the r_k values falls outside the 95% confidence interval for H_1 to H_{10} series for station 7024280. For station 7025280, there is one r_k value that falls outside the 95% confidence band for H_1 series. For station 7028124, there is one r_k value that falls outside the 95% confidence band for each of H_1 and H_2 series. For station 7016294, there are two r_k values outside the 95% band for H_1 series and one r_k value outside the 95% confidence band for each of the remaining series. Based on the 95% confidence bands, the only heatwave series that could be non-independent is H_1 at station 7016294. In general, the above indicated significant r_k values are not extremely large because nearly all of them fall near the boundaries of confidence interval. Furthermore, the occurrences of significant r_k values are within the expected variability, i.e. one out of 20 values is expected to fall outside the 95% confidence band. At a more conservative confidence interval (i.e. 90%), the occurrences of significant r_k values are within the expected variability for heatwave data of stations 7024280, 7025280 and 7028124. However, for heatwave data of station 7016294, there are two time series (i.e. H_3 and H_{10}) which could be regarded as non-independent. In general, it seems that the quality of heatwave data of station 7016294 is relatively poor in comparison to other three stations.

The parametric tests used for testing stability/equality of variance and mean are the F-test and the t-test, respectively. These tests are based on the normality assumption. The requirement for normality is much less stringent for the t-test than for the F-test. One can apply the F-test for stability of variance and the t-test for stability of mean to data that belong to any frequency distribution, but the length of the sub-samples should be equal, or approximately so, if the distribution is skewed (Dahmen and Hall, 1990). Hence to apply these tests, the time series needs to be divided into two or three split, non-overlapping, sub-samples of approximately equal length. The heatwave time series are mildly skewed, so it is expected that the F-test will give an acceptable indication of stability of variance and the t-test for stability of mean. The t-test for stability of mean assumes that the variances of the two sub-samples are statistically similar. The variances, can, however, differ only because of sampling variability. The test for stability of variance is usually done first because instability of the variance implies that the time series is not stationary and is not suitable for a modeling exercise where one assumes stationarity of observations.

The results of F-test for stability of variance are provided in Table A6a for the case of two sub-samples and in Table A6b for the case of three sub-samples. None of the heatwave series at all stations indicates instability of variance at 5% significance level (the same is true even at 10% significance level) when the time series were divided into two non-overlapping sub-samples of

approximately equal length. However, when using three sub-samples, the variance of H_1 series, at station 7028124, is found instable at 5% significance level with one marginal case of H_1 series at station 7016294. The results of t-test for stability of mean are provided in Table A7a for the case of two sub-samples and in Table A7b for the case of three sub-samples. None of the heatwave series at all stations indicates instability of mean (at 5% significance level) when the time series are divided into two non-overlapping sub-samples of approximately equal length. There are two significant cases at 10% significance level for station 7016294. When using three sub-samples, none of the time series showed any sign of instability of mean at 5% and 10% significance levels. Combined results of both F-test and t-test are provided in Tables A8a and A8b, respectively. Based on the results of F-test for stability of variance and the t-test for stability of mean, the H_1 series at station 7028124 is probably non-stationary at 5% significance level. The number of non-stationary cases increases when using a more conservative significance level, e.g. 10%.

The results of statistical tests presented in this section indicate that there is one data series (H_1 series at station 7016294) which could be non-independent at 5% significance level. However, the dependence is not very strong because firstly this non-independence vanishes at 1% significance level and secondly the same series becomes independent at 10% significance level because the number of r_k values outside the 90% confidence interval is within allowed statistical variability. Furthermore, the values of the significant autocorrelation coefficients are within 6-22% of the 95% confidence intervals. Another data series (H_1 series at station 7028124) is found non-stationary in the variance but stationary in the mean. However, the non-stationarity is dependent on the number of non-overlapping sub-samples the series is divided into in order to perform statistical tests for stability of variance. The non-stationarity vanishes when using two non-overlapping sub-samples. Thus, it is conjectured that inclusion of these two series into frequency analysis will have 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.

3.3 Basic Statistics of H_D Series

Basic statistics (i.e. mean, standard deviation, and coefficients of variation and skewness) of all H_D series are provided in Table A9 and plotted in Figure B4. It is useful to plot these statistics as a function of dependent variable to study their mutual variability and to formulate appropriate hypotheses. Plots of mean values in Figure B4 indicate that the mean μ_D is a decreasing function of duration and it decreases in a parabolic fashion. However, standard deviation σ_D does not depict a regular increasing/decreasing pattern as a function of duration. It fluctuates around a mean value randomly except for station 7016294 where it shows a regular decreasing trend after D=2-day duration. Small values for the range of σ_D for stations 7024280, 7025280 and 7028124 (i.e. 0.12, 0.09 and 0.12, respectively) and increase in coefficient of variation as duration increases (Figure B4) suggest that the assumption of duration-independent value of σ_D would be appropriate for these stations. For station 7016294, where a decreasing trend after D=2-day duration is realized, the assumption of duration-independent value of σ_D is appropriate for a subset of durations. However, overall small range of σ_D (i.e. 0.20) for this station also suggests that the assumption of duration-independent value of standard deviation may be applicable.

Both Table A9 and Figure B4 indicate that H_D series are just moderately skewed, i.e. coefficient of skewness ranges between 0.39 and 0.84 for station 7024280, between 0.12 and 0.37 for station 7016294, between 0.09 and 0.39 for station 7025280 and between 0.23 and 0.53 for station 7028124. One could expect that H_{10} series should present minimum value of coefficient of skewness because each value in this series is an average of 10 consecutive values. The minimum value of skewness is observed for H_4 series at station 7024280 and 7028124, for H_3 series at station 7025280, and for H_{10} series at station 7025280. A regular decrease in skewness as duration increases can probably be hypothesized in the case of station 7028124. However, small range of coefficient of skewness indicates that it could be possible to assume a duration-independent value of skewness for all heatwave series at each station. The above presented behaviours of basic statistics, 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.

3.4 SELECTION OF A STATISTICAL DISTRIBUTION FOR H_D SERIES

Extreme value plots of heatwaves are provided in Figure B5 for the four studied stations. Extreme values are plotted against Extreme Value type 1 (EV1) reduced variate using Gringorton plotting position formula: $p = (i - 0.44)/(n + 0.12)$, where i is the rank in ascending order and n is the number of observations. Such plots help one to decide about the suitability of the EV1 (or Gumbel) distribution for modelling purposes. A straight line plot would suggest that the EV1 distribution is a strong candidate model. The plots of Figure B5 show moderate downward curvature except for station 7024280, for which plots tend to follow straight lines. Another obvious observation is that the plots of extreme values of various durations demonstrate a parallel behaviour, except for few observations at both tails. This behaviour suggests that scale is a dominant factor and if the effect of scale is removed then all the plots could overlap each other. The effect of scale is studied in a subsequent section.

In order to model values of H_D series, some of the 2-parameter and 3-parameter statistical distributions are analyzed and no attempt has been made to perform an exhaustive analysis in this respect. The 2-parameter distributions include the Normal (NOR), Lognormal (LN2), and Extreme Value Type I (EV1). The 3-parameter distributions include the Generalized Extreme Value (GEV), Lognormal (LN3), Pearson Type III (P3) and Log Pearson Type III (LP3). For 2-parameter distributions, only maximum likelihood (ML) estimators are used. For the LN3 and P3, both ML and ordinary moment (MOM) estimators are used. For the GEV, both ML and probability weighted moments (PWM) estimators are used. LP3 distribution is fitted using the procedure described by Bobée (1975). All of the parameter estimations are performed using HYFRAN (2003) software. The choice of a best fitting model is performed on the basis of three model selection criteria, i.e. Akaike Information Criterion (AIC), Chi-squared test statistic and Root Mean Square Error (RMSE).

The Chi-squared goodness-of-fit statistic is well documented in most of the statistical books (e.g. Snedecor and Cochran, 1989), and is not repeated here. This test is very sensitive to the number of classes in which the empirical distribution is divided into. It is possible that if the test statistic is significant with n_1 class intervals then it may not be significant with n_2 (where n_2 being greater

than or less than n_1) class intervals. For calculating class intervals, HYFRAN (2003) software is used and no attempt has been made to analyze the effect of class intervals because the chi-squared test statistic is not used to perform a test of significance rather its magnitude is used for ranking purposes to distinguish a statistical model that gives the smallest value of the test statistic in a manner similar to AIC and RMSE criteria. Thus among a group of competing models, the one that gives minimum value of the Chi-squared test statistic is the best fitting model on the basis of Chi-squared test statistic criterion and the one that gives the minimum value of the AIC is the best fitting model on the basis of AIC. Similar argument is applicable for RMSE criterion.

Various forms of AIC have been used in literature. The form of AIC varies with respect to how the test statistic is penalized for the estimated parameters and record length. The form of AIC used in this work is given below:

$$\mathrm{AIC} = -2\ln(\mathrm{ML}) + 2m \tag{3.1}$$

where ML denotes the maximum likelihood and m denotes the number of fitted parameters. The AIC is employed to select the best fitting model. The best fitting model is the one with the smallest value of AIC. Although the AIC goodness-of-fit criterion is constrained to maximum likelihood estimation, it has been used regardless of the estimation method (Strupczewski et al., 2001; Mutua, 1994).

The use of RMSE for model selection is demonstrated by Cunnane (1989) and a variant of it by Yu et al. (1993). This criterion is evaluated as follows:

$$\mathrm{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{X}_i)^2\right]^{1/2} \tag{3.2}$$

where X_i and $\hat{X}_i$ are respectively the observed and estimated quantiles corresponding to ith plotting position and n is the record length. In order to estimate the value of this criterion, one needs to select an appropriate plotting position formula. A number of plotting position formulae have been proposed in the literature for various distributions. Cunnane (1989) analyzed that the choice of a best fitting distribution using RMSE criterion is independent of the plotting position formula and method of parameter estimation (i.e. maximum likelihood and moments). Cunnane (1978) proposed a single distribution free formula as a reasonable compromise for mildly skewed data and same has been used to compute RMSE in the present work. The current dataset is mildly skewed and hence, the choice of Cunnane plotting position formula seems reasonable. A model that gives smallest value of RMSE is considered the best fitting model among a group of competing models.

For the 2-parameter distributions, the values of AIC, Chi-squared and RMSE are given in Table A10. All three criteria suggest that LN2 is the overall best fitting distribution. However, EV1 is another equally likely choice for station 7024280. The choice of best fitting distribution is made as follows: for each record, each distribution is assigned a rank between 1 and 3, rank 1 for the best fitting distribution (e.g. the distribution having lowest value of Chi-squared test statistic).

For each distribution, the ranks are summed over all 8-durations and these totals of ranks are used to choose a best fitting distribution. Table A11 presents the values of AIC and Chi-squared criteria for 3-parameter distributions. As commented earlier, the value of Chi-squared test statistic is very sensitive to the number of class intervals. Hence, the Chi-squared test statistic criterion is not used to distinguish best fitting 3-parameter distributions. Based on AIC criterion, first four best fitting distributions are GEV-ML, LN3-ML, P3-ML and GEV-PWM. The RMSE for these four best fitting distributions are given in Table A12. According to RMSE, the order of best fitting distribution for stations 7024280 and 7016294 is GEV-PWM, GEV-ML, LN3-ML, and P3-ML, with the last two being at the same rank. For station 7025280, the order is GEV-PWM, GEV-ML, LN3-ML, and P3-ML. For station 7028124, the order is GEV-PWM, GEV-ML, LN3-ML, and P3-ML, with the first two of the same rank. Combined results of AIC and RMSE for 2-parameter best fitting distribution (i.e. LN2) and 3-parameter distributions are given in Table A13. Based on AIC, overall best fitting distribution is the GEV-ML and based on RMSE the best fitting distribution is the GEV-PWM.

The L-moment ratio diagram for H_D series is provided in Figure B6. The definitions and methods of computing L-moments, which are linear combination of order statistics, 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. In this diagram, two parameter distributions are defined as points and three parameter distributions are defined as curves. The L-moment ratio diagram given in Figure B6 suggest that the points of some of the H_D series are located more close to Generalized Pareto distribution (GPD) curve as compared to curves of LN3, P3 and GEV distributions. However, none of those series which fall near GPD curve could be fitted by GPD when its parameters were estimated by method of moments and probability weighted moments because the upper bound of the distribution was always smaller than the maximum value of the observed sample. The points of H_D series of 7016294 and 7025280 surround LN3, P3 and GEV curves indicating that any one of them could be a suitable choice for these two stations. However, this behavior is less obvious for the remaining two stations particularly in the case of 7024280. The L-moment ratio diagram is provided here for the sake of completeness rather being used as a model selection criterion.

Based on the above presented results and discussion, the GEV-PWM distribution is selected for modeling H_D series for all four stations. One obvious conclusion is that the Normal distribution that has been used by Wettstein and Mearns (2002) to model daily maximum temperature series cannot be selected as the best fitting distribution on the basis of AIC and RMSE criteria for the current dataset.

3.5 HEATWAVE MAGNITUDE-DURATION-FREQUENCY (HDF) APPROACH

The HDF analysis presented here is restricted to D=1-, 2-, ..., 7-, and 10-day durations. It is possible to extend the analysis beyond 10-day duration but in that case the usefulness of heatwave may become meaningless. The aim of the HDF modeling is to describe the H_D distributions in a parsimonious manner by adopting the methodology of regional index-flood frequency analysis. Similar approach has been adopted by Javelle et al. (2002, 2003) for QDF analysis. 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 while the growth curve is assumed to be valid for the entire group of H_D distributions. This can be explained by the following relationship:

$$H(D,T) = \mu(D)g(T) \tag{3.3}$$

where $H(D,T)$ is the modeled value of $H_D(T)$ and $\mu(D)$ is the modeled value of μ_D and $g(T)$ is a dimensionless parent fitted distribution, with a mean of 1. 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 behaviors of basic statistics presented in section 3.3 and further verified in section 4.2. Based on the analysis presented in the above section, $g(T)$ is taken to be GEV distribution with probability distribution function given by

$$f_Y(y;u,\alpha,k) = (1/\alpha)[1-k(y-u)/\alpha]^{-1+1/k} \times \exp\{-[1-k(y-u)/\alpha]^{1/k}\} \tag{3.4}$$

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 return level which is exceeded once every T years on average, is given by

$$Y(T) = u + (\alpha/k)[1-(-\ln(1-1/T))^k]. \tag{3.5}$$

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 standardized PWMs averaged across durations. The later approach, which is well documented by Hosking and Wallis (1997) for regional frequency analysis, is adopted here. The averaged standardized probability weighted moments are defined below:

$$b(r) = \frac{1}{N_D} \sum_{\substack{D=1 \\ D\neq 8,9}}^{D=10} \frac{b_D(r)}{b_D(1)} \tag{3.6}$$

where $b(r)$ is the rth PWM of $g(T)$, $b_D(r)$ the rth PWM of H_D distribution, $b_D(1)$ is the first PWM of 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 modeled as a function of D similarly as in Javelle et al. (2002, 2003) for QDF analysis and as in Ferro and Porto (1999) for regional IDF analysis. The general form of this relationship is given by

$$\mu(D) = f(D). \tag{3.7}$$

Six different forms of $f(D)$ are assumed and analyzed. These different forms of $f(D)$ are given below (where a and b are parameters of the function):

$$f_1(D) = a_1 D^{b_1} \tag{3.8}$$

$$f_2(D) = a_2 D^{b_2}, a_2 = \mu_1 \tag{3.9}$$

$$f_3(D) = a_3 D^{b_3} \tag{3.10}$$

$$f_4(D) = \frac{a_4}{1 + \ln(D)/b_4} \tag{3.11}$$

$$\begin{cases} f_5(1) = \mu_1 & \text{for } D = 1 \\ f_5(D) = a_5 D^{b_5} & \text{for } D = 2\text{-},3\text{-},\ldots,7\text{-}, \text{and } 10\text{-day} \end{cases} \tag{3.12}$$

$$f_6(D) = \frac{a_6}{1 + D/b_6} \tag{3.13}$$

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 and the form $f_3(D)$ is fitted following the minimization approach introduced by Javelle et al. (2002). The rational behind studying six functional forms of $f(D)$ is explained below.

The functional 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 if same functional form could also be used for HDF analysis. However both of these studies also noticed the duration dependent functional form of $f(D)$ relation, i.e. multiple functional forms govern the $\mu(D) = f(D)$ relationship. This possibility of multiple relations is also investigated in the case of HDF analysis by hypothesizing $f_5(D)$ functional form for modeling μ_2 to μ_{10} and μ_1 independently. Thus the functional form $f_5(D)$ has total of three parameters as opposed to all other forms. The functional form $f_2(D)$ is designed to study if the μ_D values for D=2- to 10-day can be scaled from μ_1. The parameter b_2 of this form can be estimated through least squares approach first by dividing each μ_D value by μ_1, performing log-transformation and setting constant to zero. The functional form $f_3(D)$ is exactly same as $f_1(D)$ but differs with respect to the fitting procedure. The $f_1(D)$ functional form is fitted through least squares method and $f_3(D)$ is fitted using the method of Javelle et al. (2002), i.e. the parameter b_3 is estimated by minimizing the following error criterion:

$$\varepsilon = \frac{1}{n}\frac{1}{N_D}\sum_{j=1}^{n}\sum_{i=1}^{N_D}\left[\frac{x_{D_i}(j)-\bar{x}(j)}{\bar{x}(j)}\right]^2 \tag{3.14}$$

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

$$\bar{x}(j) = \frac{1}{N_D}\sum_{i=1}^{N_D} x_{D_i}(j) \tag{3.15}$$

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 $\bar{x}(j)$ series. The above equation implies that $\bar{x}(j)$ is computed by taking average of duration scaled ordered values across durations. This implies that sample size has to be same for each H_D series. If the sample size is different for each H_D series (which could result in case of peaks-over-threshold approach), then this approach cannot be applied and recourse has to be made to an alternate method given in Javelle et al. (2003) or to some other appropriately designed methods. The above described parameter estimation procedure is included here to find the effect of parameter estimation procedure on the estimated parameters. It is necessary to mention here that Javelle et al. (2002) used $\bar{x}(j)$ series to estimate quantiles of the averaged series for scaling purposes to estimate quantiles corresponding to various durations. Their approach for estimating $H(D,T)$ quantiles is retained here for comparison purposes. The functional 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. The functional form $f_6(D)$ is included here for comparison purposes and to investigate if same could be used for HDF analysis. It is important to mention here that no physical meanings are associated with parameters of any of the functional form $f(D)$ and focus has been on how well the μ_D values can be modeled.

Site independent comparison of $f(D)$ functions can also be performed by defining the following quantities:

$$f^*(D) = f(D)/a \text{ and} \tag{3.16}$$

$$D^* = 1/D^b \text{ and } D^* = 1/(1+\ln D/b)\,. \tag{3.17}$$

The quantile estimates of heatwave of any desired return period T and any desired duration D depend upon the choice of scaling function (i.e. $f_1(D)$ to $f_6(D)$) and upon the type of quantiles being scaled, i.e. dimensionless quantiles given by $g(T)$ or $H(1,T)$ quantiles or quantiles of averaged series $\bar{x}(j)$. For the sake of presentation convenience, these combinations are given acronyms as explained below:

Model 1 (M1) – The $H(D,T)$ quantiles for this case are estimated using $f_1(D)$ functional form of $\mu(D)$ and $g(T)$ dimensionless quantiles, i.e.

$$H(D,T) = f_1(D) \times g(T) = a_1 D^{b_1} \times \left\{ u + \frac{\alpha}{k}\left[1 - (-\ln(1-1/T))^k\right]\right\}. \tag{3.18}$$

Model 2 (M2) – The $H(D,T)$ quantiles for this case are scaled from $H(1,T)$ quantiles using $f_2(D)$ functional form of $\mu(D)$. $H(1,T)$ quantiles are estimated by fitting GEV distribution to H_1 series using PWMs.

Model 3 (M3) – The $H(D,T)$ quantiles for this case are estimated by scaling the quantiles obtained by fitting GEV distribution to $\bar{x}(j)$ series with $f_3(D)$ functional form of $\mu(D)$.

Model 4 (M4) – The $H(D,T)$ quantiles for this case are estimated similarly as in the case of M1 but using $f_4(D)$ functional form of $\mu(D)$.

Model 5 (M5) – The $H(D,T)$ quantiles for this case are estimated similarly as in the case of M1 but using $f_5(D)$ functional form of $\mu(D)$.

Model 6 (M6) – The $H(D,T)$ quantiles for this case are estimated similarly as in the case of M1 but using $f_6(D)$ functional form of $\mu(D)$.

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

3.6 PATTERN OF OCCURRENCES OF HEATWAVES

The pattern of occurrences of heatwaves (i.e. values of H_D series) is studied by extracting the time series of starting date (denoted by SD_D) of 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's rank correlation and Mann-Kendall non-parametric tests. The results of these tests are further supported using Sen' slope and linear regression slope estimates.

4.0 RESULTS

The HDF approach presented above is applied to DMT series observed at 4 stations in Quebec and results for these stations are presented with respect to validation of HDF approach, i.e. fitting of functional forms of $\mu(D)$ function, development of a common at-site growth curve, heatwave

magnitude-duration-relationships and their adequacy in preserving base model quantiles. Finally, the pattern of occurrences of heatwaves is presented.

4.1 FUNCTIONAL FORMS OF $\mu(D)$ FUNCTION

The estimated parameters of all functional forms of $\mu(D)$ are given in Table A14. The R^2 values are also provided in this table when the parameters are estimated through least squares approach. The Relative Root Mean Square Error (rRMSE) values are also provided as an alternate measure for goodness-of-fit of the functional form. It is interesting to note that R^2 values are quite high (> 0.95) for all functional forms indicating that any one functional form can be assumed to model $\mu(D) = f(D)$ relationship. To visually compare this relationship, the plots of observed and modeled values of μ_D are provided in Figure B7. In this figure, the functional forms are also compared with respect to extrapolation, i.e. how well they perform beyond the range used for estimating their parameters. The relationships are extrapolated up to D=30 days. It is obvious from Figure B7 that the functional form $f_6(D)$ does not follow the trend of μ_D curve. Hence, this form is not an appropriate choice to model $\mu(D) = f(D)$ relationship for heatwaves and the results of analysis using this form are not included in the remainder of this work. The estimated parameters of $f_1(D)$ and $f_3(D)$ are almost similar and hence they perform similarly. There is almost negligible effect of estimation method on the estimated parameters. The $f_2(D)$ functional form preserves μ_1 but generally underestimates μ_2 to μ_5 values and overestimates in the extrapolated region. The $f_4(D)$ functional form performs almost similarly as $f_1(D)$ and $f_3(D)$. The functional form $f_5(D)$ preserves μ_1, performs very well both in the parameter estimation region and in the extrapolated region. Although μ_{15} and μ_{30} are not used for fitting $f_5(D)$ function, it provides reasonably good estimates of those values. The well behaved nature of $f_5(D)$ can be partly explained as follows. The H_D series can be divided into two domains i.e. one without any averaging (i.e. H_1 series) and another with averaging (i.e. H_2 to H_{30} series). It seems that these two domains need to be modeled independently – this observation supports the findings of Ferro and Porto (1999) and Porras and Porras (2001) for noticing more than one functional form for some datasets.

Site independent comparison of $f(D)$ functions is provided in Figure B8 where $f^*(D)$ and D^* values are plotted. The closer the plotted points are to the $1/D^*$ line the better are the estimated values. The extrapolated region for the $f_5(D)$ case is also shown in this figure. The lower part of the curve presents the extrapolated region where the plotted points are very close to $1/D^*$ line.

From the results presented above, it is obvious that the $f_5(D)$ functional form performs very well. For all the stations, the relative error between μ_D and corresponding modeled 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%. The rRMSE given in Table A14 is lowest for $f_5(D)$. Hence, it is concluded that the functional form $f_5(D)$ is more

appropriate for modeling $\mu(D) = f(D)$ relationship for the current dataset while other forms, $f_1(D)$, $f_3(D)$ and $f_4(D)$, also provide reasonable estimates of μ_D. All functional forms of $\mu(D)$ are considered for the remaining analysis presented here.

4.2 Development of At-Site Growth Curve

The fitting of GEV distribution to a selected set of H_D series is provided in Figure B9. It is obvious that the GEV distribution provides a good fit to observed data at all four stations. The quantitative measures of goodness-of-fit have already been presented and discussed in a previous section on model selection. The quantiles $H_D(T)$ estimated by fitting the GEV distribution to each individual H_D series are referred to as base model quantiles for comparison purposes while assessing the validity of HDF modelling approach.

It is hypothesized in the previous section that if the effect of scale is removed by standardizing H_D extreme values then all plots would overlap each other. This is demonstrated in Figure B10 where standardized plots along with fitted GEV distribution are provided. All standardized empirical distributions overlap each other and the GEV distribution provides a good fit to standardized data. Thus it is safe to assume that various $H_D(T)$ distributions are linked through a scale factor. The validity of this assumption is investigated using Hosking and Wallis (1993)'s homogeneity test, which was specifically devised for defining homogeneous group 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 C. To apply this test for at-site H_D series, vector Bootstrap technique (Efron and Tibshirani, 1994) was used to evaluate W_k statistics. It was found that W_k statistics were much less than 1 at all four studied stations. 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 suitability of a common at-site growth curve is again verified by comparing $H_D(T)$ and $H(D,T)$ quantiles in section 4.3.

The GEV produced $g(T)$ growth curves for all four stations are plotted together in Figure B11 and they differ from each other in extreme upper and lower tails only. Although there are small differences from one growth curve to the other, it is not possible to assume a single growth curve for these four stations without providing any physical justification with respect to weather producing mechanisms that prevail over these stations during summer season. This aspect is beyond the scope of 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 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 curve using standardized pooled sample.

4.3 Heatwave Magnitude-Duration-Frequency Relations

Estimated quantiles for models M1 to M5 are plotted in Figure B12 against base model quantiles, which are estimated by fitting the GEV distribution to each H_D series. The line of perfect fit help

quickly determine the under- or over-estimation. Overall best estimates are provided by model M5 while model M1, M3 and M4 perform similarly but always over-estimate $H(1,T)$ quantiles. Model M2 provides exact estimates of $H(1,T)$ quantiles, relatively better estimates of $H(10,T)$ and poor estimates of quantiles of heatwave of other in-between durations (i.e. $H(2,T)$ to $H(7,T)$) and difference between base and estimated quantiles increases as return period increases. The significance of over- or under-estimation of base quantiles can be assessed by plotting all estimated quantiles within 90% or 95% or 99% confidence bands. 95% confidence interval is selected here. All estimated and base quantiles along with 95% confidence band for 2 and 50 year return periods are plotted in Figure B13. The 95% confidence intervals are estimated using vector Bootstrap technique (Efron and Tibshirani, 1994). It is hoped that the Bootstrap technique would provide reasonable estimates because of relatively large sample sizes used in this work. In order to determine the number of replicates required to obtain stable estimates of confidence intervals, estimates were obtained for 200, 500, 1000, 2000, 5000, 10000 replicates and assessed against those obtained from 20000 replicates. It was found that 10000 replicates are sufficient to obtain stable estimates of confidence intervals. Hence, the values located at 2.5% and 97.5% position were taken as lower and upper 95% confidence intervals. Figure B13 indicates that the quantiles estimated with model M5 generally lie right inside the 95% confidence band for all durations at all sites. The $H(1,T)$ quantiles estimated with models M1, M3 and M4 lie at the boundary of upper 95% confidence interval and those for other durations inside the confidence band. The $H(2,T)$ to $H(6,T)$ quantiles estimated with model M2 lie at or near the lower boundary of 95% confidence interval. It is expected that if the confidence interval is expanded to 99% then they might fall inside the confidence interval.

The adequacy of M1 to M5 models in preserving the base model quantiles is also assessed on the basis of rRMSE criterion. The plots of rRMSE for various models are provided in Figure B14. The rRMSE varies between 0.77% (7025280) and 0.88% (7016295) for M1, between 1.1% (7025280) and 1.39% (7028124) for M2, between 0.78% (7025280) and 0.88% (7016294) for M3, between 0.85% (7025280) and 0.98% (7016294) for M4 and between 0.47% (7024280) and 0.59% (7028124) for M5. The rRMSE is generally small for small return periods and large for large return periods. The rRMSE is the lowest for M5 for all the four stations and almost indifferent for M1 and M3. Although the rRMSE for M2 is highest among all models for 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 to derive quantiles of heatwave of higher durations from quantiles of heatwave of 1-day duration only.

4.4 PATTERN OF OCCURRENCES OF HEATWAVES

Time series of starting dates of first occurrence of heatwave, SD_D, were extracted from the historical observations and are plotted in Figure B15. It was observed that in some years the heatwave of same magnitude had occurred more than once. Therefore, the starting dates of additional occurrences were also extracted and plotted in Figure B15. It can be seen from this figure that in some years, at most three occurrences of the same heatwave had occurred and this observation is uniform across all four stations. In general, the frequency of second occurrence either decreases or stays almost constant as the duration of heatwave increases. Third occurrence

is limited to heatwaves of: 1-day and 2-day durations for stations 7016294 and 7028124, 1- to 4-day durations for station 7024280, and 1-day and 10-day durations for station 7025280. In some years, first and second occurrence had occurred in association meaning that heatwave persisted for an additional day (e.g. if the starting dates of first and second occurrences had been August 14 and 15, respectively). Also, there are some years where third occurrence had occurred in association with first and second occurrence meaning that heatwave persisted for two additional days. Because of lack of uniformity of additional occurrences, the starting dates of first occurrence, SD_D, are considered for further analysis presented in this section. The analysis of dates of first occurrences would be useful to answer questions like: are we experiencing early summers? An early summer could trigger an early snowmelt season and that may have severe implications in certain regions because of early snowmelt flooding season.

4.4.1 Trend Analysis

Time series of starting dates of first occurrence of heatwave are plotted in Figure B16 along with linear slope estimates (i.e. linear regression slopes). Results of Spearman's rank correlation and Mann-Kendall tests are provided in Table A15 and those of Sen's slope and linear regression slope in Table A16. Unlike the H_D series, some of the SD_D series do represent trend in their occurrence pattern. The results of slope estimates are not uniform for all stations and hence every station is discussed separately.

For station 7024280, both Sen's slope and linear slope methods produce negative slopes for SD_1 to SD_4 series, positive slope for SD_6 to SD_{10} series and opposite values for SD_5. Taking May 1st as the reference point, negative slope means that the starting date of heatwave is moving towards beginning of May and positive slope means that starting date is moving away from beginning of May. Both Sen' slope and regression slope tests produce statistically significant results, at 5% significance level, for SD_3 series. Spearman rank correlation and Mann-Kendall tests confirm this result at 5% significance level. The values of slope estimate for SD_1 to SD_4 series are quite high as compared to remaining series. If the significance level is raised to 15% then, according to Spearman rank correlation and Mann-Kendall tests, more significant cases emerge out.

For station 7016294, both Sen's slope and linear regression slope methods produce negative slopes for all series but only one of them (i.e. slope for SD_5 series) is significant at 5% significance level. Both Spearman rank correlation and Mann-Kendall tests indicate that the trend for SD_5 is statistically significant at 5% significance level. If the significance level is raised to 15% then the trend in SD_4 series becomes significant also. For this station, slope estimates are not as high as in the case of station 7024280.

For station 7025280, both Sen's slope and linear slope methods produce negative slopes for 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 5% significance level. Both Spearman rank correlation and Mann-Kendall tests indicate that trends for SD_D series are statistically non-significant at 5% significance level. If the significance level is raised to 15% then trends for SD_1, SD_2 and SD_{10} series become significant.

For station 7028124, both Sen's slope and linear slope methods produce negative slopes for SD_1 to SD_5 series and positive slopes for the remaining series. None of the tests produced statistically significant results at 5% significance level.

4.4.2 Quartile Plots of Values of SD_D Series

Quartile plots of values of SD_D series are provided in Figure B17. Median heatwave has occurred during the third week of July at station 7025280 and during the second and third week of July at the remaining stations. The range of dates of occurrence decreases as duration of heatwave increases. For long duration heatwaves (e.g. heatwaves of 7- and 10-day durations), the range of occurrence is generally concentrated over the interval from the third week of May to the last week of August at station 7016294 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-day durations), the range of occurrence is generally concentrated between the third week of May and the fourth week of August at station 7016294 and between the third week of May and the third week of September at other three stations.

4.4.3 Association of Heatwave Magnitudes and Dates of Occurrences

The relationship between heatwave magnitude and date of occurrence is analyzed by plotting magnitude against date of occurrence (Figure B18) and estimating linear regression slopes. This relationship is marginally significant (non-significant) at 5% (1%) significance level for two heatwave series (i.e. H_1 and H_5) of station 7024280 indicating that there is a tendency for intense heatwaves to occur later during the summer season. For the remaining heatwave series of station 7024280 and for all heatwave series of all other stations, no significant relationship was found at 5% significance level indicating that there is no tendency for larger (smaller) magnitudes of heatwave to occur later (earlier) in the summer season. Having noted these results, it is possible to determine the time window where the majority of heatwaves were concentrated through relative frequency estimates over non-overlapping time windows. For this purpose, each month was divided into three time windows (first window: 1 to 10 days, second window: 11 to 20 days, third window: 21 to 30 or 21 to 31 days) and counts of occurrences were performed over these non-overlapping time intervals. The counts of occurrences were converted into relative frequencies and results are shown in Figure B18. Additional occurrences of heatwave, as indicated in section 4.4, were also included into counts while estimating relative frequencies. Majority of these distributions could be considered as having a symmetric pattern with the center located in the vicinity of third week of July. The majority of heatwaves had occurred during the time interval from last ten days of June to first ten days of August.

It was also interesting to analyze the relationship between the most extreme heatwaves and their corresponding dates of occurrences. For this purpose, 10 most extreme heatwaves and their corresponding dates of occurrence are plotted in Figure B19. It is possible that the 10^{th} extreme heatwave had occurred more than once. Therefore, these additional occurrences are also plotted in Figure B19. It is clear from the scatter of plotted data that no relation exists between extreme heatwave and the corresponding date of occurrence within the summer season except that July is the most probable month for their occurrence. This result was further confirmed by performing linear regression and analyzing the significance of slope estimates at 5% significance level.

5.0 DISCUSSION AND CONCLUSIONS

5.1 DISCUSSION

The results obtained confirm that adopting the regional index-flood frequency analysis approach for HDF analysis, similar to the one used by Javelle et al. (2002, 2003) for QDF analysis, describes the behavior of heatwave of 1-10 days duration in a valuable and parsimonious manner. Two or three-parametric form of $\mu(D)$ function along with 3-parameteric at-site growth curve can model various quantiles of heatwave in an acceptable manner when assessed on the basis of rRMSE and 95% Bootstrap confidence intervals. The HDF approach is demonstrated using annual maximum series only. However, the approach is equally applicable for peaks-over-threshold series and it will be interesting to explore the suitability of Generalized Pareto distribution (Smith, 2001) for the current dataset. Future work should address this investigation. The current study attempts to model heatwave by explicitly taking into account its duration and in that respect the study findings will complement the work of Kharin and Zweirs (2000), who used L-moment and GEV approach to model daily maximum temperature series only. Heatwave events are solely defined 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 health industry. The proposed HDF methodology can be adapted to this definition of heatwave or to any other definition of heatwave based on suitably chosen combination of meteorological variables.

In this study, $\mu(D)$ and $g(T)$ functions are modeled independently from each other. 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 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 the empirical distributions or on the basis of descriptive and predictive ability of the assumed distribution along with a parameter estimation procedure. So, whatever the choice of $g(T)$ is, the $\mu(D)$ function remains unaffected. Six different functional forms of $\mu(D)$ function were analyzed. It is possible to assume other functional forms, e.g. a second degree polynomial would provide an equal fitting to the one observed with $f_5(D)$. However, it is more interesting to investigate the suitability of $\mu(D)$ functional forms that have been used for modeling other hydro-meteorological variables in an endeavor 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 which allow site independent comparison. The study analyzed the $\mu(D)$ functional forms used for IDF and QDF analyses, in addition to three other intuitively appealing forms, and reached at the conclusion that the $\mu(D)$ function for the current dataset can best be modeled by using three parametric $f_5(D)$ form. This implies that there is more than one relation that governs the $\mu(D)$ function. The presence of more than one functional form has been noticed by Ferro and Porto (1999) and

Porras and Porras (2001) for rainfall data. The two parametric functional form $f_1(D)$ is the second best choice. The $g(T)$ function is modeled using the GEV distribution when its parameters are estimated using the PWM method. This choice is made on the basis of the RMSE as a goodness-of-fit criterion. It is possible that a different distribution and parameter estimation procedure may emerge as the best fitting model if an exhaustive analysis has been undertaken. On the basis of limited analysis carried out in this work, even the AIC model selection criterion favors the GEV distribution when its parameters are estimated using the ML procedure. According to Hosking et al. (1985), the PWM method for the GEV is not sensitive to outliers and its efficiency is comparable to that of ML method. Thus selecting the GEV for current analysis does not seem to be an unreasonable choice. Although the L-moment ratio diagram is not used for model selection, it indicates that the GEV, LN3 and P3 distributions are equally likely candidate models for the current dataset. It is noted in the results section that negligible differences were found between dimensionless quantiles obtained by fitting the GEV distribution using averaged standardized PWMs and PWMs of standardized pooled sample. Hence, the $g(T)$ function can also be modeled 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 the results section that heatwaves of 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 adequate.

This study models heatwave as average maximum daily temperature that prevails over D-day duration and hence the T-year magnitude of heatwave is the average maximum temperature that prevails over D consecutive days. Multiplying T-year quantile of heatwave of duration D-day by the corresponding duration D, gives the sum of D-day maximum temperature. It is possible to disintegrate this sum over D-day duration using an appropriate disintegration technique e.g. by modeling the duration-dependent standardized typical heatwave profiles similarly as in the case of rainfall disintegration (see for example Acreman, 1990). This point remains outside the scope of present study and further analysis would be required to incorporate this aspect in the proposed model.

Combined plots of dimensionless growth curves (Figure B11) suggest 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 homogenous 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. Again this point remains outside the scope of present study and future work should address this point in the context of homogeneous regions.

The pattern of occurrence of heatwaves is studied using time series of starting dates of first occurrence of heatwaves of various durations. It is found that heatwaves of short durations (1- to 5-day) exhibit stronger negative slopes and those of long durations (> 5-day) exhibit moderately positive slopes. This implies that over the studied period there had been shift in the occurrence of heatwaves of small duration. There are more significant cases at 15% significance level than at

5%. Combined results of Spearman rank correlation and Mann-Kendall tests and Sen' slope and linear regression slope estimates confirm that the trends in some of the SD_D series are not just because of chance observations. Although multiple occurrences (generally two or three) of heatwave of same magnitude in the same year have been noticed, it is deemed sufficient to study the patterns of first date of occurrence for the current analysis. This decision is based on the observation of non-uniformity of additional occurrences across durations. Except for two cases, it is found that no association exists between heatwave magnitude and date of occurrence. Thus there is insufficient evidence to state that heatwave tends to occur in the beginning or in the middle or at the end of the summer season.

During data screening, one heatwave series has shown signs of non-independence and another has shown signs of second-order non-stationarity while being first-order stationary. Hence, the results of these series should be interpreted with caution. 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. In reality, a generalized non-stationary approach needs to be devised and that should reduce to stationary approach as a special case.

5.2 CONCLUSIONS

Motivated by the regional flood frequency analysis and QDF analysis, this report presented HDF analysis for modeling at-site daily maximum temperature data in an endeavor to interpret heatwave. Two main components of the HDF approach are $\mu(D)$ function to model the duration dependent scale factor and $g(T)$ function to model the at-site dimensionless growth curve. Hence many of the conclusions drawn are directly dependent upon the form of these two functions. The main findings of the study are summarized below:

It was found that the $\mu(D)$ function can best be modeled using $f_5(D)$ functional form. This functional form preserved μ_1 and provided good estimates for μ_2 to μ_{10}. It also provided good estimates of μ_D in the extrapolated region. The functional forms $f_1(D)$ and $f_3(D)$ provided similar estimates of μ_D in the studied region but over-estimated values in the extrapolated region. The functional form $f_6(D)$, while being suitable for QDF analysis, did not follow the trend of μ_D as a function of D and hence was not a suitable choice for current dataset. The estimated parameters of $f_1(D)$ and $f_3(D)$ were almost same. This indicates that there is negligible effect of parameter estimation procedure on the estimated parameters. In general, $\mu(D)$ function can best be modeled using a relationship of the form $\mu(D) = aD^b$. This type of relation has been used by Ferro and Porto (1999) and Porras and Porras (2001) for IDF analyses. It would be interesting to investigate its validity for modeling $\mu(D)$ function in the case of QDF analysis.

The GEV distribution was selected to model at-site growth curve. This choice was made on the basis of its good fit to empirical observations. Previous studies had also used the GEV (e.g. Kharin and Zwiers, 2000), but few studies had used other distributions. The LN3 and P3 could

be alternate choices for modeling at-site growth curves as shown by AIC model selection criterion. The L-moment ratio diagram indicated that the GEV, LN3 and P3 distributions could be used to model current dataset. The Normal distribution, which was used in some of the previous studies to model daily maximum temperature data, could not be selected as a suitable candidate model for any of the data series on the basis of AIC and RMSE criteria. If one wishes to use 2-parameter distribution for the current dataset then the best choice is LN2.

The models M1 and M3 provided almost similar results but slightly better than M4. Although these two models are very different with respect to quantile estimation procedures, the approach of model M1 is straightforward and should be preferred over the latter. While designing model M2, it was hoped that it would be very useful if quantiles of heatwave of larger durations than 1-day, could be scaled from quantiles of heatwave of 1-day duration. This scaling procedure, while preserving quantiles of 1-day heatwave, was strictly applicable only for heatwaves of 7- and 10-day duration. The model M5 provided the best performance and estimated quantiles were very similar to those of base quantiles. Base quantiles were estimated by fitting a GEV-PWM distribution to each heatwave series independently. These results indicated that adopting regional index-flood frequency analysis methodology for modeling at-site heatwaves is a reasonable approach.

By performing an analysis of pattern of occurrence of heatwaves, it was found that heatwaves of small duration had shifted over the studied period and had advanced towards beginning of May. There were more significant cases that indicated such shifts at 15% significance level than at 5%. The analysis indicated that there were stronger negative slopes for small duration heatwaves and moderately positive slopes for large duration heatwaves. Considering H_1 to H_5 series at four studied stations, it was observed that 95% of series showed downward trend and 55% of them had negative slope magnitude >0.1 day/year. This implies that over the studied period there had been a shift in the occurrence of heatwaves of small duration. 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 (7025280) and during the second and third week of July at the other three stations. July and August are peak summer months but occasionally the individual days having highest daily maximum temperature and group of days having highest average maximum daily temperature had been observed as early as third week of May and as late as last week of September for the dataset analyzed in this study. It was noted that heatwave of same magnitude had occurred more than once in some of the years. In general, additional occurrences were non-uniform across durations of heatwave because third additional occurrence was limited to heatwaves of 1- and 2-day durations only. Except for two cases, no association was found between heatwave magnitude and the corresponding date of occurrence meaning there is no tendency for heatwaves to occur during the early part or 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 in the vicinity of third week of July. The pattern of occurrences of ten largest heatwaves did not exhibit any temporal behavior within the summer season except that July is the most probable month for their occurrence.

6.0 REFERENCES

Acreman, M.C. 1990. A simple stochastic model for hourly rainfall for Farnborough, England. *Hydrol. Sci. J.* 35: 119-148.

Anderson, R.L. 1942. Distribution of the Serial Correlation Coefficient. *Ann. Math. Stat.* 13: 1-13.

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

Bonsal, B.R., Zhang, X., Vincent L.A., and Hogg, W.D. 2001. Characteristics of daily and extreme temperatures over Canada. *J. Clim.* 14: 1959-1976.

Box, G.P. and Jenkins, G.M. 1976. Time Series Analysis, Forecasting and Control. Holden-day, San Francisco.

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

Cunnane, C. 1978. Unbiased plotting positions – a review. *J. Hydrol.* 37 (3/4): 205-222.

Dahmen, E.R. and Hall, M.J. 1990. Screening of Hydrological Data. International Institute for Land Reclamation and Improvement (ILRI), Netherlands. Publication No. 49, 58 pp.

Dalrymple, T. 1960. Flood frequency methods. U.S. Geological Survey Water Supply paper 1543 A, pp. 11-51.

Duenas, C., Fernandez, M.C., Canete, S., Carretero, J. and Liger, E. 2002. Assessment of Ozone variations and meteorological effects in an urban area in the mediteranean coast. *The Sience of Total Environment* 299: 97-113.

Drouin, L. and King, N. 2004. Changements climatiques et santé publique. Poster presented at 1st Ouranos symposium on climate changes, June 9-10, 2004, Montreal, QC, Canada.

Efron, B. and Tibshirani, R.J. 1994. An introduction to the Bootstrap. Monographs on applied probability and statistics. Chapman and Hall.

Ferro, V. and Porto, P. 1999. Regional analaysis of rainfall-depth-duration equation for South Itally. *J. Hydrol. Eng.*, 4(4): 326-336.

GREHYS. 1996. Presentation and review of some methods for regional flood frequency analysis. *J. Hydrol.* 186(1-4): 63-84

Haan, C.T., 1977. Statistical methods in hydrology. The Iowa State University Press. Ames, Iowa.

Hosking, J.R.M. and Wallis, J.R. 1997. Regional frequency analysis. Cambridge university press, U.K.

Hosking, J.R.M. and Wallis, J.R. 1993. Some statistics useful in regional frequency analysis. *Water Resour. Res.*, 29(2): 271-281.

Hosking, J.R.M. 1990. L-moments: analysis and estimation of distributions using linear combinations of ordered statistics. *J. R. Statis. Soc. Ser. B.* 52(1): 105-124.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F. 1985. Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics* 27(3): 251-261.

Huth, R., Kysely, J. and Pokorna, L. 2000. A GCM simulation of heatwaves, dry spells, and their relationships to circulation. *Clim. Change* 46:29-60.

HYFRAN. 2003. A software package for statistical modeling. INRS-Eau, University of Quebec, Sainte-Foy, Quebec, Canada.

Javelle, P. Ouarda, T.B.M.J. and Bobée, B. 2003. Spring flood analysis using the flood-duration-frequency approach: application to the provinces of Quebec and Ontario, Canada. *Hydrol. Process.* 17: 3717-3736.

Javelle, P. Ouarda, T.B.M.J., Lang, M. Bobée, B., Galéa, G. and Grésillon, J.M. 2002. Development of regional flood-duration-frequency curves based on the index flood method. *J. Hydrol.* 258(1-4): 249-259.

Javelle, P. Galéa, G. and Grésillon, J.M. 2000. The flow-duration-frequency approach: former and new developments in French. *Revue des Sciences de l'Eau* 13(3): 305-323.

Katz, R.W., Parlang, M.B. and Naveau, P. 2002. Statistics of extremes in hydrology. *Adv. Water Resour.*, 25: 1287-1304.

Kendall, M.G. 1975. Rank correlation methods. Charless Griffin, London.

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

Lang, M., Ouarda, T.B.M.J. and Bobee, B. 1999. Towards operational guidelines for over-threshold modeling. *J. Hydrol.*, 225: 103-117.

Madsen, H., Rasmussen, P.F. and Rosbjerg, D. 1997a. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. At-site modeling. *Water Resour. Res.*, 33(4): 747-757.

Madsen, H., Pearson, C.P. and Rosbjerg, D. 1997b. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 2. Regional modeling. *Water Resour. Res.*, 33(4): 759-769.

Mann, H.B. 1945. Non-parametric tests against trend. *Econometrica* 13: 245-259.

Mutua, F.M. 1994. The use of Akaike Information Criterion in the identification of an optimum flood frequency model. *Hydrol. Sci. J.* 39 (3): 235-244.

Porras (Sr.), P.J. and Porras (Jr.), P.J. 2001. New perspective on rainfall frequency curves. *J. Hydrol. Eng.* 6(1): 82:85.

Rainham, D.G.C. and Smoyer-Tomic, K.E. 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, J.D., Delleur, J.W., Yevjevich, V.M., and Lane, W.L. 1980. Applied modeling of hydrologic time series. Littleton, Colorado, Water Resources Publications, 484 pp.

Sen, P.K. 1968. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* 63: 1379-1389.

Smakhtin, V.U. 2001. Low flow hydrology: a review. *J. Hydrol.*, 240: 147-186.

Smith, R.L. 2001. Extreme value statistics in meteorology and the environment. Chapter 8: Environmental Statistics: 300-357. [available at http://www.stat.unc.edu/postscript/rs/envstat/env.html]

Snedecor, G.W. and Cochran, W.G. 1989. Statistical methods. Macmillan Publishing Company, New York.

Subak, S., Palutikof, J.P., Agnew, M.D., Watson, S.J., Bentham, C.G. Hulme, M., McNally, S., Thornes, J.E., Waughray, D. and Woods, J.C. 2000. The impact of anomolous weather of 1995 on the U.K. Economy. *Climatic Change* 44(1/2): 1-26.

Strupczewski, W.G., Singh, V.P. and Mitosek, H.T. 2001. Non-stationary approach to at-site flood frequency modeling. III. Flood analysis of Polish rivers. *J. Hydrol.* 248: 152-167.

Sveinsson, O.G.B., Salas, J.D. and Boes, D.C. 2002. Regional frequency analysis of extreme precipitation in Northeastern Colorado and Fort Collins flood of 1997. *J. Hydrol. Eng.* 7(1): 49-63.

UK Guardian. 2003. Global warming may be speeding up, fears scientist. UK Guardian of August 12, 2003.

Viessman, W., Knapp, J.W., Lewis, G.L. and Harbaugh, T.E. 1977. Introduction to Hydrology, second edition. Harper and Row Publishers, Inc., New York.

Walpole, R.E. and Myers, R.H. 1989. Probability and statistics for engineers and scientists, 4th edition. Macmillan publishing company, New York. 765 pp.

World Meteorological Organization/WMO. 1966. Climatic Change. J.M. Mitchell (Editor). WMO No. 195. TP. 100. WMO, Geneva.

Wettstein, J.J. and Mearns, L.O. 2002. The influence of North Atlantic Oscillation on mean , variance, and extremes of temperature in the Notheastern United States and Canada. *J. Clim.* 15: 3586-3600.

Yu, F.X., Li, W.Q., Singh, V.P. and Naghavi, B. 1993. Estimating LP3 parameters using a combination of the method of moments and least squares. *J. Environmental Hydrology* 1(2): 9-19.

Zhang, X., Vincent, L.A., Hogg, W.D. and Niitsoo, A. 2000. Temperature and precitipation trends in Canada during the 20th century. *Atmos.-Ocean* 38: 395-429.

7.0 APPENDIX A

This appendix contains tables referred to in the study.

Table A1. Station codes and record lengths used in the study.

Station Code	Period of observations	Record length (years)
7024280	1915 - 1995	81
7016294	1895 - 1995	101
7025280	1895 - 1993	98
7028124	1994 - 1994	91

Table A2. Results of Spearman's rank correlation and Mann-Kendall tests for heatwave data observed at four studied stations.

Station: 7024280				
Heatwave duration (days)	Spearman rank correlation test		Mann-Kendall test	
	Test statistic	*p*-value	Test statistic	*p*-value
1	0.070	0.533	0.615	0.539
2	0.072	0.525	0.707	0.48
3	0.051	0.649	0.559	0.576
4	0.022	0.847	0.314	0.753
5	0.052	0.646	0.522	0.602
6	0.061	0.587	0.543	0.587
7	0.071	0.528	0.677	0.498
10	0.096	0.392	0.84	0.401

Note: A small *p*-value indicates a significant correlation, i.e. there is a trend in the time series (e.g. *p*-value<= 0.05 means significant correlation at significance level alpha=5%)

Table A2. Contd.

Station: 7016294				
Heatwave duration (days)	Spearman rank correlation test		Mann- Kendall test	
	Test statistic	*p*-value	Test statistic	*p*-value
1	0.097	0.335	0.993	0.321
2	0.164	0.100	1.633	0.102
3	0.191	0.056	1.858	0.063
4	0.162	0.106	1.623	0.105
5	0.110	0.274	1.089	0.276
6	0.088	0.383	0.833	0.405
7	0.099	0.325	0.968	0.333
10	0.100	0.321	1.018	0.309

Table A2. Contd.

Station: 7025280				
Heatwave duration (days)	Spearman rank correlation test		Mann- Kendall test	
	Test statistic	*p*-value	Test statistic	*p*-value
1	-0.049	0.630	-0.553	0.580
2	-0.012	0.905	-0.126	0.900
3	-0.007	0.944	-0.111	0.912
4	0.016	0.873	0.187	0.851
5	0.032	0.755	0.252	0.801
6	0.030	0.769	0.258	0.797
7	0.048	0.637	0.473	0.636
10	0.055	0.591	0.574	0.566

Table A2. Contd.

Station: 7028124				
Heatwave duration (days)	Spearman rank correlation test		Mann- Kendall test	
	Test statistic	*p*-value	Test statistic	*p*-value
1	-0.074	0.487	-0.662	0.508
2	-0.026	0.808	-0.209	0.834
3	-0.001	0.989	-0.014	0.989
4	0.011	0.920	0.113	0.910
5	0.014	0.896	0.182	0.856
6	0.018	0.869	0.151	0.880
7	0.026	0.806	0.237	0.813
10	0.041	0.701	0.247	0.805

Table A3. Linear regression and Sen's slope estimates for heatwaves observed at four studied stations.

Station: 7024280						
Heatwave duration (days)	Linear slope			Sen slope		
	Slope	95% LL	95% UL	Slope	95% LL	95% UL
1	0.001	-0.014	0.016	0.000	-0.006	0.017
2	0.001	-0.014	0.016	0.004	-0.008	0.019
3	0.000	-0.015	0.014	0.004	-0.011	0.018
4	0.000	-0.015	0.014	0.002	-0.013	0.018
5	0.002	-0.013	0.017	0.004	-0.012	0.019
6	0.002	-0.013	0.017	0.005	-0.012	0.020
7	0.003	-0.013	0.018	0.005	-0.010	0.020
10	0.003	-0.011	0.017	0.006	-0.009	0.021

LL: Lower confidence limit; UL: Upper confidence limit

Table A3. Contd.

Station: 7016294						
Heatwave duration (days)	Linear slope			Sen slope		
	Slope	95% LL	95% UL	Slope	95% LL	95% UL
1	0.005	-0.006	0.015	0.000	0.000	0.015
2	0.008	-0.003	0.019	0.009	0.000	0.021
3	0.010	-0.001	0.020	0.011	0.000	0.023
4	0.007	-0.004	0.018	0.009	-0.002	0.020
5	0.004	-0.007	0.015	0.006	-0.005	0.017
6	0.003	-0.007	0.014	0.005	-0.006	0.016
7	0.003	-0.007	0.014	0.005	-0.006	0.016
10	0.003	-0.007	0.013	0.006	-0.005	0.015

Table A3. Contd.

Station: 7025280						
Heatwave duration (days)	Linear slope			Sen slope		
	Slope	95% LL	95% UL	Slope	95% LL	95% UL
1	-0.003	-0.013	0.007	-0.001	-0.013	0.005
2	-0.002	-0.012	0.008	0.000	-0.011	0.009
3	-0.002	-0.012	0.008	0.000	-0.011	0.010
4	-0.001	-0.011	0.010	0.001	-0.010	0.011
5	0.001	-0.010	0.011	0.001	-0.010	0.012
6	0.000	-0.011	0.010	0.001	-0.010	0.013
7	0.000	-0.011	0.010	0.003	-0.010	0.013
10	0.001	-0.009	0.011	0.003	-0.008	0.014

Table A3. Contd.

Station: 7028124						
Heatwave duration (days)	Linear slope			Sen slope		
	Slope	95% LL	95% UL	Slope	95% LL	95% UL
1	-0.007	-0.017	0.004	-0.003	-0.015	0.007
2	-0.003	-0.015	0.008	-0.001	-0.013	0.010
3	-0.002	-0.013	0.010	0.000	-0.013	0.012
4	-0.001	-0.012	0.011	0.001	-0.012	0.013
5	0.000	-0.011	0.012	0.001	-0.011	0.013
6	0.001	-0.011	0.012	0.001	-0.010	0.013
7	0.000	-0.011	0.012	0.001	-0.009	0.014
10	0.000	-0.011	0.011	0.002	-0.010	0.013

Table A4. Lag-1 autocorrelation coefficients and their 95% confidence intervals.

Heatwave duration (days)							
1	2	3	4	5	6	7	10
Station: 7024280 LCL: -0.230, UCL: 0.205							
-0.125	-0.134	-0.136	-0.140	-0.145	-0.092	-0.070	-0.128
Station: 7016294 LCL: -0.205, UCL: 0.185							
-0.117	-0.017	0.031	-0.008	-0.011	-0.007	0.011	0.008
Station: 7025280 LCL: -0.208, UCL: 0.188							
-0.117	-0.158	-0.148	-0.132	-0.116	-0.141	-0.128	-0.070
Station: 7028124 LCL: -0.217, UCL: 0.194							
0.054	-0.015	-0.060	-0.064	-0.127	-0.108	-0.085	-0.120

Note: None of the lag-1 autocorrelation coefficient is significant at 5% level of significance; LCL: lower confidence limit; UCL: upper confidence limit

Table A5. The numbers of r_k values, that falls outside the confidence bands.

Level of significance, α (%)	Heatwave duration (days)							
	1	2	3	4	5	6	7	10
Station: 7024280								
10	0	0	0	0	0	1	0	0
5	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
Station: 7016294								
10	2	1	3	1	2	2	2	3
5	2	1	1	1	1	1	1	1
1	0	1	1	1	0	0	0	0
Station: 7025280								
10	1	0	1	1	1	1	2	1
5	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
Station: 7028124								
10	2	1	0	0	0	1	0	0
5	1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0

Note: Shaded cell indicate that the time series could be persistent.

Table A6a. *p*-values for F-tests of stability/equality of variance of a time series considering two sub-samples.

Heatwave duration (days)							
1	2	3	4	5	6	7	10
Station: 7024280; Sample sizes: 1: 40 (1-40), 2: 41 (41-81)							
0.106	0.181	0.263	0.846	0.670	0.551	0.367	0.533
Station: 7016294; Sample sizes: 1: 50 (1-50), 2: 51 (51-101)							
0.623	0.866	0.518	0.572	0.620	0.968	0.963	0.859
Station: 7025280; Sample sizes: 1: 49 (1-49), 2: 49 (50-98)							
0.572	0.961	0.704	0.885	0.789	0.649	0.424	0.890
Station: 7028124; Sample sizes: 1: 45 (1-45), 2: 46 (46-91)							
0.138	0.251	0.158	0.371	0.596	0.709	0.650	0.612

Note: a small *p*-value (e.g., ≤ 0.05) indicates a significant test.
*: significant at α=10%; **: significant at α=5%; ***: significant at α=1%

Table A6b. *p*-values for F-tests of stability/equality of variance of a time series considering three sub-samples.

Sub-samples	Heatwave duration (days)							
	1	2	3	4	5	6	7	10
Station: 7024280; Sample sizes: 1: 27 (1-27), 2: 27 (28-54), 3: 27 (55-81)								
1, 2	0.709	0.684	0.765	0.820	0.948	0.518	0.534	0.717
1, 3	0.359	0.315	0.356	0.770	0.507	0.493	0.394	0.543
2, 3	0.585	0.548	0.532	0.604	0.466	0.185	0.143	0.332
Station: 7016294; Sample sizes: 1: 34 (1-34), 2: 34 (35-68), 3: 33 (69-101)								
1, 2	0.712	0.893	0.993	0.963	0.994	0.718	0.521	0.896
1, 3	0.050*	0.147	0.105	0.239	0.460	0.991	0.943	0.788
2, 3	0.109	0.114	0.107	0.222	0.464	0.711	0.572	0.690
Station: 7025280; Sample sizes: 1: 33 (1-33), 2: 33 (34-66), 3: 32 (67-98)								
1, 2	0.627	0.830	0.556	0.458	0.427	0.580	0.541	0.182
1, 3	0.172	0.413	0.168	0.365	0.724	0.787	0.945	0.551
2, 3	0.374	0.303	0.422	0.864	0.664	0.780	0.590	0.464
Station: 7028124; Sample sizes: 1: 30 (1-30), 2: 30 (31-60), 3: 31 (61-91)								
1, 2	0.111	0.161	0.258	0.371	0.518	0.965	0.853	0.849
1, 3	0.017**	0.100	0.071*	0.301	0.220	0.265	0.276	0.253
2, 3	0.422	0.816	0.502	0.145	0.563	0.247	0.203	0.340

Note: a small *p*-value (e.g., $p \leq 0.05$) indicates a significant test.
*: significant at α=10%; **: significant at α=5%; ***: significant at α=1%

Table A7a. *p*-values for t-tests of stability/equality of mean of a time series considering two sub-samples.

Heatwave duration (days)							
1	2	3	4	5	6	7	10
Station: 7024280; Sample sizes: 1: 40 (1-40), 2: 41 (41-81)							
0.540	0.408	0.310	0.349	0.371	0.374	0.450	0.574
Station: 7016294; Sample sizes: 1: 50 (1-50), 2: 51 (51-101)							
0.130	0.057*	0.069*	0.158	0.251	0.275	0.209	0.271
Station: 7025280; Sample sizes: 1: 49 (1-49), 2: 49 (50-98)							
0.459	0.714	0.716	0.842	0.942	0.981	0.986	0.779
Station: 7028124; Sample sizes: 1: 45 (1-45), 2: 46 (46-91)							
0.136	0.194	0.194	0.206	0.260	0.252	0.261	0.454

Note: a small *p*-value (e.g., ≤ 0.05) indicates a significant result.
*: significant at α=10%; **: significant at α=5%; ***: significant at α=1%

Table A7b. *p*-values for t-tests of stability/equality of mean of a time series considering three sub- samples.

Sub-samples	Heatwave duration (days)							
	1	2	3	4	5	6	7	10
Station: 7024280; Sample sizes: 1: 27 (1-27), 2: 27 (28-54), 3: 27 (55-81)								
1, 2	0.445	0.532	0.500	0.569	0.404	0.435	0.496	0.522
1, 3	0.917	0.835	0.979	0.892	0.789	0.782	0.669	0.633
2, 3	0.468	0.637	0.478	0.651	0.534	0.569	0.740	0.828
Station: 7016294; Sample sizes: 1: 34 (1-34), 2: 34 (35-68), 3: 33 (69-101)								
1, 2	0.680	0.254	0.275	0.330	0.358	0.617	0.696	0.563
1, 3	0.469	0.197	0.134	0.288	0.546	0.628	0.621	0.749
2, 3	0.780	0.990	0.779	1.000	0.713	0.977	0.936	0.786
Station: 7025280; Sample sizes: 1: 33 (1-33), 2: 33 (34-66), 3: 32 (67-98)								
1, 2	0.621	0.590	0.593	0.689	0.618	0.741	0.810	0.899
1, 3	0.604	0.686	0.622	0.805	0.984	0.866	0.787	0.857
2, 3	0.274	0.336	0.263	0.486	0.622	0.607	0.600	0.736
Station: 7028124; Sample sizes: 1: 30 (1-30), 2: 30 (31-60), 3: 31 (61-91)								
1, 2	0.671	0.918	0.791	0.709	0.590	0.593	0.646	0.637
1, 3	0.163	0.318	0.476	0.657	0.866	0.987	0.954	0.972
2, 3	0.249	0.298	0.261	0.353	0.430	0.565	0.653	0.574

Note: a small *p*-value (e.g., ≤ 0.05) indicates a significant test.
*: significant at α=10%; **: significant at α=5%; ***: significant at α=1%

Table A8a. An overall summary of F-tests based on two and three sub-samples.

Level of significance, α (%)	Heatwave duration (days)							
	1	2	3	4	5	6	7	10
Station: 7024280								
10	S	S	S	S	S	S	S	S
5	S	S	S	S	S	S	S	S
1	S	S	S	S	S	S	S	S
Station: 7016294								
10	NS	S	S	S	S	S	S	S
5	S^m	S	S	S	S	S	S	S
1	S	S	S	S	S	S	S	S
Station: 7025280								
10	S	S	S	S	S	S	S	S
5	S	S	S	S	S	S	S	S
1	S	S	S	S	S	S	S	S
Station: 7028124								
10	NS	S	NS	S	S	S	S	S
5	NS	S	S	S	S	S	S	S
1	S	S	S	S	S	S	S	S

S: Time series is probably stationary; NS: Time series is probably non-stationary;
m: A marginal case

Table A8b. An overall summary of t-tests based on two and three sub-samples.

Level of significance, α (%)	Heatwave duration (days)							
	1	2	3	4	5	6	7	10
Station: 7024280								
10	S	S	S	S	S	S	S	S
5	S	S	S	S	S	S	S	S
1	S	S	S	S	S	S	S	S
Station: 7016294								
10	S	NS	NS	S	S	S	S	S
5	S	S	S	S	S	S	S	S
1	S	S	S	S	S	S	S	S
Station: 7025280								
10	S	S	S	S	S	S	S	S
5	S	S	S	S	S	S	S	S
1	S	S	S	S	S	S	S	S
Station: 7028124								
10	S	S	S	S	S	S	S	S
5	S	S	S	S	S	S	S	S
1	S	S	S	S	S	S	S	S

S: Time series is probably stationary; NS: Time series is probably non-stationary

Table A9. Basic statistics (mean, standard deviation and coefficients of variation and skewness) of heatwaves observed at four studied locations.

Station: 7024280				
HW duration (days)	Mean	Std. deviation	Coeffi. of variation	Coeffi. of skewness
1	32.69	1.55	0.047	0.69
2	31.93	1.58	0.050	0.84
3	31.28	1.55	0.050	0.63
4	30.70	1.52	0.050	0.39
5	30.18	1.56	0.052	0.42
6	29.82	1.61	0.054	0.46
7	29.44	1.58	0.054	0.53
10	28.70	1.48	0.052	0.41

Table A9. Contd.

Station: 7016294				
HW duration (days)	Mean	Std. deviation	Coeffi. of variation	Coeffi. of skewness
1	31.96	1.55	0.048	0.25
2	31.15	1.65	0.053	0.18
3	30.49	1.63	0.053	0.12
4	29.85	1.60	0.054	0.38
5	29.28	1.57	0.054	0.37
6	28.85	1.55	0.054	0.32
7	28.48	1.51	0.053	0.35
10	27.77	1.44	0.052	0.24

Table A9. Contd.

Station: 7025280				
HW duration (days)	Mean	Std. deviation	Coeffi. of variation	Coeffi. of skewness
1	32.83	1.40	0.043	0.26
2	32.05	1.37	0.043	0.32
3	31.42	1.41	0.045	0.39
4	30.90	1.44	0.047	0.30
5	30.43	1.44	0.047	0.26
6	30.05	1.46	0.048	0.18
7	29.73	1.45	0.049	0.25
10	28.97	1.40	0.048	0.09

Table A9. Contd.

Station: 7028124				
HW duration (days)	Mean	Std. deviation	Coeffi. of variation	Coeffi. of skewness
1	31.80	1.36	0.043	0.47
2	30.89	1.40	0.045	0.53
3	30.19	1.44	0.048	0.43
4	29.65	1.47	0.050	0.23
5	29.14	1.44	0.050	0.25
6	28.74	1.48	0.051	0.29
7	28.38	1.45	0.051	0.39
10	27.60	1.42	0.051	0.26

Table A10. Goodness-of-fit analysis of two parametric distributions to heatwaves observed at four studied stations.

Station: 7024280					
Heatwave duration (days)	Distribution	Criterion			
		AIC	Chi-Squared		RMSE (%)
			Statistic	*p*-value	
1	NOR	303.40	20.99	0.01	33.18
	LN2	301.03	20.99	0.01	29.34
	EV1	294.25	26.69	0.00	21.17
2	NOR	307.32	17.19	0.03	38.66
	LN2	304.26	15.56	0.05	34.20
	EV1	293.62	15.01	0.06	19.21
3	NOR	303.91	13.65	0.09	31.64
	LN2	301.63	11.21	0.19	27.64
	EV1	295.11	5.23	0.73	22.19
4	NOR	301.18	13.38	0.10	24.56
	LN2	299.81	15.28	0.05	21.44
	EV1	298.09	6.05	0.64	30.03
5	NOR	305.27	9.58	0.30	25.24
	LN2	303.74	12.30	0.14	21.65
	EV1	301.99	9.85	0.28	28.80
6	NOR	309.69	5.78	0.67	25.03
	LN2	307.97	3.06	0.93	20.75
	EV1	305.96	6.86	0.55	26.42
7	NOR	307.30	6.05	0.64	27.56
	LN2	305.28	6.86	0.55	23.25
	EV1	301.84	9.04	0.34	23.79
10	NOR	296.68	4.96	0.76	20.12
	LN2	295.24	3.06	0.93	16.39
	EV1	295.38	3.60	0.89	26.45
Overall rank					
	NOR	3	2		3
	LN2	2	1		1
	EV1	1	1		2

Note: HYFRAN (2003) software uses a scaling procedure to map AIC values on to positive domain. These AIC values are very different from those one would normally expect using AIC equation.

Table A10. Contd.

Station: 7016294					
Heatwave duration (days)	Distribution	Criterion			
		AIC	Chi-Squared		RMSE (%)
			Statistic	*p*-value	
1	NOR	378.03	25.06	0.00	20.56
	LN2	377.14	25.06	0.00	18.90
	EV1	384.06	21.97	0.01	43.44
2	NOR	390.35	13.42	0.14	15.58
	LN2	389.76	13.89	0.13	13.78
	EV1	397.71	10.80	0.29	46.11
3	NOR	388.42	11.04	0.27	12.03
	LN2	388.11	7.95	0.54	10.96
	EV1	397.28	11.51	0.24	48.72
4	NOR	385.06	3.20	0.96	19.50
	LN2	383.38	8.43	0.49	15.46
	EV1	384.95	6.29	0.71	31.81
5	NOR	381.32	6.76	0.66	20.14
	LN2	379.67	6.05	0.74	16.31
	EV1	380.87	15.55	0.08	31.41
6	NOR	378.45	7.24	0.61	19.42
	LN2	377.04	11.75	0.23	16.08
	EV1	378.93	5.81	0.76	33.72
7	NOR	373.06	10.56	0.31	20.92
	LN2	371.51	10.80	0.29	17.75
	EV1	372.97	11.51	0.24	32.60
10	NOR	363.91	4.62	0.87	14.68
	LN2	362.99	6.05	0.74	12.38
	EV1	369.63	7.48	0.59	38.35
Overall rank					
	NOR	2	1		2
	LN2	1	2		1
	EV1	3	2		3

Table A10. Contd.

Station: 7025280					
Heatwave duration (days)	Distribution	Criterion			
		AIC	Chi-Squared		RMSE (%)
			Statistic	*p*-value	
1	NOR	346.55	18.57	0.03	14.84
	LN2	345.70	19.31	0.02	12.87
	EV1	351.98	17.84	0.04	35.81
2	NOR	342.54	2.90	0.97	12.85
	LN2	341.47	2.90	0.97	9.98
	EV1	346.50	8.04	0.53	30.81
3	NOR	348.15	8.53	0.48	19.51
	LN2	346.67	8.04	0.53	17.00
	EV1	347.12	5.10	0.83	30.09
4	NOR	353.18	12.45	0.19	23.53
	LN2	351.99	11.22	0.26	21.70
	EV1	352.20	10.98	0.28	36.49
5	NOR	351.98	8.29	0.51	20.71
	LN2	350.96	8.29	0.51	19.07
	EV1	352.74	7.80	0.55	37.21
6	NOR	354.83	7.55	0.58	15.71
	LN2	354.19	3.88	3.88	14.49
	EV1	359.51	5.84	0.76	40.80
7	NOR	354.04	8.04	0.53	16.82
	LN2	353.10	7.8	0.55	14.98
	EV1	357.55	12.94	0.17	37.75
10	NOR	347.06	11.96	0.22	15.01
	LN2	346.94	15.14	0.09	14.53
	EV1	359.66	11.71	0.23	49.07
Overall rank					
	NOR	2	3		2
	LN2	1	2		1
	EV1	3	1		2

Table A10. Contd.

Station: 7028124					
Heatwave duration (days)	Distribution	Criterion			
		AIC	Chi-Squared		RMSE (%)
			Statistic	*p*-value	
1	NOR	317.29	9.09	0.43	18.90
	LN2	315.79	9.35	0.41	16.35
	EV1	318.94	11.46	0.25	27.83
2	NOR	323.04	4.60	0.87	21.26
	LN2	321.21	6.19	0.72	18.33
	EV1	321.81	9.09	0.43	25.46
3	NOR	327.95	9.35	0.41	21.88
	LN2	326.31	7.24	0.61	18.55
	EV1	324.55	11.20	0.26	25.59
4	NOR	331.28	17.26	0.04	20.25
	LN2	330.42	10.41	0.32	18.30
	EV1	332.74	19.37	0.02	36.69
5	NOR	324.14	11.20	0.26	22.33
	LN2	327.19	10.67	0.30	20.74
	EV1	328.62	14.89	0.09	38.10
6	NOR	332.28	20.96	0.01	21.84
	LN2	331.15	20.96	0.01	19.63
	EV1	332.57	22.80	0.01	35.74
7	NOR	329.03	17.26	0.04	22.13
	LN2	327.47	13.31	0.15	19.15
	EV1	327.04	10.67	0.30	29.51
10	NOR	325.01	13.04	0.16	20.06
	LN2	324.04	13.04	0.16	17.96
	EV1	326.53	22.01	0.01	34.87
Overall rank					
	NOR	2	2		2
	LN2	1	1		1
	EV1	2	3		3

Table A11. Goodness-of-fit analysis of three parametric distributions to heatwave data observed at four studied locations.

Station: 7024280								
Heatwave duration (days)	Criterion	GEV-ML	LN3-ML	P3-ML	GEV-PWM	LN3-MOM	P3-MOM	LP3-BOB
1	AIC	296.13	295.83	295.02	296.21	297.19	296.54	297.00
	Chi-Sq. Stat.	26.69	26.69	28.32	26.69	24.25	23.43	24.25
	p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	AIC	295.39	294.65	293.66	295.57	297.14	295.98	296.73
	Chi-Sq. Stat.	15.01	13.65	13.11	15.01	26.42	28.32	26.42
	p-value	0.04	0.06	0.07	0.04	0.00	0.00	0.00
3	AIC	297.02	296.63	295.62	297.17	298.54	297.91	298.41
	Chi-Sq. Stat.	6.59	5.23	3.06	5.51	7.14	6.86	7.14
	p-value	0.47	0.63	0.88	0.60	0.41	0.44	0.41
4	AIC	298.84	298.97	297.67	299.12	300.21	299.95	300.22
	Chi-Sq. Stat.	7.41	9.04	6.59	5.78	9.58	9.85	15.56
	p-value	0.39	0.25	0.47	0.57	0.21	0.20	0.03
5	AIC	302.79	302.85	301.51	303.06	303.96	303.68	303.97
	Chi-Sq. Stat.	9.31	10.67	9.31	9.04	8.22	8.22	8.22
	p-value	0.23	0.15	0.23	0.25	0.31	0.31	0.31
6	AIC	306.87	306.86	305.59	307.14	307.89	307.57	307.89
	Chi-Sq. Stat.	3.06	4.15	6.86	4.69	3.06	4.69	3.06
	p-value	0.88	0.76	0.44	0.70	0.88	0.70	0.88
7	AIC	303.28	302.98	301.46	303.58	304.26	303.81	304.21
	Chi-Sq. Stat.	11.21	9.04	8.22	7.68	6.32	6.86	7.41
	p-value	0.13	0.25	0.31	0.36	0.50	0.44	0.39
10	AIC	295.15	295.41	294.85	295.32	295.81	295.61	295.81
	Chi-Sq. Stat.	3.33	2.79	1.98	4.15	4.69	3.06	3.06
	p-value	0.85	0.90	0.96	0.76	0.70	0.88	0.88
Overall rank								
	AIC	3	2	1	4	7	5	6
	Chi-Sq.	6	3	1	2	4	4	5

Notes: ML - method of maximum likelihood, PWM - method of probability weighted moments, MOM - method of moments, BOB - Bobée (1975)'s method of parameter estimation
'-': parameter estimation procedure did not converge

Table A11. Contd.

Station: 7016294								
Heatwave duration (days)	Criterion	GEV-ML	LN3-ML	P3-ML	GEV-PWM	LN3-MOM	P3-MOM	LP3-BOB
1	AIC	378.32	378.83	378.77	378.35	378.86	378.82	378.87
	Chi-Sq. Stat.	25.06	25.06	25.06	25.06	25.06	25.06	25.06
	p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	AIC	391.15	391.70	391.66	391.24	391.74	391.71	-
	Chi-Sq. Stat.	23.87	22.21	23.87	15.08	13.89	13.89	-
	p-value	0.00	0.00	0.00	0.06	0.08	0.08	-
3	AIC	389.17	390.11	390.08	389.26	390.14	390.12	390.12
	Chi-Sq. Stat.	9.14	7.95	7.48	9.14	7.95	9.14	9.14
	p-value	0.33	0.44	0.49	0.33	0.44	0.33	0.33
4	AIC	383.30	383.77	383.34	383.44	384.10	383.91	384.12
	Chi-Sq. Stat.	6.76	9.38	7.48	5.57	7.00	5.57	5.57
	p-value	0.56	0.31	0.49	0.69	0.54	0.69	0.69
5	AIC	379.50	379.96	379.41	379.70	380.39	380.19	380.41
	Chi-Sq. Stat.	4.86	7.95	6.52	5.57	4.39	5.81	4.39
	p-value	0.77	0.44	0.59	0.69	0.82	0.67	0.82
6	AIC	377.02	377.73	377.19	377.23	378.19	378.02	378.21
	Chi-Sq. Stat.	8.90	8.43	5.57	5.81	5.57	7.24	6.29
	p-value	0.35	0.39	0.69	0.67	0.69	0.51	0.62
7	AIC	371.30	371.93	371.40	371.47	372.37	372.19	372.40
	Chi-Sq. Stat.	15.08	13.42	13.89	10.33	15.32	13.18	11.28
	p-value	0.06	0.10	0.08	0.24	0.05	0.11	0.19
10	AIC	363.77	364.67	364.60	364.02	364.75	364.70	364.76
	Chi-Sq. Stat.	10.33	8.19	9.61	6.05	8.90	8.43	7.95
	p-value	0.24	0.42	0.29	0.64	0.35	0.39	0.44
Overall rank								
	AIC	1	4	2	3	6	5	7
	Chi-Sq.	6	5	4	1	3	3	2

Table A11. Contd.

Station: 7025280								
Heatwave duration (days)	Criterion	GEV-ML	LN3-ML	P3-ML	GEV-PWM	LN3-MOM	P3-MOM	LP3-BOB
1	AIC	346.70	347.24	347.18	346.80	347.28	347.24	347.29
	Chi-Sq. Stat.	11.22	11.22	11.22	11.22	11.22	11.22	11.22
	p-value	0.19	0.19	0.19	0.19	0.19	0.19	0.19
2	AIC	342.42	342.67	342.60	342.44	342.71	342.65	342.71
	Chi-Sq. Stat.	3.63	2.90	3.63	3.63	3.88	3.63	3.88
	p-value	0.89	0.94	0.89	0.89	0.87	0.89	0.87
3	AIC	345.77	346.44	346.08	346.02	346.70	346.90	346.89
	Chi-Sq. Stat.	4.12	5.35	3.14	4.37	3.39	2.41	3.39
	p-value	0.85	0.72	0.93	0.82	0.91	0.97	0.91
4	AIC	351.19	352.06	351.00	351.58	353.00	352.82	353.03
	Chi-Sq. Stat.	10.49	7.06	7.31	6.33	11.47	11.96	11.47
	p-value	0.23	0.53	0.50	0.61	0.18	0.15	0.18
5	AIC	350.52	351.76	351.17	350.99	352.36	352.23	352.39
	Chi-Sq. Stat.	9.02	11.47	8.78	10.98	9.51	10.49	10.49
	p-value	0.34	0.18	0.36	0.20	0.30	0.23	0.23
6	AIC	354.36	355.93	355.76	354.86	356.10	356.05	356.10
	Chi-Sq. Stat.	3.63	3.39	2.16	3.14	3.39	2.65	3.63
	p-value	0.89	0.91	0.98	0.93	0.91	0.95	0.89
7	AIC	353.42	354.53	354.35	353.72	354.73	354.65	354.74
	Chi-Sq. Stat.	11.47	12.94	15.39	15.39	11.47	10.00	10.00
	p-value	0.18	0.11	0.05	0.05	0.18	0.27	0.27
10	AIC	348.00	348.90	348.90	348.55	348.91	348.91	348.90
	Chi-Sq. Stat.	12.69	11.96	11.22	16.37	11.96	11.22	11.22
	p-value	0.12	0.15	0.19	0.04	0.15	0.19	0.19
Overall rank								
	AIC	1	3	2	2	5	4	6
	Chi-Sq.	3	4	1	4	3	2	3

Table A11. Contd.

Station: 7028124								
Heatwave duration (days)	Criterion	GEV-ML	LN3-ML	P3-ML	GEV-PWM	LN3-MOM	P3-MOM	LP3-BOB
1	AIC	316.32	316.26	316.27	316.33	316.26	316.29	316.26
	Chi-Sq. Stat.	11.99	11.99	19.11	11.99	11.99	15.15	11.99
	p-value	0.15	0.15	0.01	0.15	0.15	0.06	0.15
2	AIC	321.02	320.86	320.70	321.29	320.88	320.74	320.84
	Chi-Sq. Stat.	8.03	9.62	6.45	11.20	10.41	10.41	7.24
	p-value	0.43	0.29	0.60	0.19	0.24	0.24	0.51
3	AIC	324.98	325.14	324.06	325.21	326.10	325.78	326.08
	Chi-Sq. Stat.	7.77	6.19	8.30	6.71	6.45	6.19	5.92
	p-value	0.46	0.63	0.41	0.57	0.60	0.63	0.66
4	AIC	330.59	331.65	331.13	330.90	332.09	331.99	332.11
	Chi-Sq. Stat.	16.74	12.78	13.84	11.99	10.67	14.36	11.20
	p-value	0.03	0.12	0.09	0.15	0.22	0.07	0.19
5	AIC	326.95	328.08	327.38	327.39	328.72	328.60	328.74
	Chi-Sq. Stat.	12.25	9.09	9.35	13.31	11.20	11.20	11.20
	p-value	0.14	0.33	0.31	0.10	0.19	0.19	0.19
6	AIC	331.06	331.94	331.37	331.35	332.49	332.35	332.52
	Chi-Sq. Stat.	14.89	17.53	17.79	18.32	19.11	19.90	17.79
	p-value	0.06	0.03	0.02	0.02	0.01	0.01	0.02
7	AIC	326.80	327.17	326.44	327.00	327.87	327.62	327.88
	Chi-Sq. Stat.	12.78	11.20	11.73	11.73	12.52	13.57	13.84
	p-value	0.12	0.19	0.16	0.16	0.13	0.09	0.09
10	AIC	324.36	325.26	324.78	324.58	325.63	325.52	325.65
	Chi-Sq. Stat.	11.20	12.52	10.93	10.14	10.93	10.14	10.93
	p-value	0.19	0.13	0.21	0.26	0.21	0.26	0.21
Overall rank								
	AIC	2	3	1	4	6	5	7
	Chi-Sq.	5	1	3	4	3	6	2

Table A12. Goodness-of-fit analysis using root mean square error (RMSE) criterion. The RMSE values are given for those distributions and estimation procedures which ranked best on the basis of AIC.

Station: 7024280									
Distribution	Heatwave duration (days)								Overall rank
	1	2	3	4	5	6	7	10	
GEV-ML	19.28	21.44	19.93	17.63	17.26	14.97	16.59	11.76	2
LN3-ML	19.97	20.77	21.31	21.67	20.36	17.72	19.13	12.44	3
P3-ML	18.11	17.02	54.92	23.02	22.68	19.86	20.48	13.06	3
GEV-PWM	19.43	20.18	20.09	17.50	16.81	14.37	15.83	11.39	1

Table A12. Contd.

Station: 7016294									
Distribution	Heatwave duration (days)								Overall rank
	1	2	3	4	5	6	7	10	
GEV-ML	18.08	13.63	10.23	11.44	12.60	13.16	14.64	10.98	2
LN3-ML	18.46	13.73	11.06	12.58	13.81	14.91	16.09	11.86	3
P3-ML	18.42	13.71	11.05	12.70	14.28	15.52	16.40	11.75	3
GEV-PWM	17.99	13.80	9.91	11.04	12.29	12.83	14.41	10.68	1

Table A12. Contd.

Station: 7025280									
Distribution	Heatwave duration (days)								Overall rank
	1	2	3	4	5	6	7	10	
GEV-ML	11.78	7.69	13.33	19.12	16.63	12.39	13.03	14.19	2
LN3-ML	11.80	7.17	15.42	23.37	20.20	14.99	14.95	14.58	3
P3-ML	32.48	29.46	34.06	36.25	36.22	31.23	31.98	26.36	4
GEV-PWM	11.49	7.54	13.90	19.64	17.28	12.32	12.96	14.02	1

Table A12. Contd.

Station: 7028124									
Distribution	Heatwave duration (days)								Overall rank
	1	2	3	4	5	6	7	10	
GEV-ML	13.13	14.04	13.57	16.36	18.41	17.31	15.74	16.10	1
LN3-ML	12.48	13.37	15.80	18.50	22.21	19.81	17.67	17.64	2
P3-ML	61.72	60.75	59.45	55.28	54.60	51.40	52.14	48.47	3
GEV-PWM	13.34	14.57	13.25	16.22	18.87	17.39	15.49	15.87	1

Table A13. Overall ranks after combining the results of two parametric and three parametric best fitting distributions.

Distribution	Criterion	Station			
		7024280	7016294	7025280	7028124
LN2-ML	AIC	4	2	2	2
	RMSE	5	4	4	3
GEV-ML	AIC	2	1	1	1
	RMSE	2	2	2	1
LN3-ML	AIC	2	5	5	4
	RMSE	3	3	3	2
P3-ML	AIC	1	3	4	2
	RMSE	4	3	5	4
GEV-PWM	AIC	3	4	3	3
	RMSE	1	1	1	1

Table A14. Estimated parameters of various forms of $\mu(D)$ function along with goodness-of-fit measures.

Station: 7024280					
$f_i(D)$	a_i	b_i	Range of D	R^2	rRMSE (%)
$f_1(D)$	33.047	-0.058	1- to 10-day	0.974	0.657
$f_2(D)$	32.688	-0.051	1- to 10-day	0.959	0.822
$f_3(D)$	33.056	-0.058	1- to 10-day	-	0.657
$f_4(D)$	33.109	16.106	1- to 10-day	0.968	0.740
$f_5(D)$	33.596	-0.067	2 - to 10-day	0.994	0.238

Table A14. Contd.

Station: 7016294					
$f_i(D)$	a_i	b_i	Range of D	R^2	rRMSE
$f_1(D)$	32.343	-0.063	1- to 10-day	0.974	0.717
$f_2(D)$	31.960	-0.056	1- to 10-day	0.959	0.896
$f_3(D)$	32.351	-0.063	1- to 10-day	-	0.717
$f_4(D)$	32.415	14.638	1- to 10-day	0.968	0.814
$f_5(D)$	32.930	-0.073	2 - to 10-day	0.994	0.263

Table A14. Contd.

Station: 7025280					
$f_i(D)$	a_i	b_i	Range of D	R^2	rRMSE
$f_1(D)$	33.143	-0.055	1- to 10-day	0.977	0.586
$f_2(D)$	32.830	-0.049	1- to 10-day	0.964	0.728
$f_3(D)$	33.152	-0.055	1- to 10-day	-	0.587
$f_4(D)$	33.199	17.013	1- to 10-day	0.971	0.663
$f_5(D)$	33.622	-0.063	2 - to 10-day	0.993	0.244

Table A14. Contd.

Station: 7028124					
$f_i(D)$	a_i	b_i	Range of D	R^2	rRMSE
$f_1(D)$	32.104	-0.062	1- to 10-day	0.981	0.597
$f_2(D)$	31.796	-0.056	1- to 10-day	0.971	0.740
$f_3(D)$	32.116	-0.062	1- to 10-day	-	0.597
$f_4(D)$	32.172	14.901	1- to 10-day	0.975	0.696
$f_5(D)$	32.576	-0.071	2 - to 10-day	0.995	0.250

Table A15. Results of Spearman's rank correlation and Mann-Kendall tests for starting dates of first occurrences of heatwaves.

Station: 7024280				
Heatwave duration (days)	Spearman rank correlation test		Mann-Kendall test	
	Test statistic	*p*-value	Test statistic	*p*-value
1	-0.163	0.146	-1.428	0.153
2	-0.173	0.123	-1.563	0.118
3	-0.263	0.018	-2.354	0.019
4	-0.113	0.317	-1.004	0.316
5	0.007	0.953	0.020	0.984
6	0.097	0.389	0.706	0.480
7	0.108	0.339	0.783	0.433
10	0.086	0.444	0.649	0.517

Note: A small *p*-value indicates a significant correlation, i.e. there is a trend in the time series (e.g. *p*-value<= 0.05 means significant correlation at significance level alpha=5%)

Table A15. Contd.

Station: 7016294				
Heatwave duration (days)	Spearman rank correlation test		Mann-Kendall test	
	Test statistic	*p*-value	Test statistic	*p*-value
1	0.001	0.995	0.050	0.960
2	-0.077	0.445	-0.740	0.459
3	-0.069	0.491	-0.705	0.481
4	-0.143	0.154	-1.409	0.159
5	-0.197	0.049	-1.920	0.055
6	-0.041	0.685	-0.385	0.701
7	-0.024	0.809	-0.188	0.851
10	-0.012	0.908	-0.103	0.918

Table A15. Contd.

Station: 7025280				
Heatwave duration (days)	Spearman rank correlation test		Mann-Kendall test	
	Test statistic	*p*-value	Test statistic	*p*-value
1	-0.163	0.108	-1.551	0.121
2	-0.155	0.128	-1.412	0.158
3	-0.055	0.589	-0.500	0.617
4	0.046	0.653	0.408	0.683
5	0.023	0.820	0.196	0.844
6	0.072	0.484	0.614	0.539
7	0.139	0.172	1.357	0.175
10	0.156	0.124	1.477	0.140

Table A15. Contd.

Station: 7028124				
Heatwave duration (days)	Spearman rank correlation test		Mann-Kendall test	
	Test statistic	*p*-value	Test statistic	*p*-value
1	-0.091	0.390	-0.785	0.432
2	-0.110	0.300	-0.926	0.354
3	-0.123	0.246	-1.180	0.238
4	-0.070	0.511	-0.665	0.506
5	-0.036	0.735	-0.333	0.739
6	0.085	0.424	0.652	0.515
7	0.081	0.444	0.610	0.542
10	0.114	0.283	0.881	0.378

Table A16. Linear regression and Sen's slope estimates for starting dates of first occurrences of heatwaves.

Station: 7024280						
Heatwave duration (days)	Linear slope			Sen slope		
	Slope	95% LL	95% UL	Slope	95% LL	95% UL
1	-0.226	-0.490	0.039	-0.202	-0.488	0.081
2	-0.207	-0.460	0.045	-0.200	-0.432	0.050
3	-0.326	-0.594	-0.058	-0.315	-0.581	-0.062
4	-0.159	-0.418	0.100	-0.147	-0.419	0.133
5	-0.007	-0.247	0.234	0.000	-0.263	0.250
6	0.080	-0.147	0.307	0.084	-0.143	0.319
7	0.089	-0.125	0.304	0.077	-0.125	0.296
10	0.075	-0.129	0.279	0.068	-0.138	0.267

LL: lower limit, UL: upper limit

Table A16. Contd.

Station: 7016294						
Heatwave duration (days)	Linear slope			Sen slope		
	Slope	95% LL	95% UL	Slope	95% LL	95% UL
1	-0.028	-0.179	0.124	0.000	-0.155	0.148
2	-0.069	-0.224	0.086	-0.054	-0.219	0.092
3	-0.070	-0.214	0.074	-0.048	-0.203	0.093
4	-0.116	-0.253	0.021	-0.096	-0.233	0.033
5	-0.154	-0.280	-0.028	-0.125	-0.256	0.000
6	-0.032	-0.155	0.091	-0.016	-0.154	0.091
7	-0.042	-0.171	0.087	0.000	-0.149	0.114
10	-0.025	-0.138	0.089	0.000	-0.125	0.105

Table A16. Contd.

Station: 7025280						
Heatwave duration (days)	Linear slope			Sen slope		
	Slope	95% LL	95% UL	Slope	95% LL	95% UL
1	-0.154	-0.319	0.011	-0.135	-0.316	0.039
2	-0.130	-0.294	0.034	-0.122	-0.284	0.048
3	-0.064	-0.244	0.117	-0.044	-0.228	0.125
4	0.012	-0.164	0.188	0.036	-0.150	0.182
5	-0.008	-0.184	0.168	0.015	-0.178	0.183
6	0.026	-0.134	0.185	0.054	-0.125	0.200
7	0.073	-0.076	0.221	0.105	-0.054	0.238
10	0.070	-0.068	0.209	0.106	-0.037	0.222

Table A16. Contd.

Station: 7028124						
Heatwave duration (days)	Linear slope			Sen slope		
	Slope	95% LL	95% UL	Slope	95% LL	95% UL
1	-0.106	-0.335	0.123	-0.086	-0.316	0.134
2	-0.118	-0.331	0.096	-0.103	-0.327	0.119
3	-0.150	-0.375	0.075	-0.148	-0.387	0.092
4	-0.089	-0.301	0.123	-0.073	-0.313	0.129
5	-0.041	-0.241	0.158	-0.026	-0.260	0.167
6	0.062	-0.124	0.248	0.065	-0.125	0.250
7	0.049	-0.131	0.229	0.045	-0.126	0.231
10	0.082	-0.090	0.255	0.083	-0.089	0.248

8.0 APPENDIX B

This appendix contains figures referred to in the study.

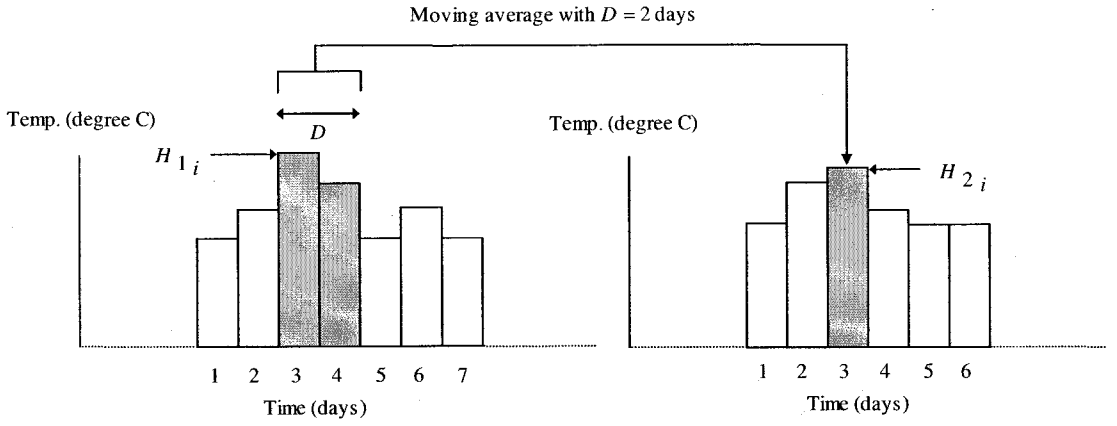


Figure B1. A definition sketch for extraction of heatwaves of 1- and 2-day durations. H_{1_i} is the highest 1-day temperature and H_{2_i} is the highest 2-day average temperature during any year *i*.

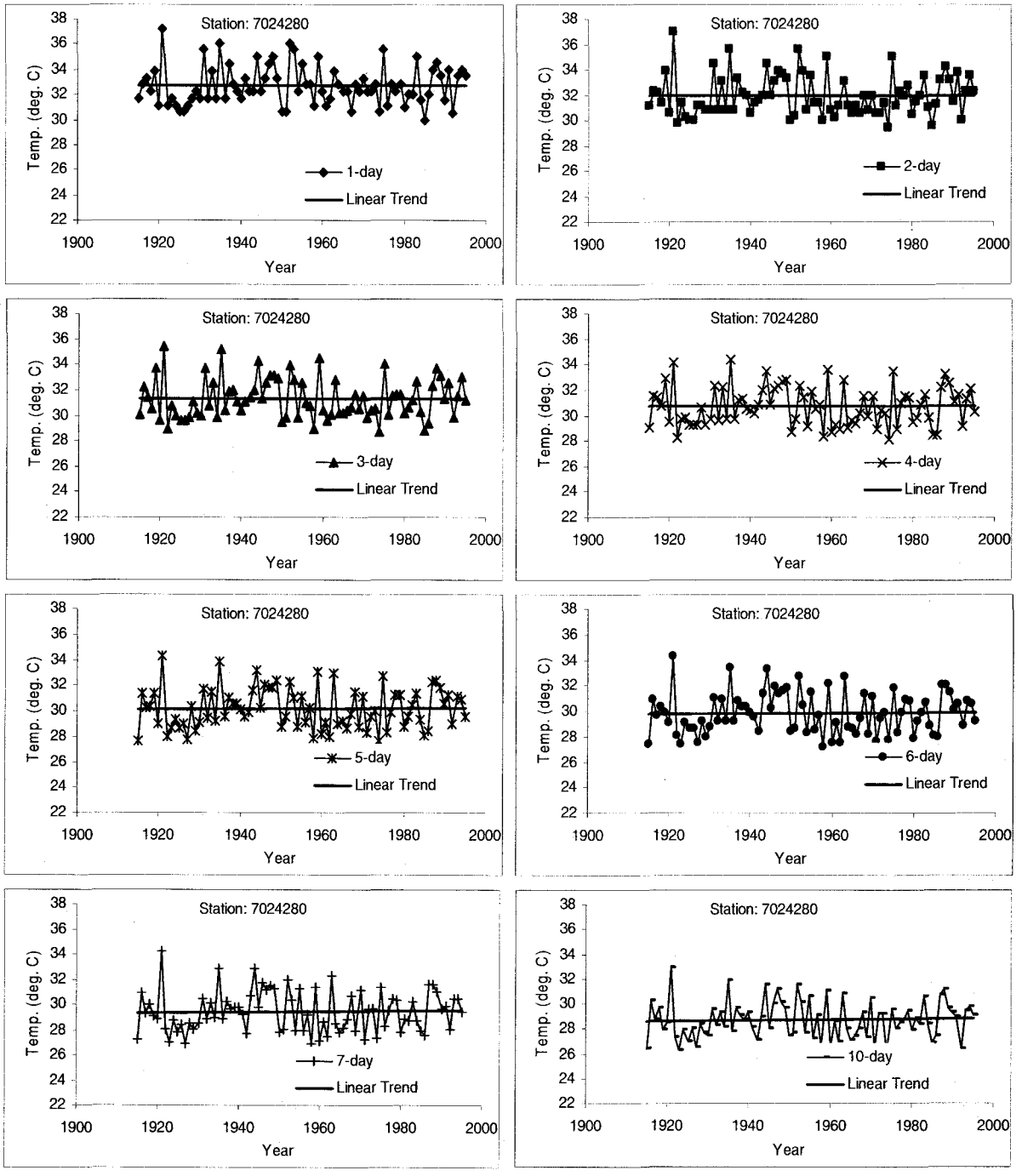


Figure B2. Time series plots of heatwaves (H_D series) of various durations along with linear trends observed at four studied stations.

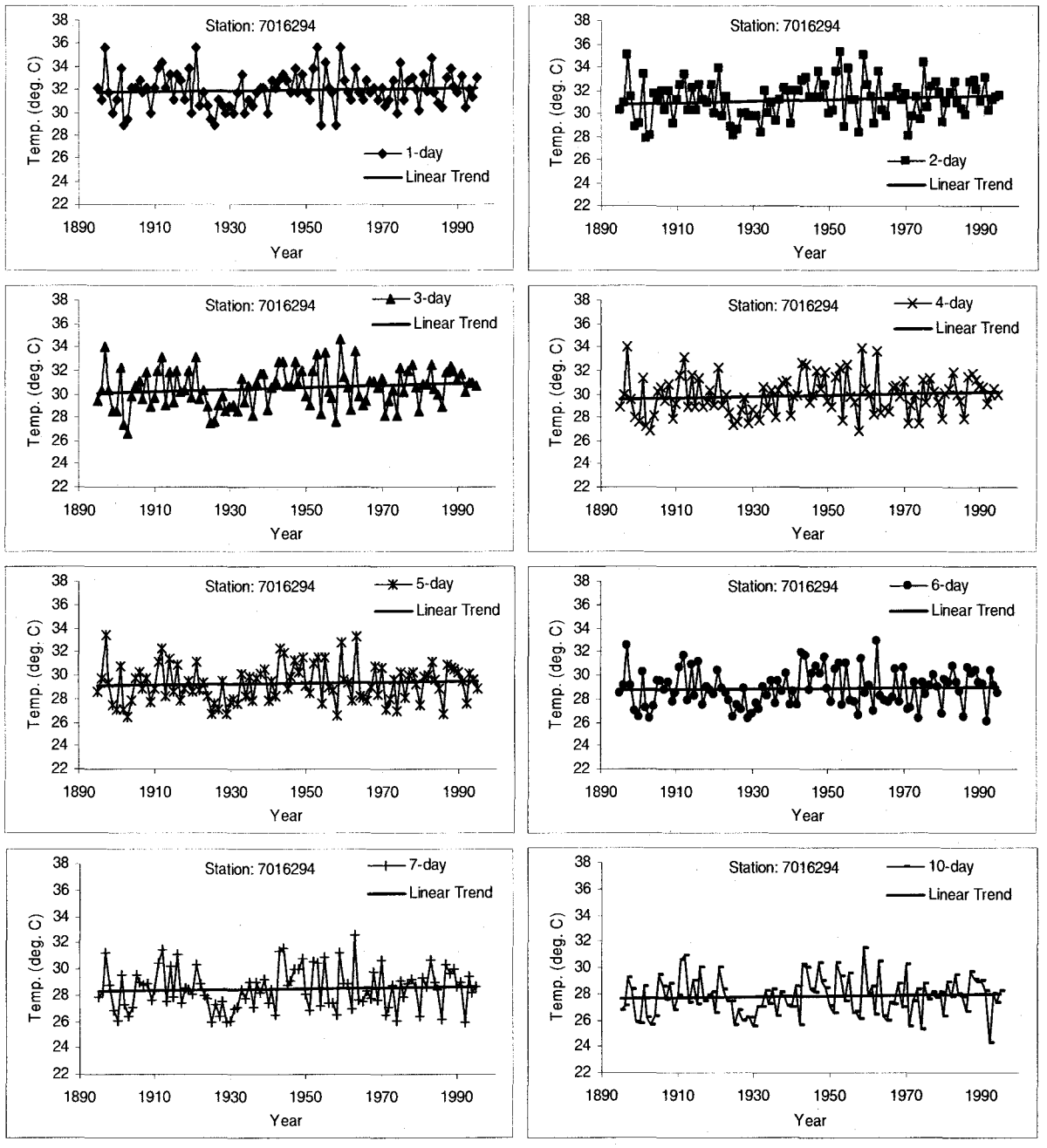


Figure B2. Contd.

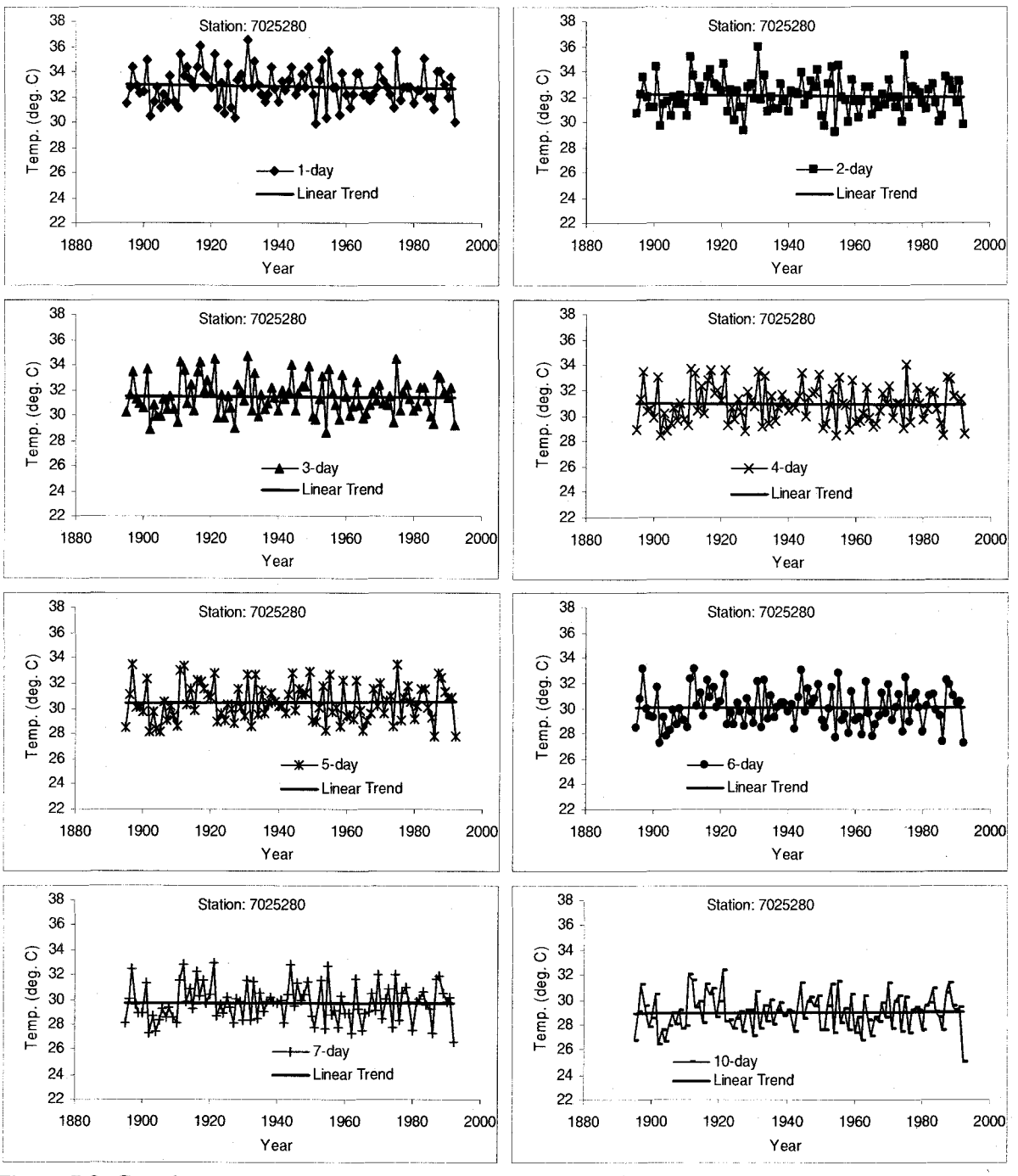


Figure B2. Contd.

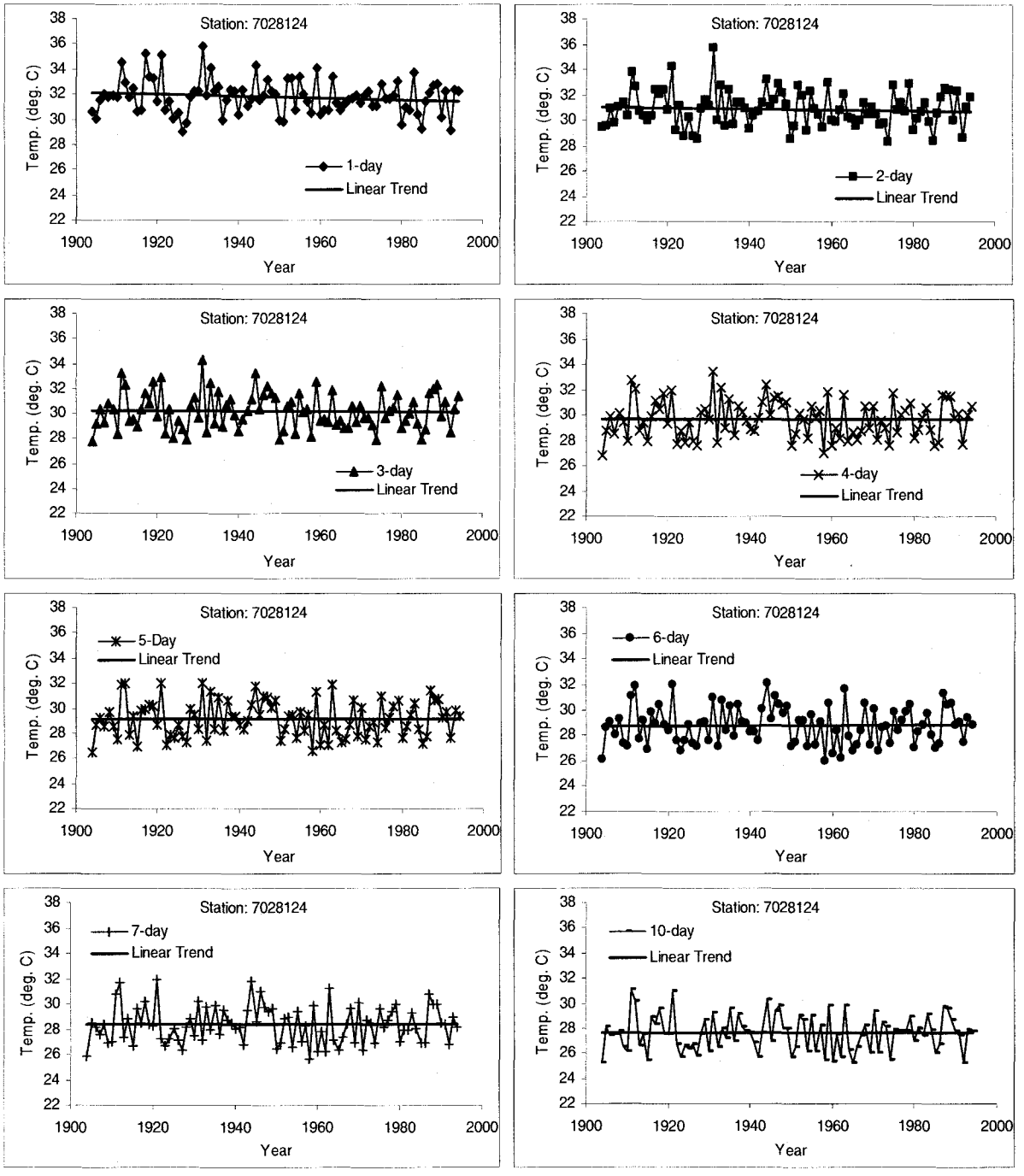


Figure B2. Contd.

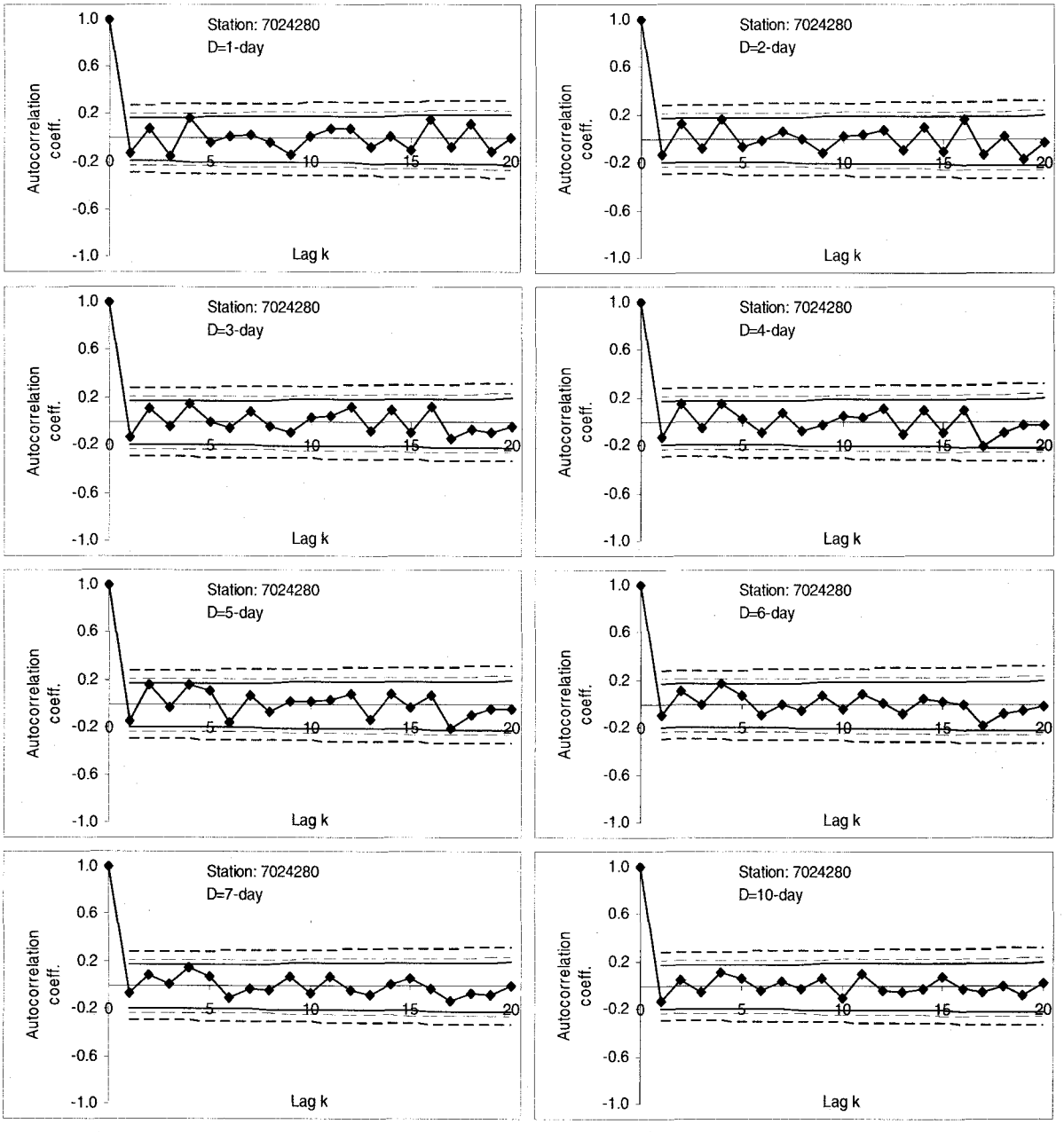


Figure B3. Correlograms of heatwaves of various durations. Solid, small dotted and large dotted lines represent 90%, 95% and 99% confidence bands, respectively.

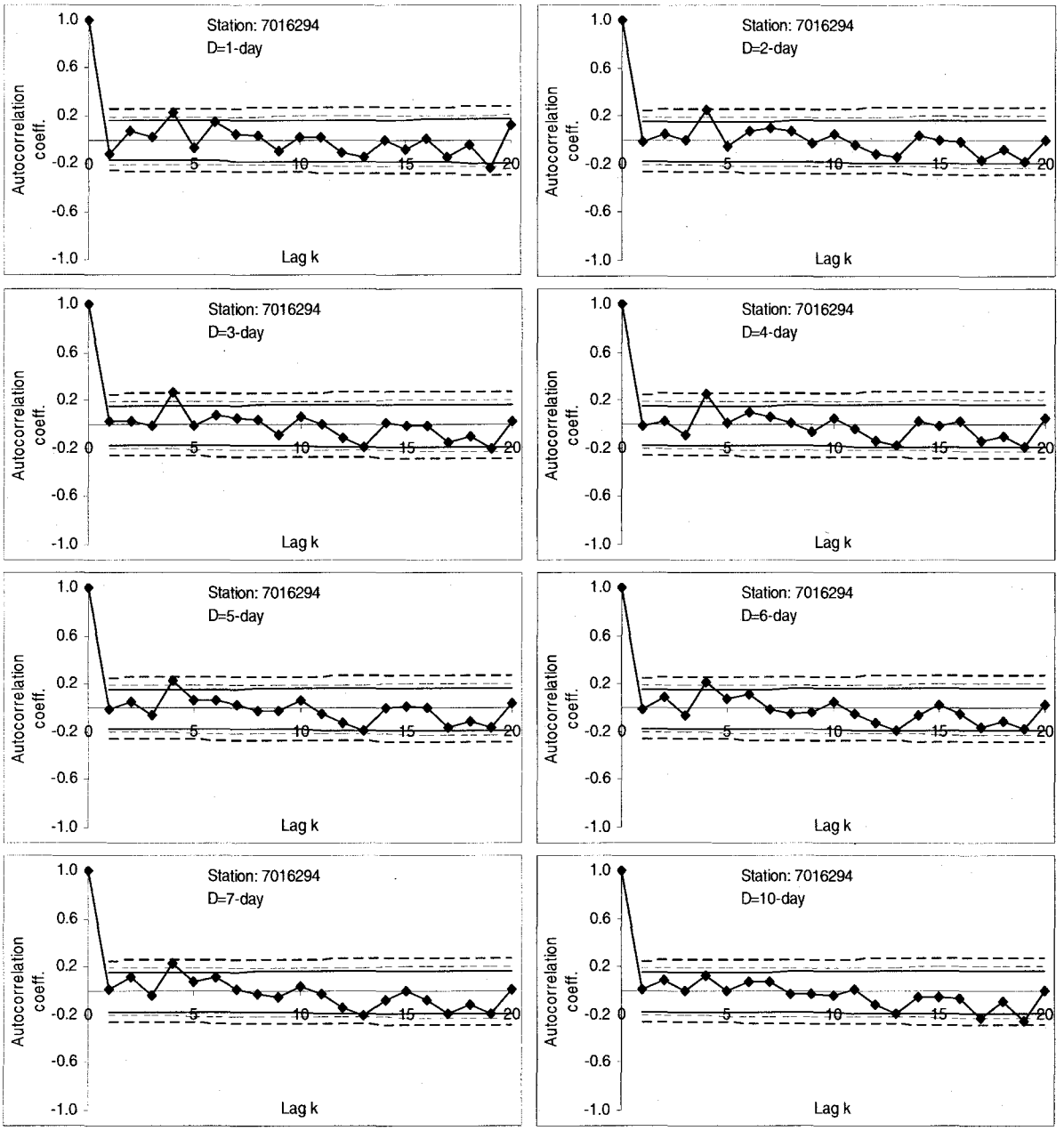


Figure B3. Contd.

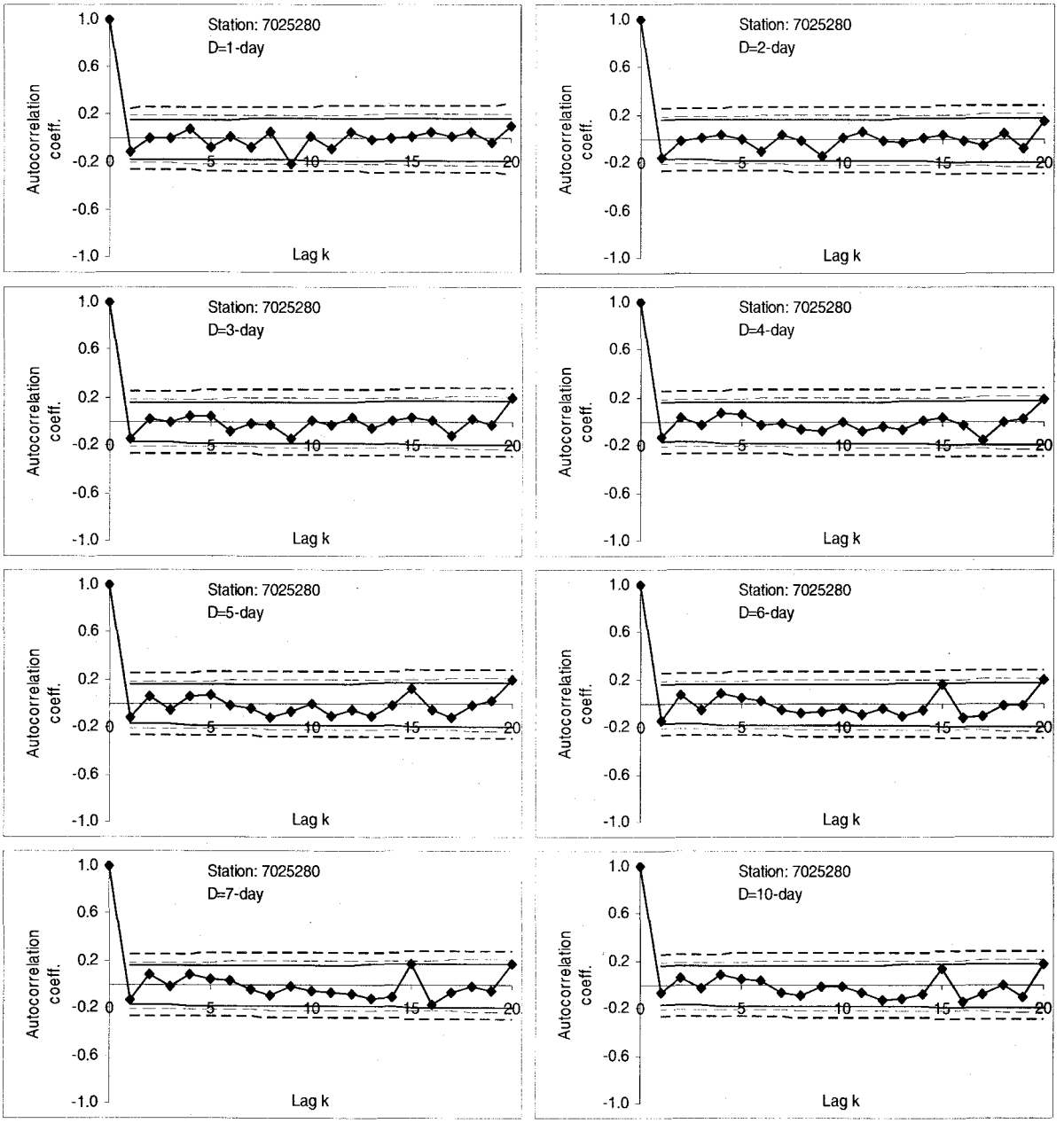


Figure B3. Contd.

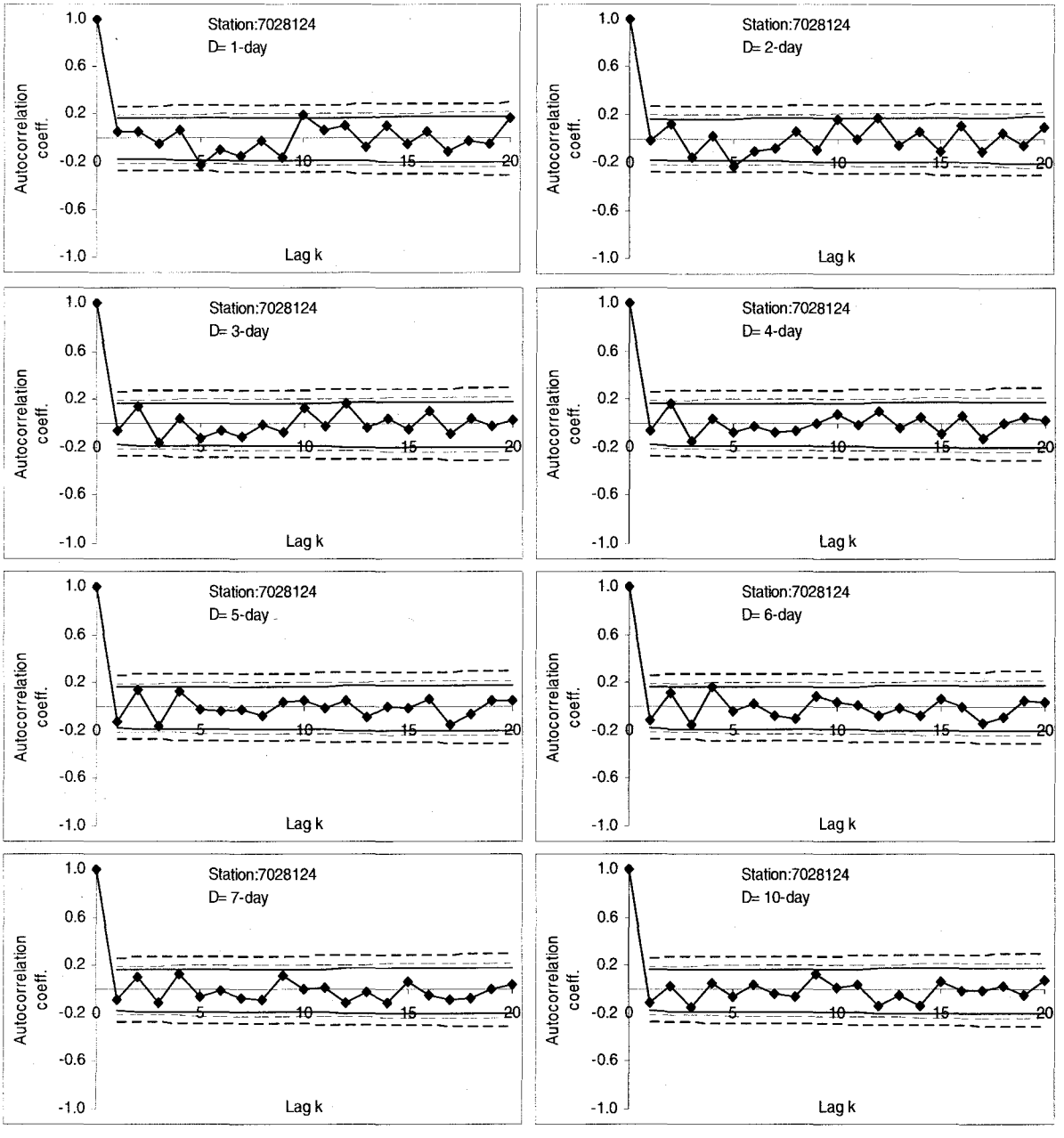


Figure B3. Contd.

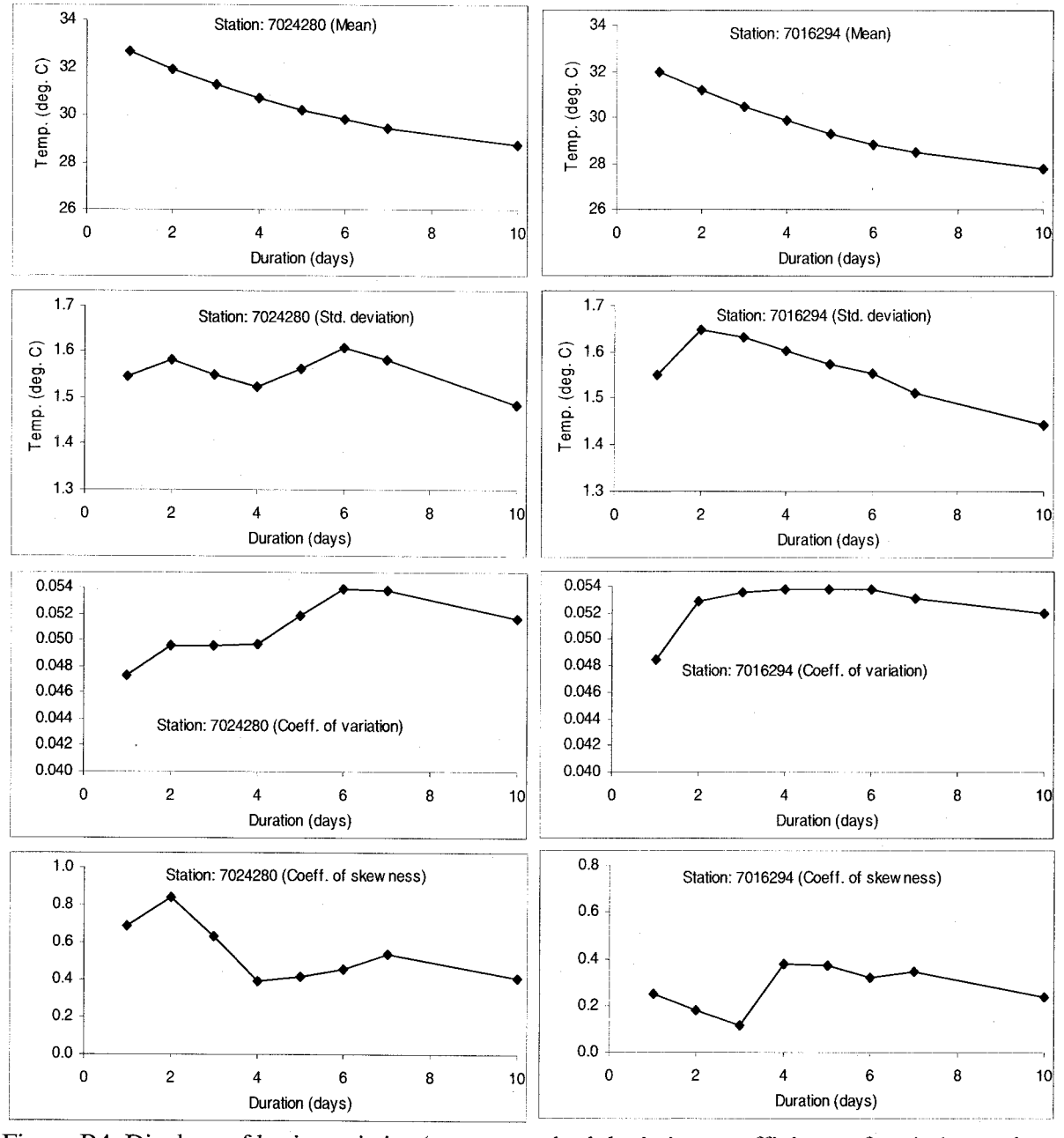


Figure B4. Displays of basic statistics (mean, standard deviation, coefficients of variation and skewness) of heatwaves (H_D series) observed at four studied stations.

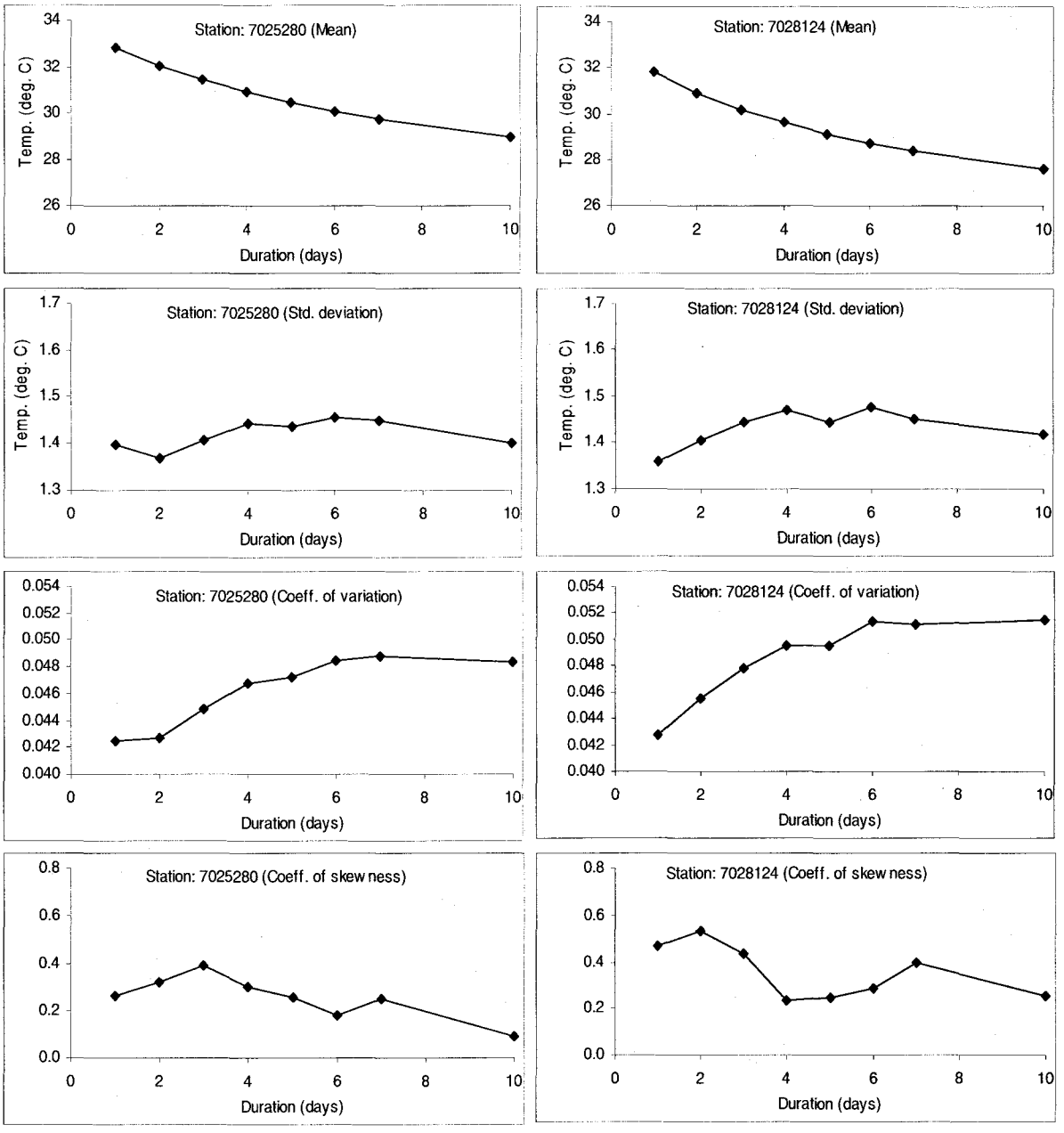


Figure B4. Contd.

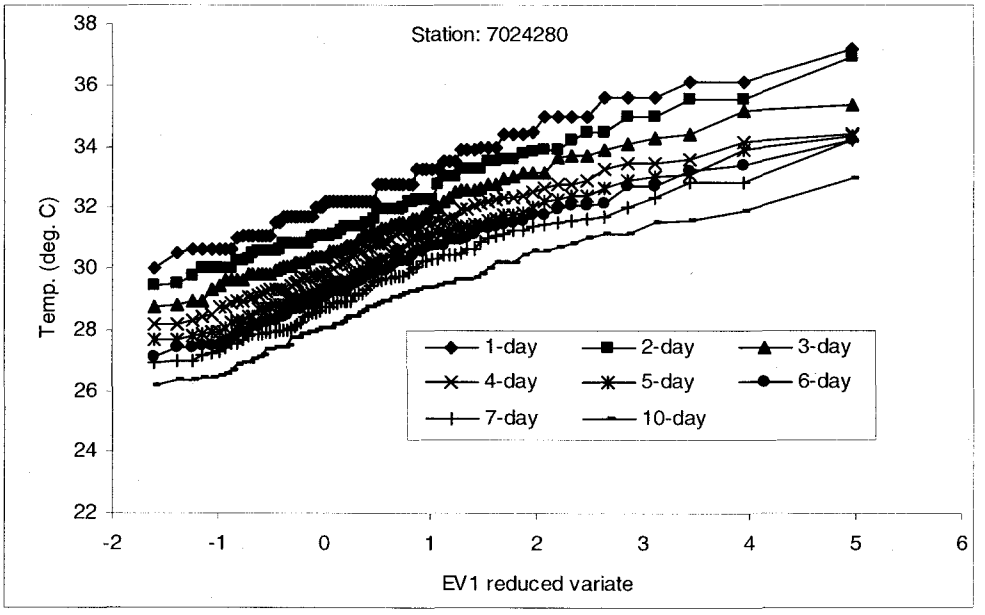


Figure B5. Extreme value plots of heatwaves of 1-day to 10-day durations.

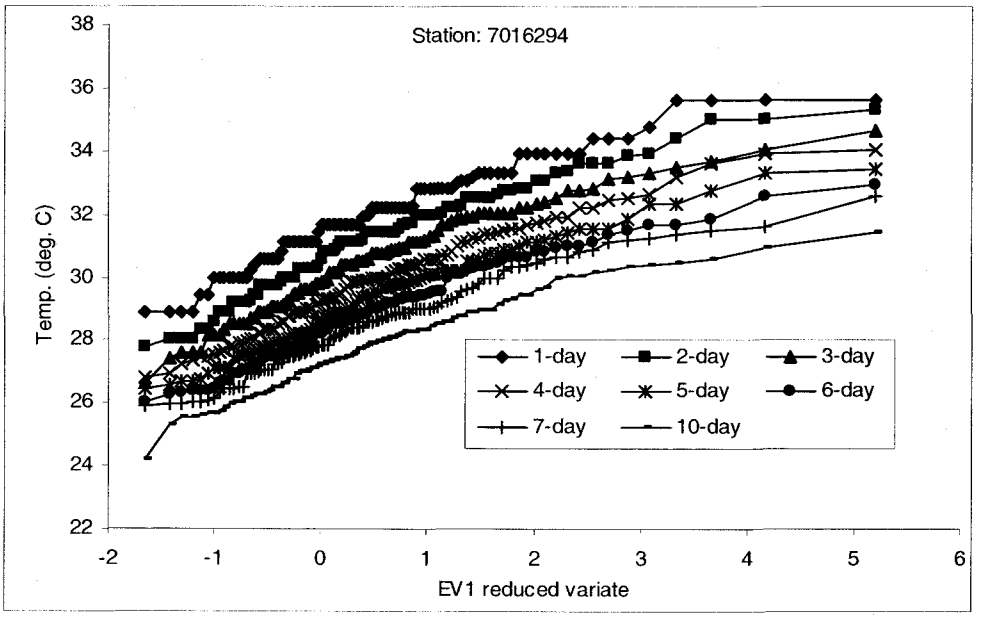


Figure B5. Contd.

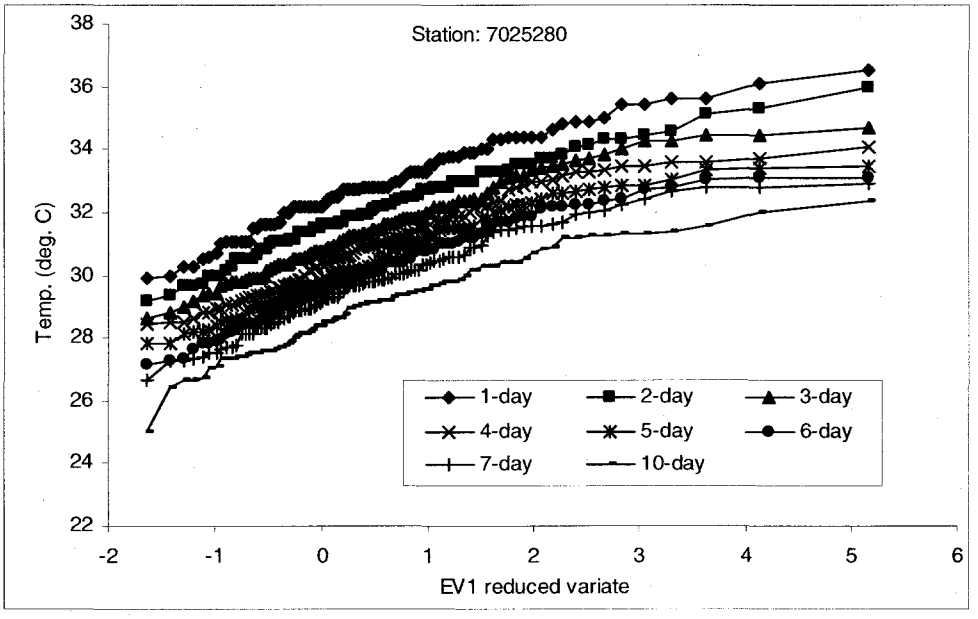


Figure B5. Contd.

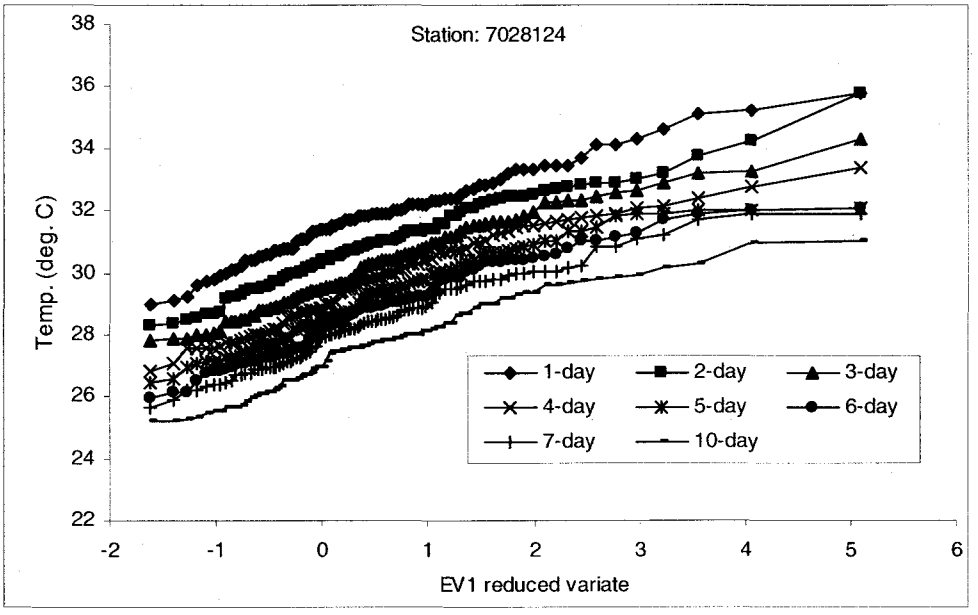


Figure B5. Contd.

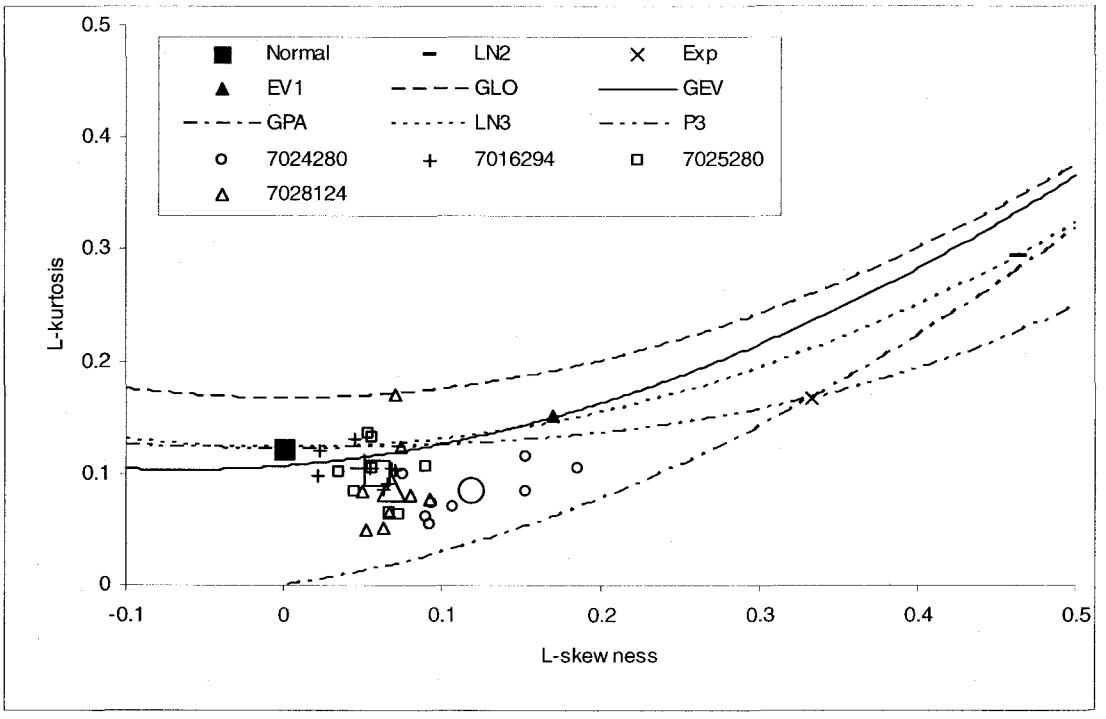


Figure B6. L-moment ratio diagram of heatwaves (H_D series) for four studied stations. Small symbols (open circle, cross, open square and open triangle) represent observed values and corresponding large symbols represent group averages.

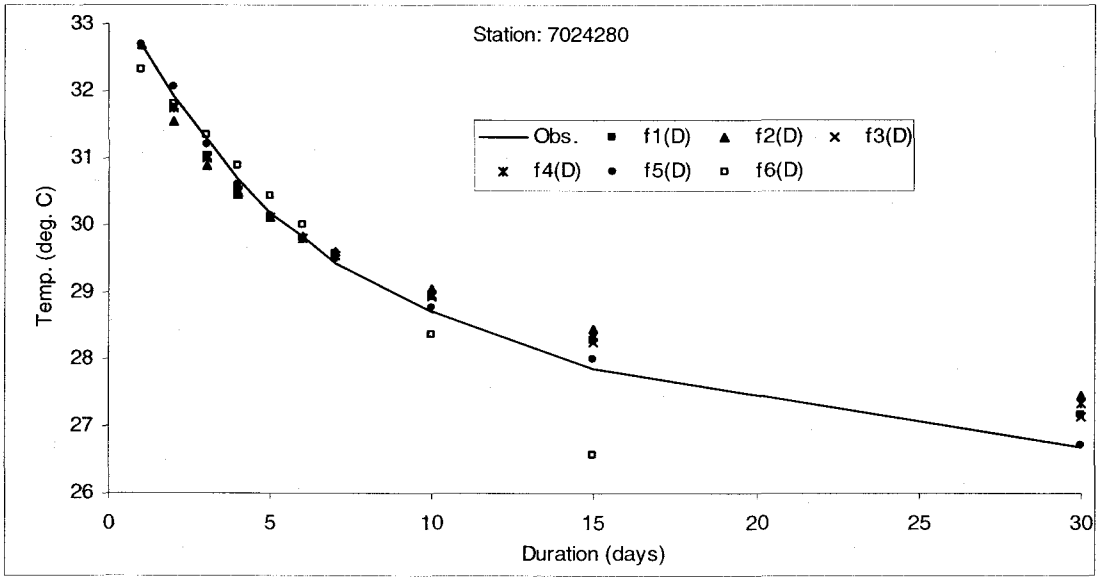


Figure B7. At-site observed mean heatwaves and fitted $\mu(D)$ functions. Fitting of $\mu(D)$ functions consisted of mean heatwaves of 1 to 10 days duration.

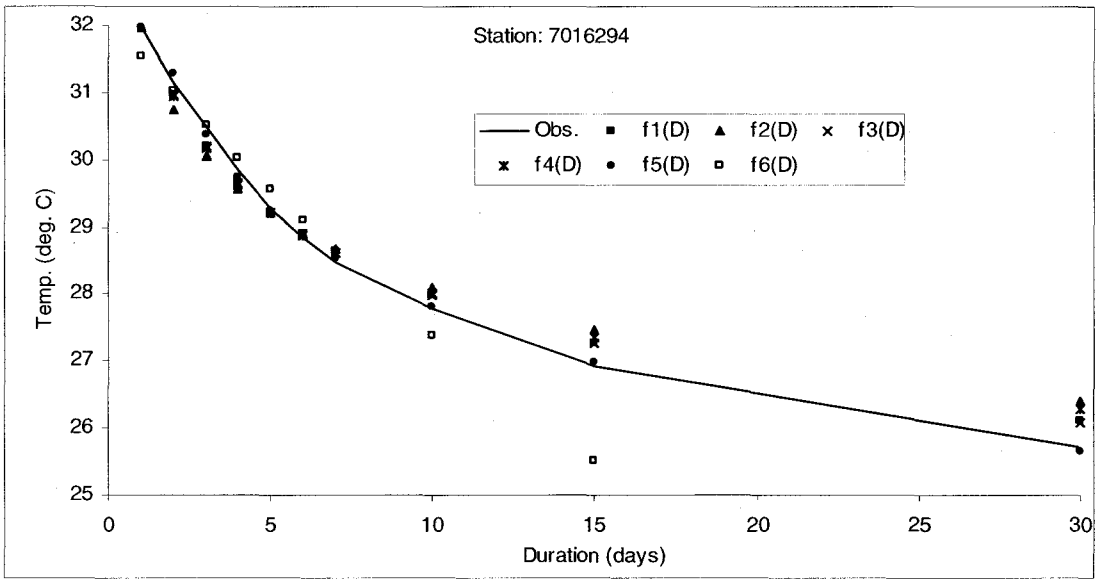


Figure B7. Contd.

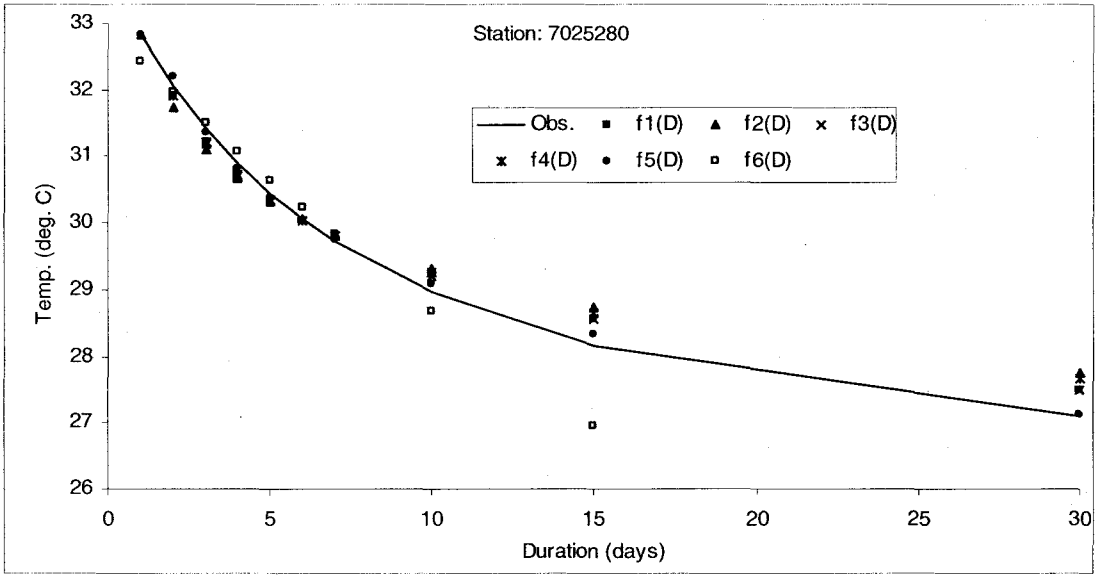


Figure B7. Contd.

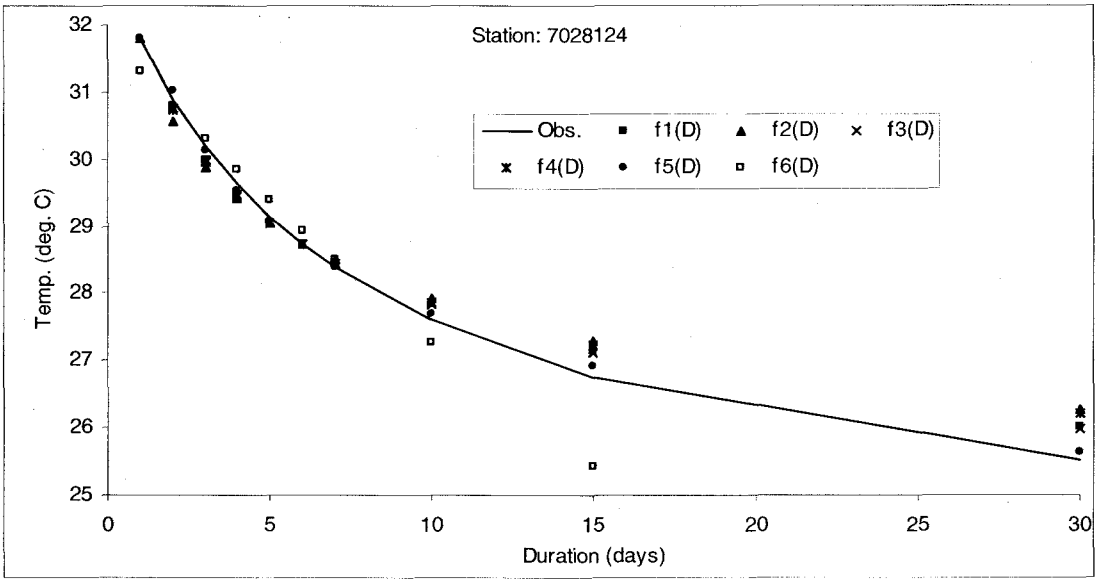


Figure B7. Contd.

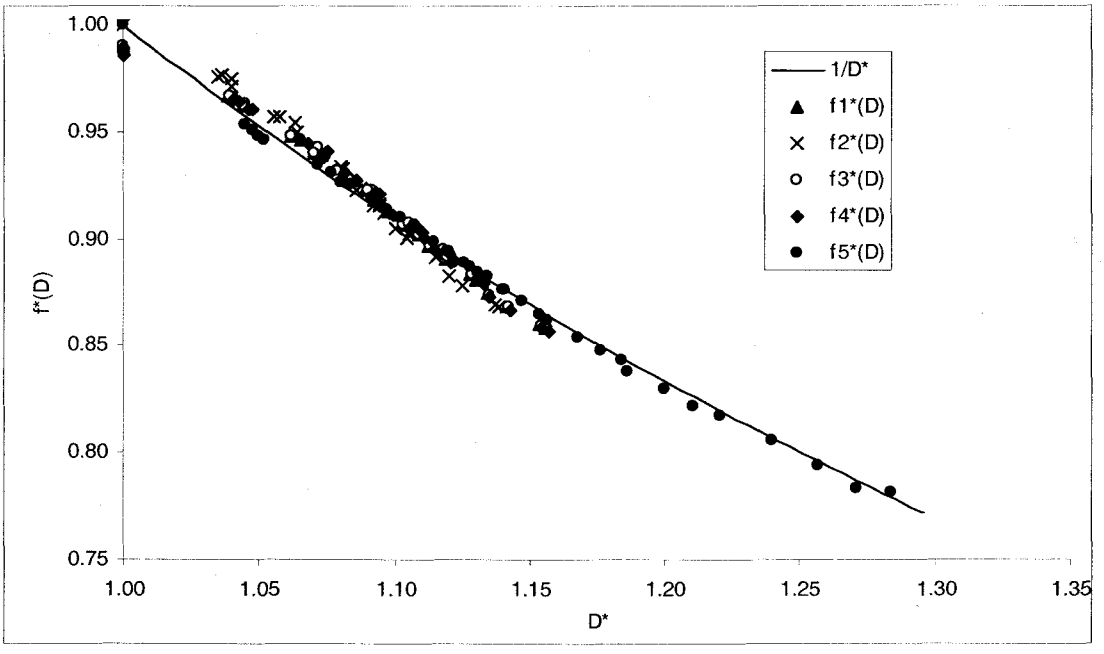


Figure B8. Site independent comparison of fitted $\mu(D)$ functions. The closer the points are to the $1/D^*$ line the better is the fitting.

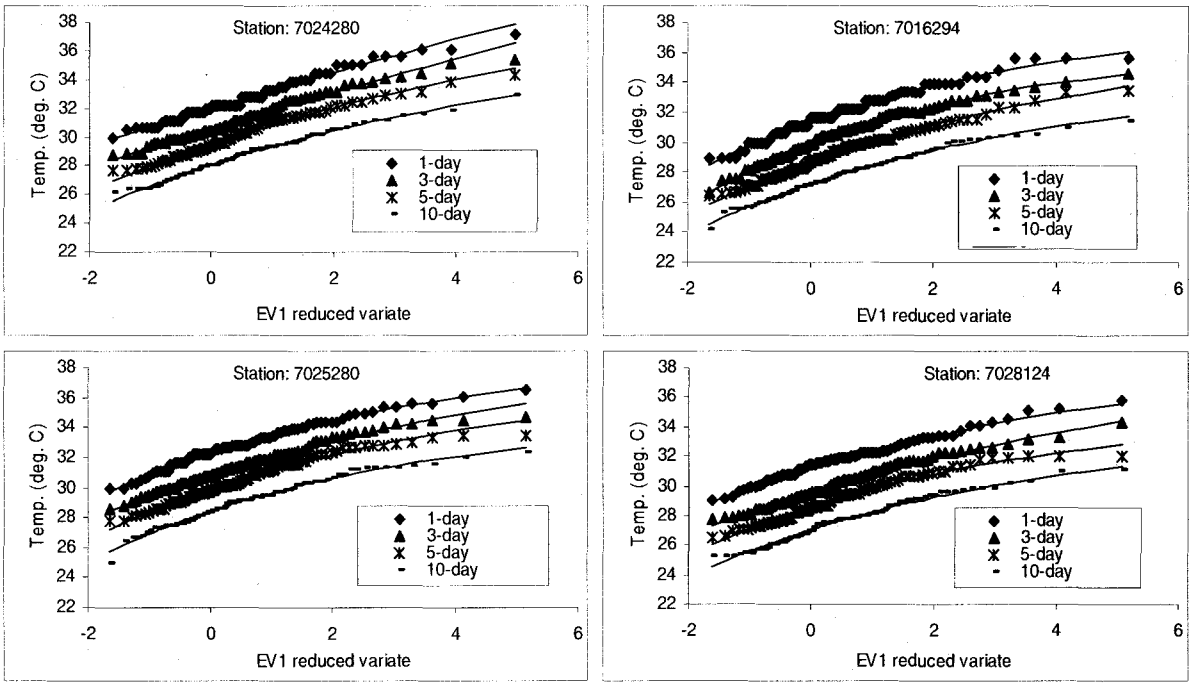


Figure B9. Fitting of GEV distribution to a selected set of observed heatwaves.

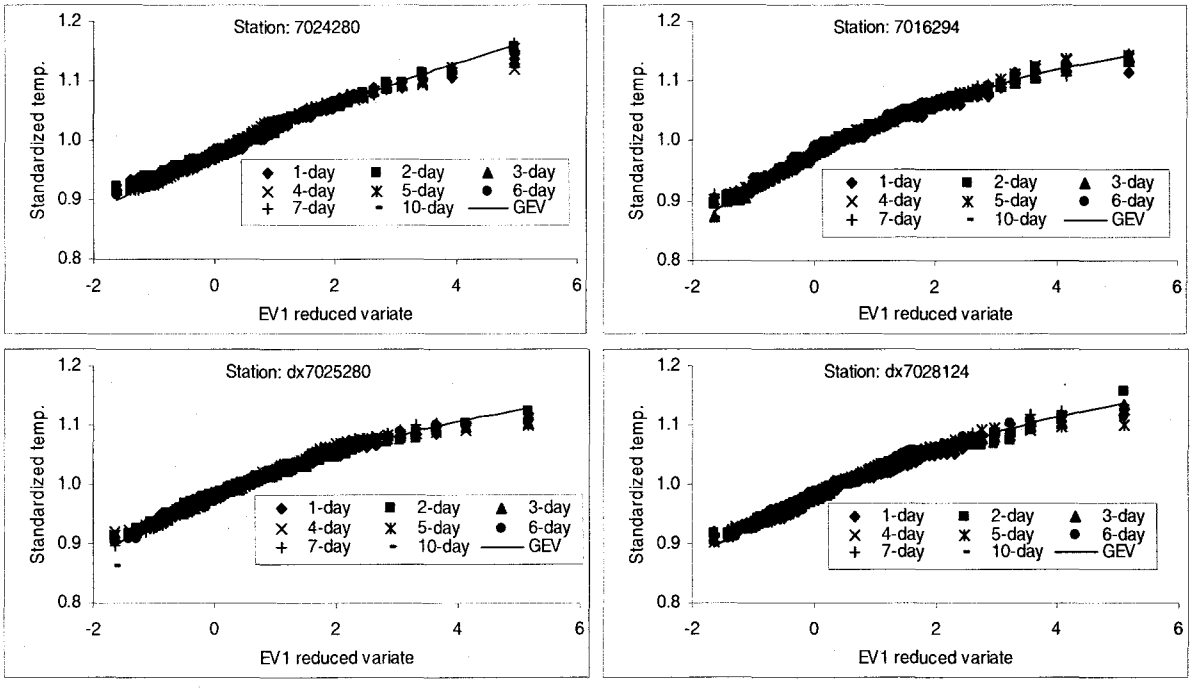


Figure B10. Plots of standardized heatwaves along with fitted GEV distribution.

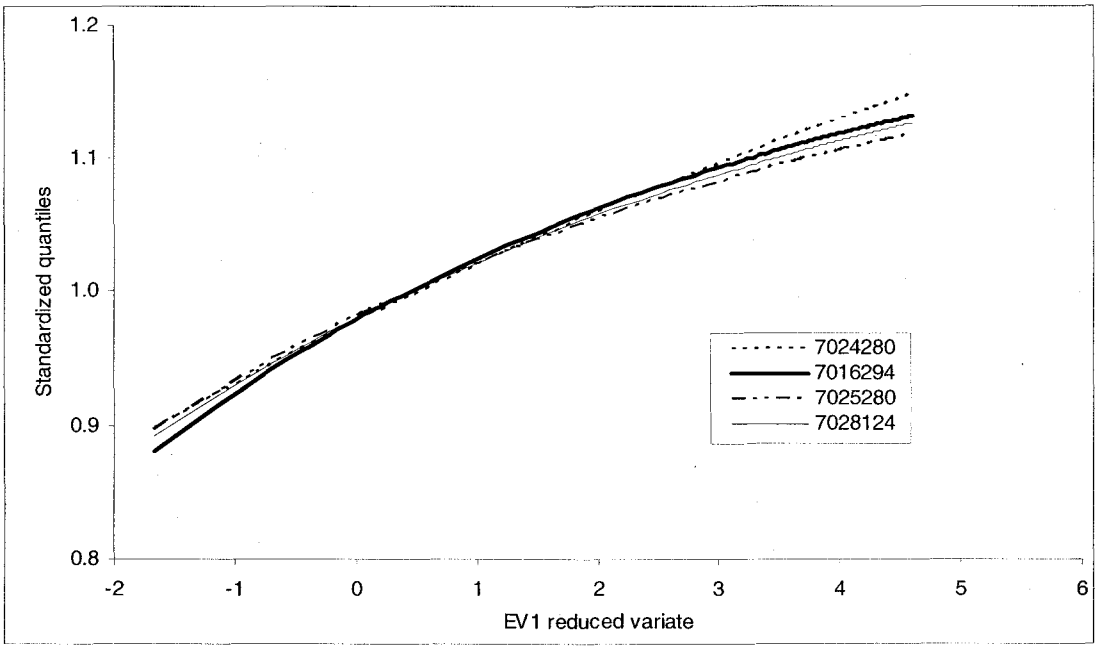


Figure B11. Comparison of at-site growth curves.

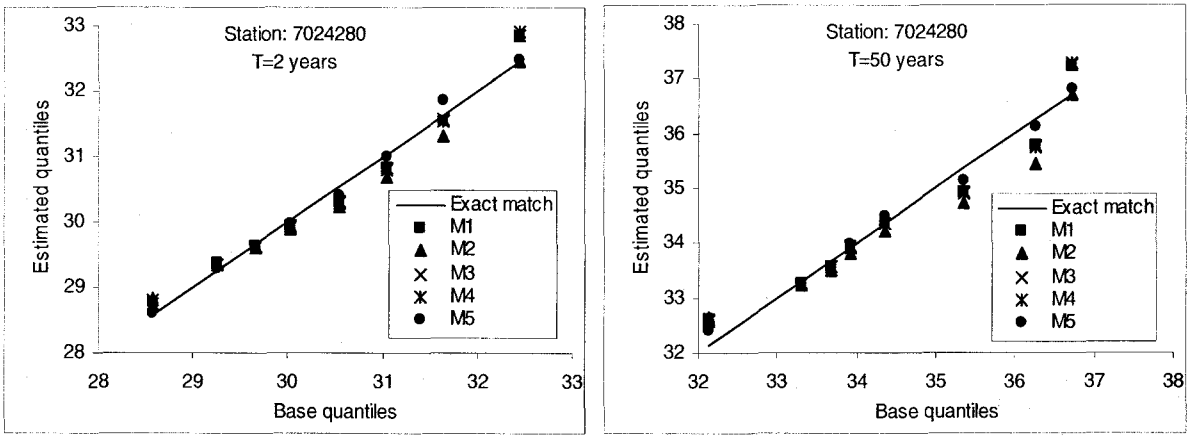
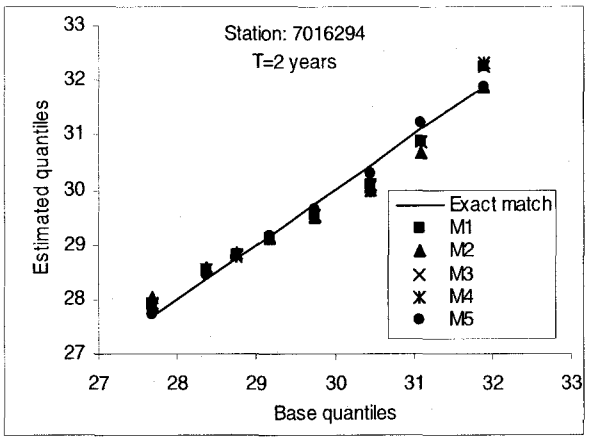


Figure B12. Comparison of estimated quantiles using models M1 to M5 with those of base model quantiles. Heatwave duration decreases along the "exact match" line. For each duration of heatwave, estimated quantiles appear as a group in the vertical direction.



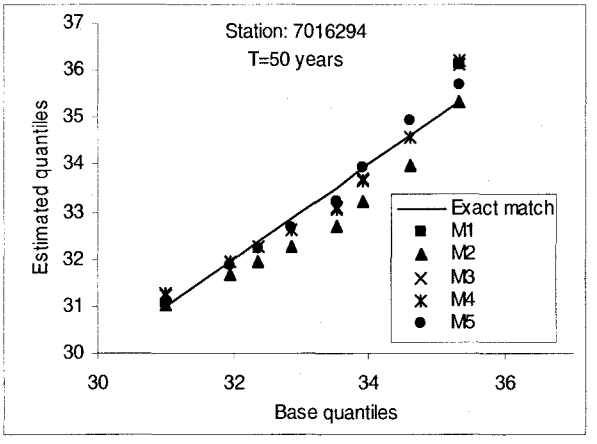
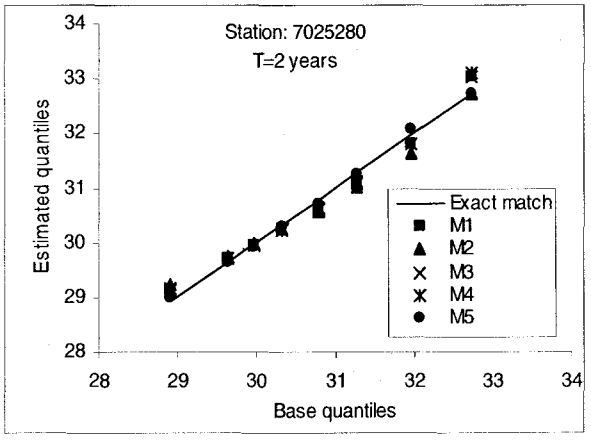


Figure B12. Contd.



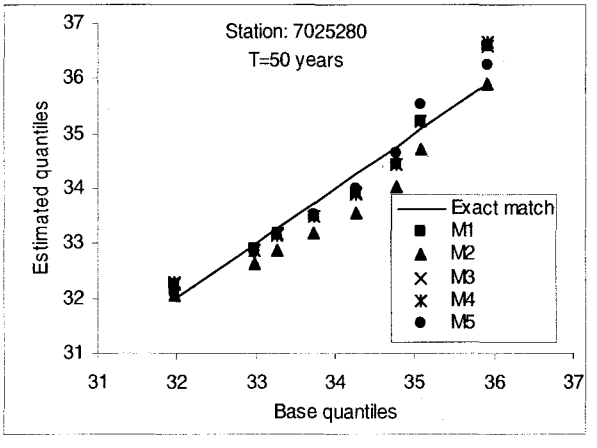
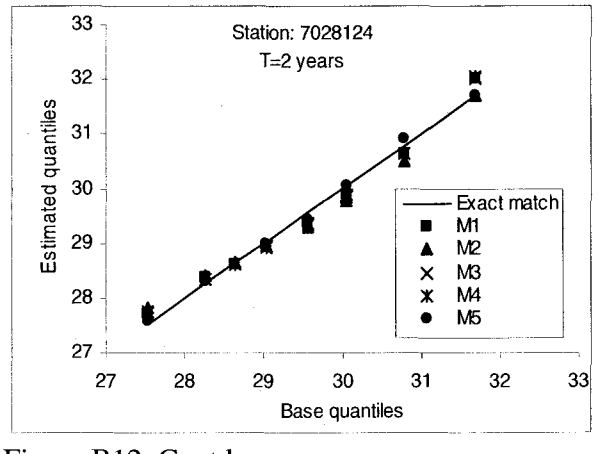


Figure B12. Contd.



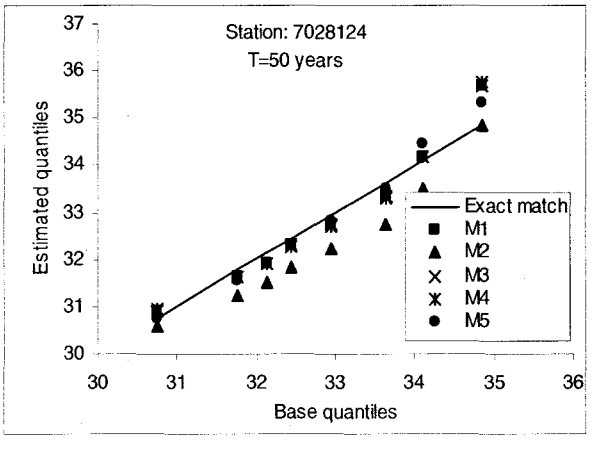


Figure B12. Contd.

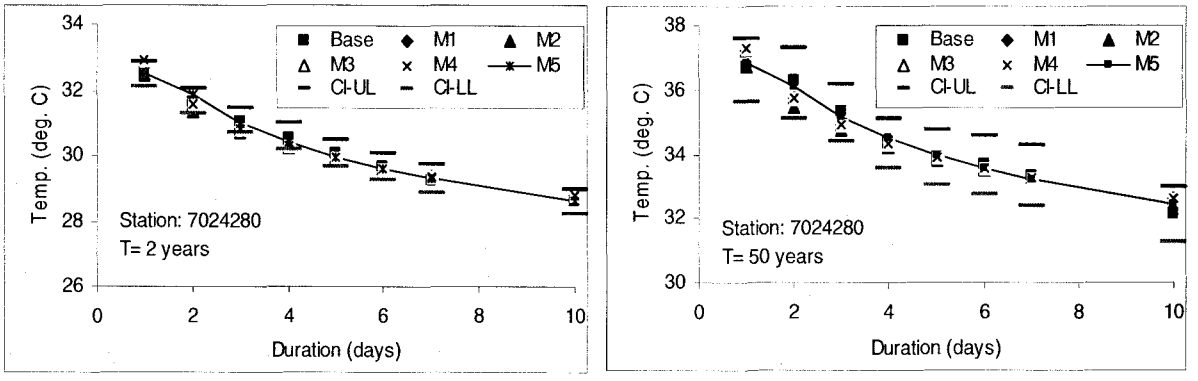


Figure B13. Plots of base and modeled quantiles along with 95% upper and lower confidence intervals. CI-LL: Confidence interval lower limit, CI-UL: Confidence interval upper limit

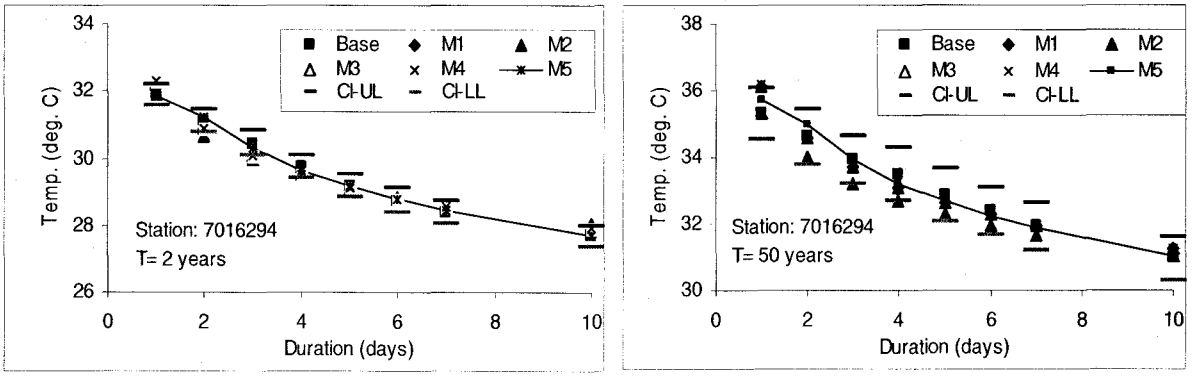


Figure B13. Contd.

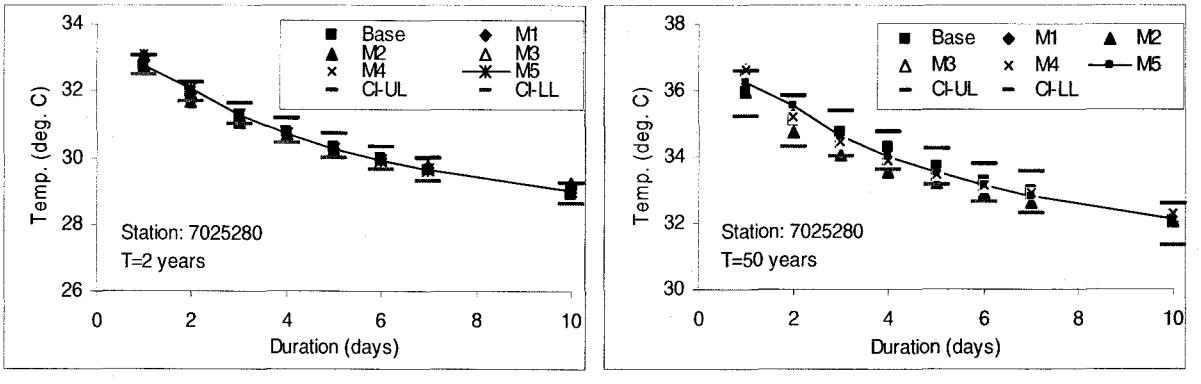


Figure B13. Contd.

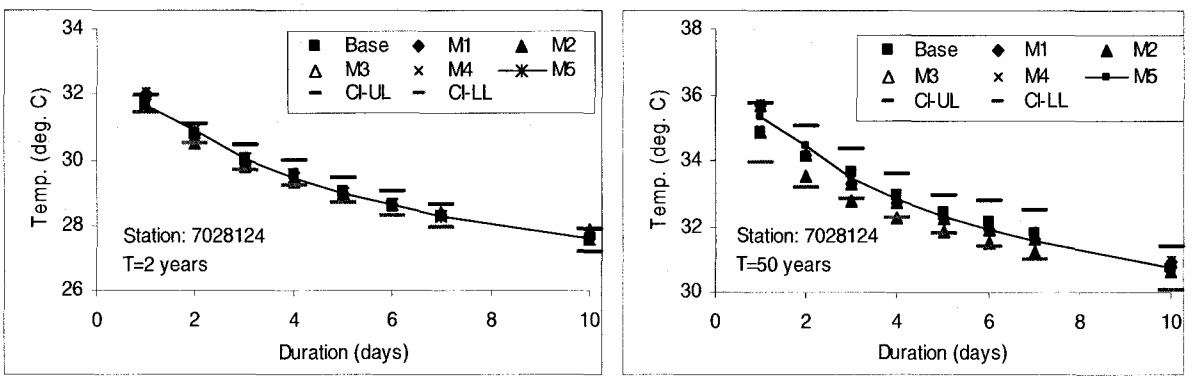


Figure B13. Contd.

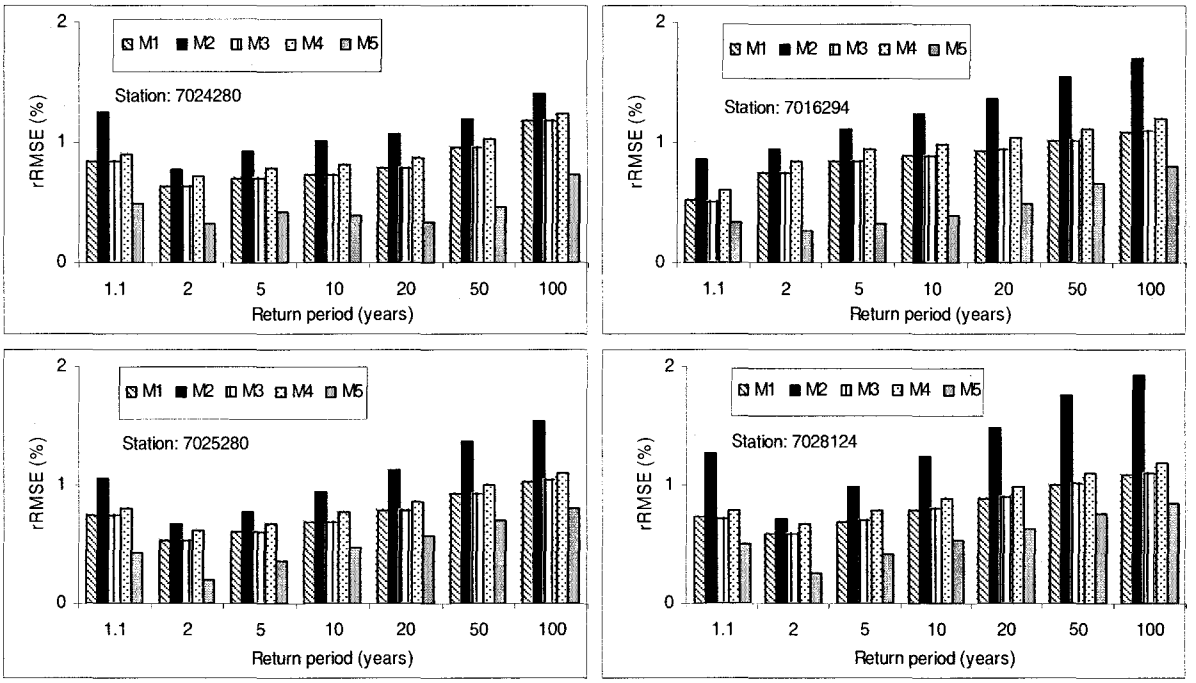


Figure B14. Comparison of adequacy of models M1 (extreme left bar) to M5 (extreme right bar) in preserving base model quantiles on the basis of relative root mean square error criterion.

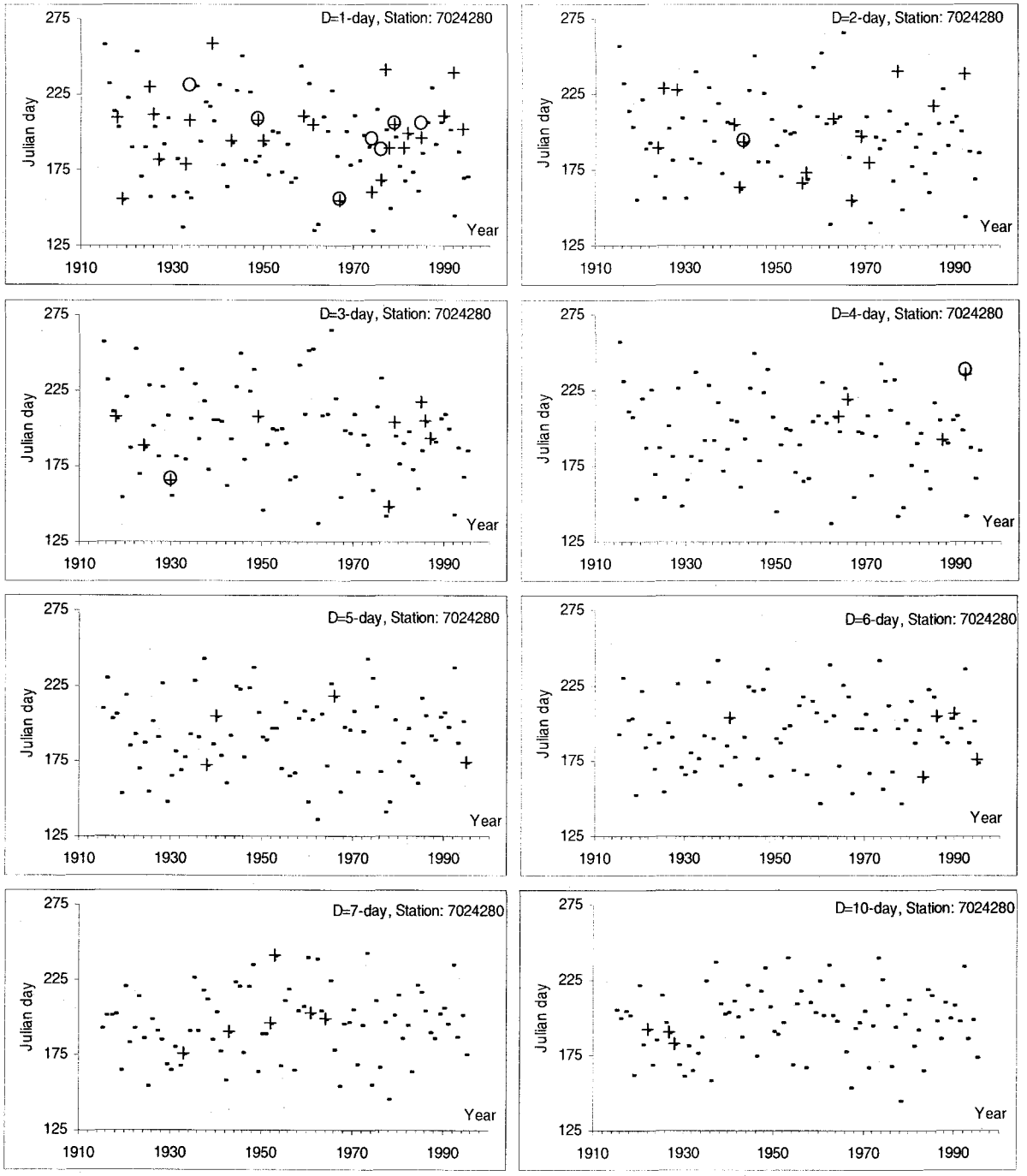


Figure B15. Dates of first occurrence and additional occurrences of heatwaves of same magnitude in a year for heatwave durations of 1 to 10 days. Dot: first occurrence; Cross: second occurrence; Circle: third occurrence.

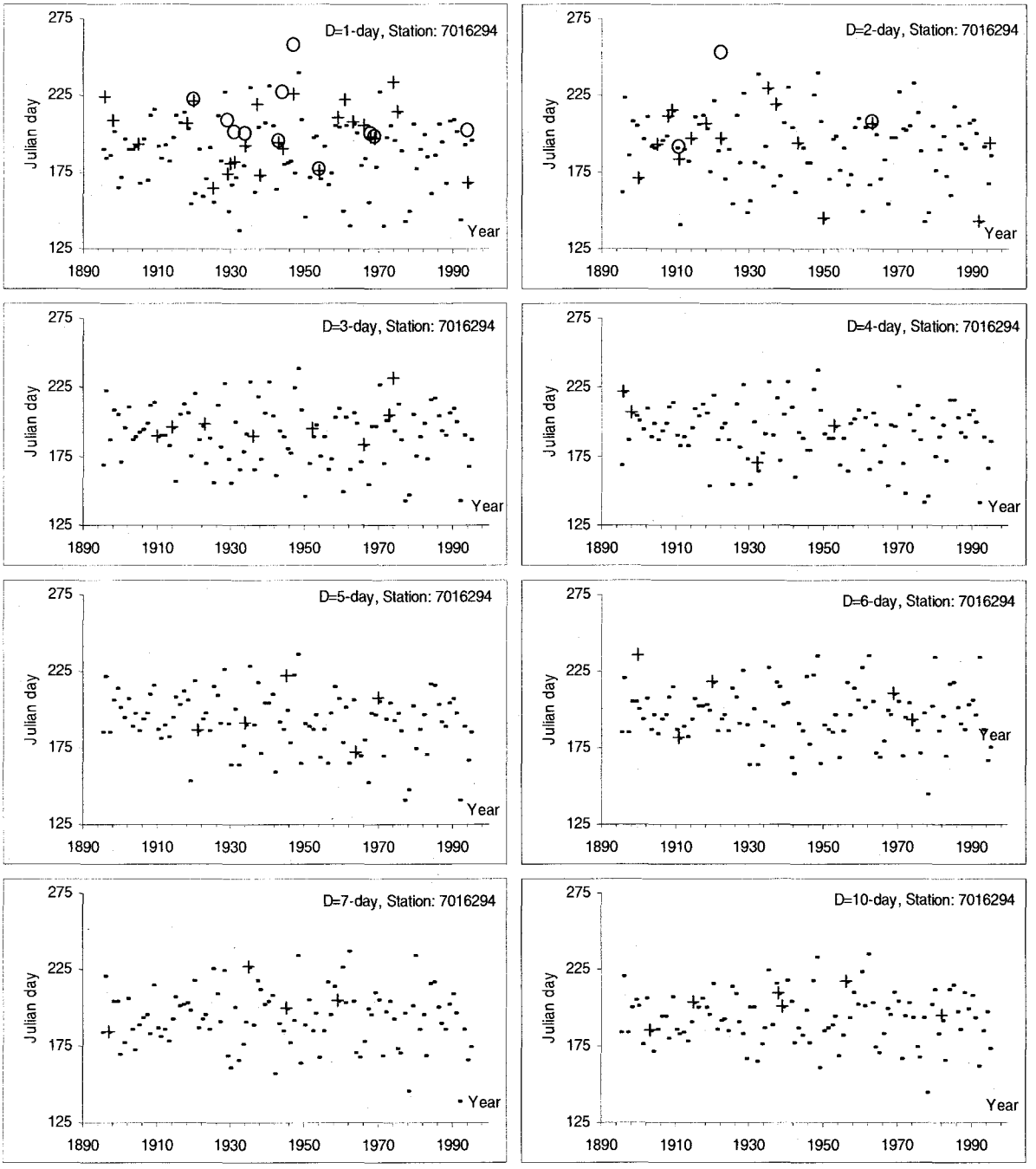


Figure B15. Contd.

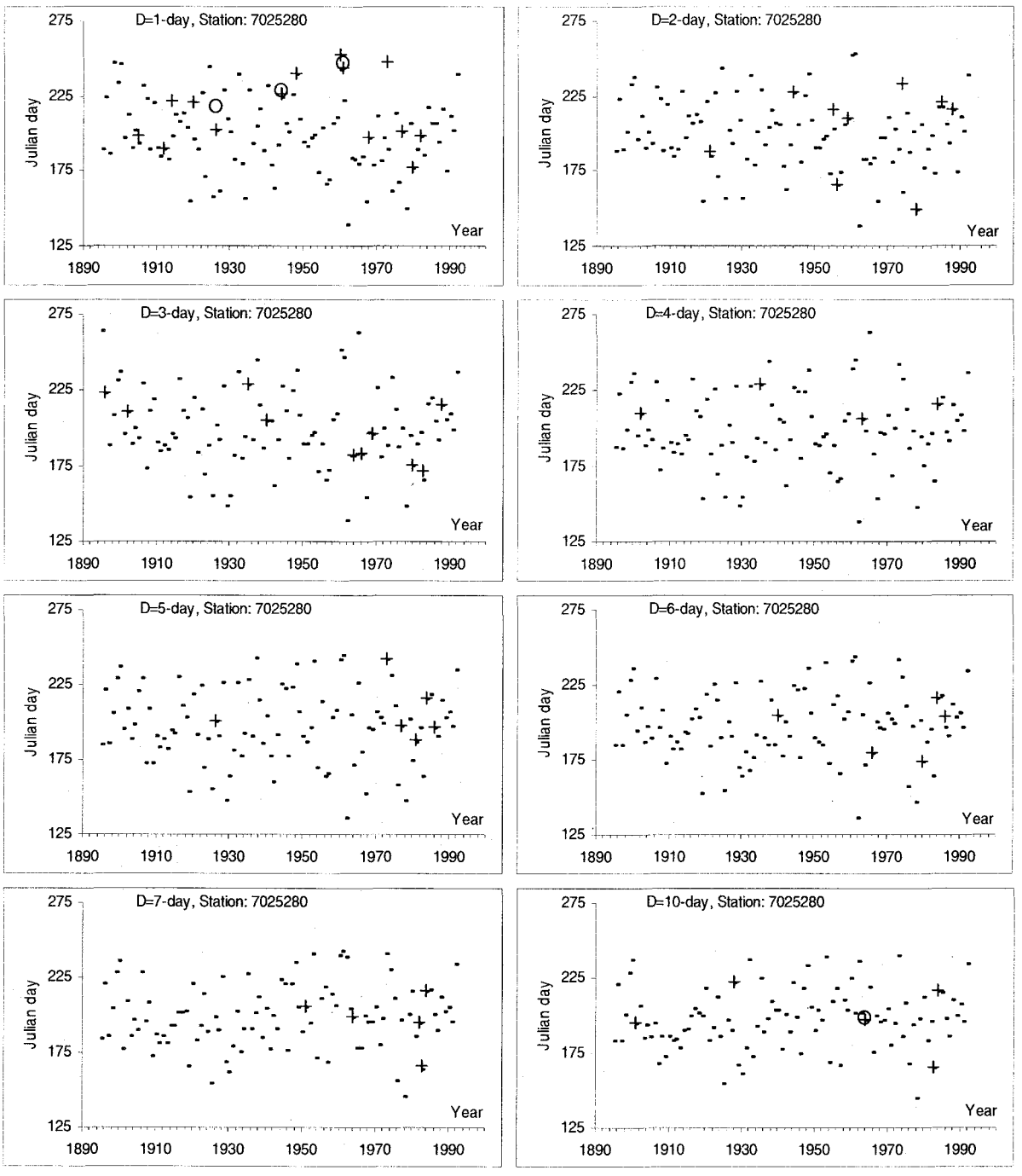


Figure B15. Contd.

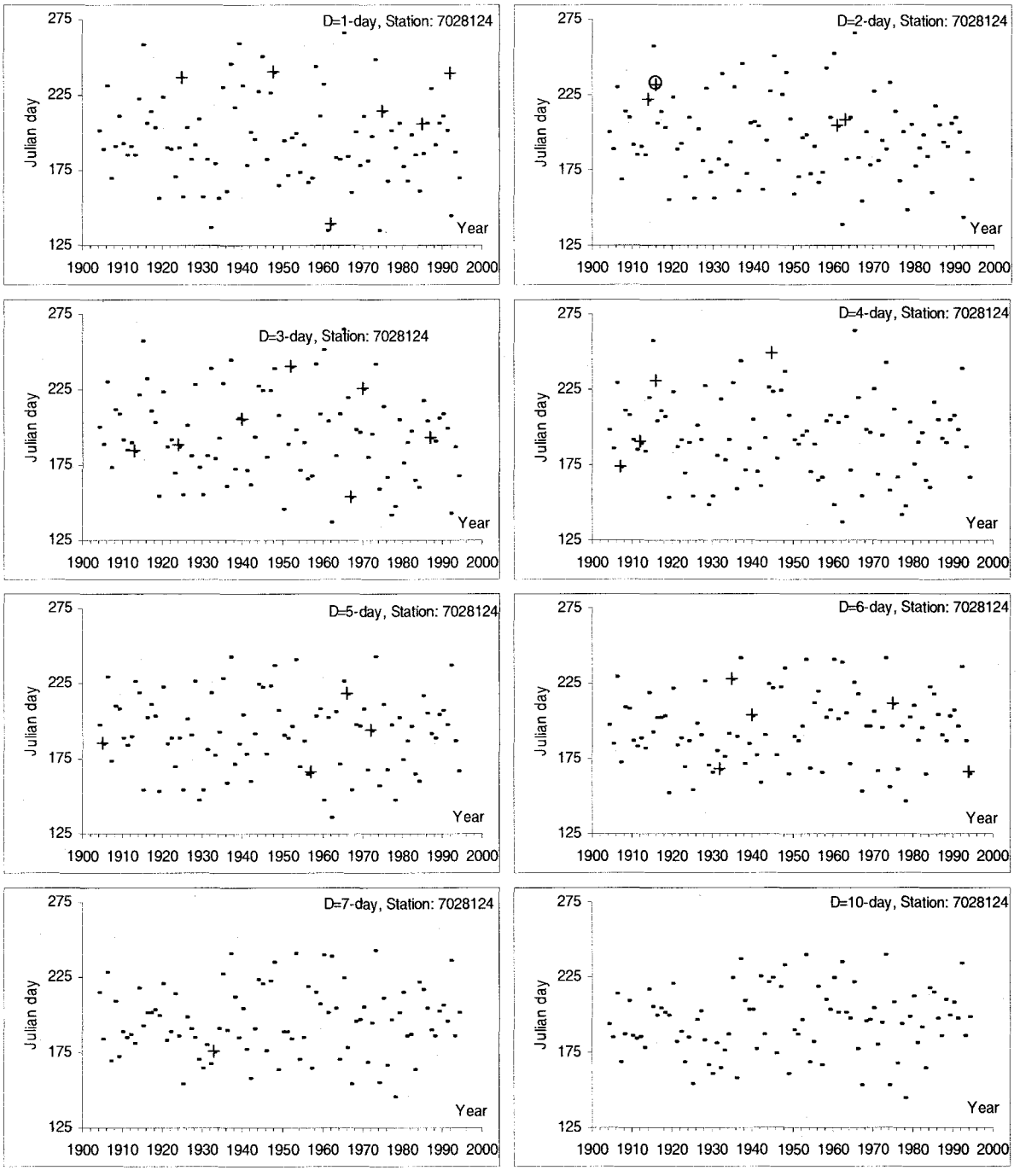


Figure B15. Contd.

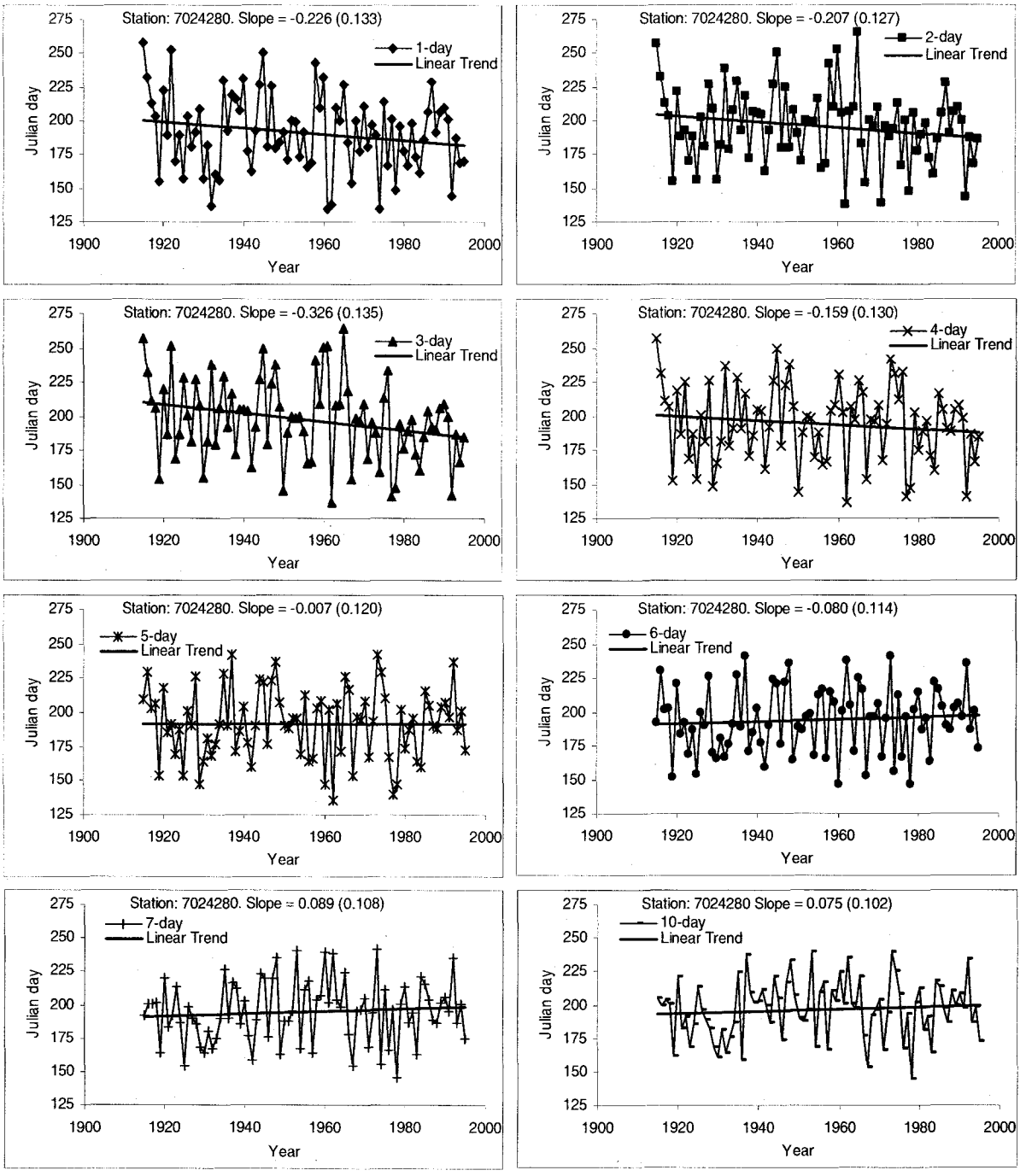


Figure B16. Time series plots of dates (converted to Julian days) of first occurrence of heatwaves of indicated durations. Bracketed value is standard error of estimated slope.

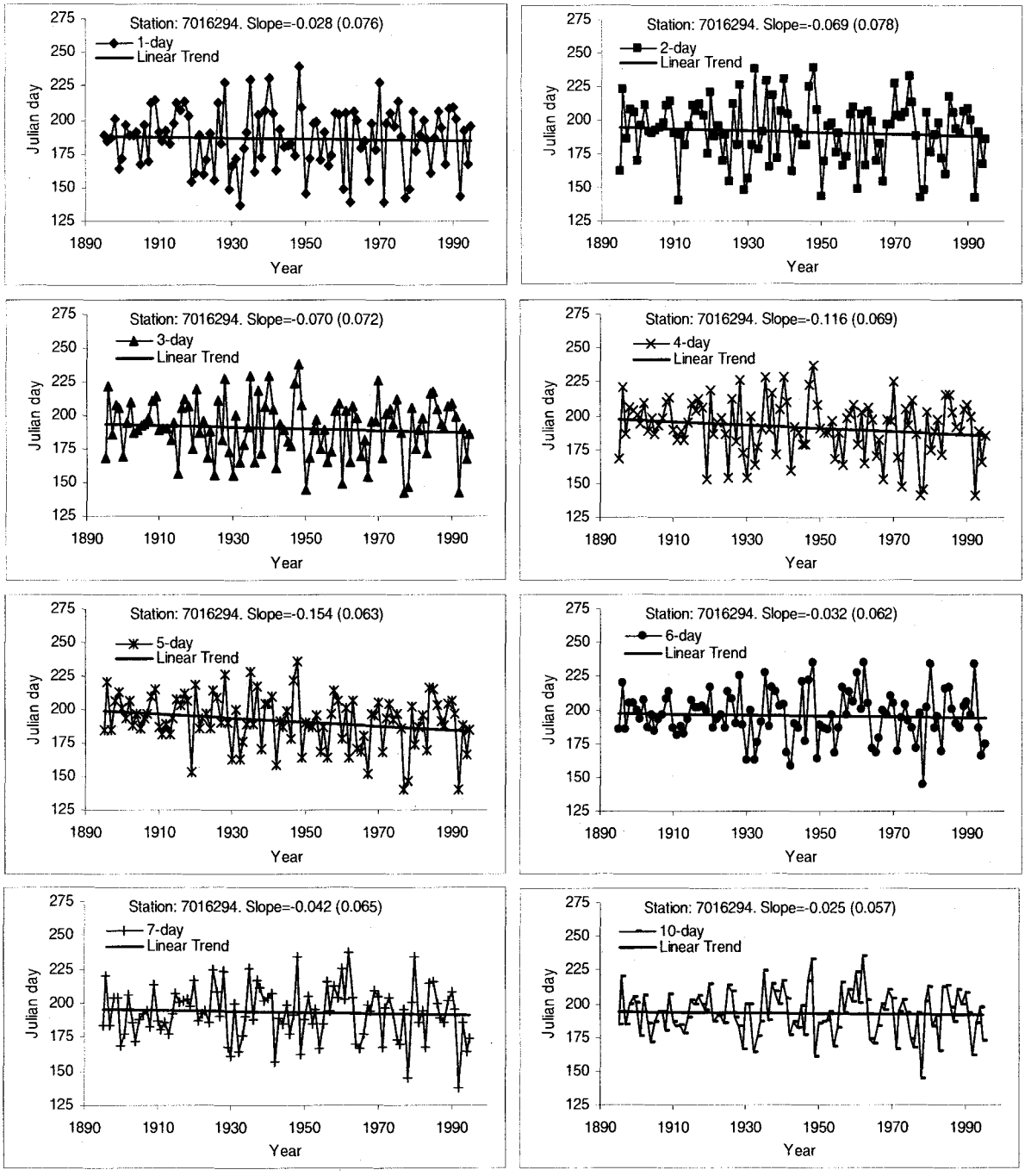


Figure B16. Contd.

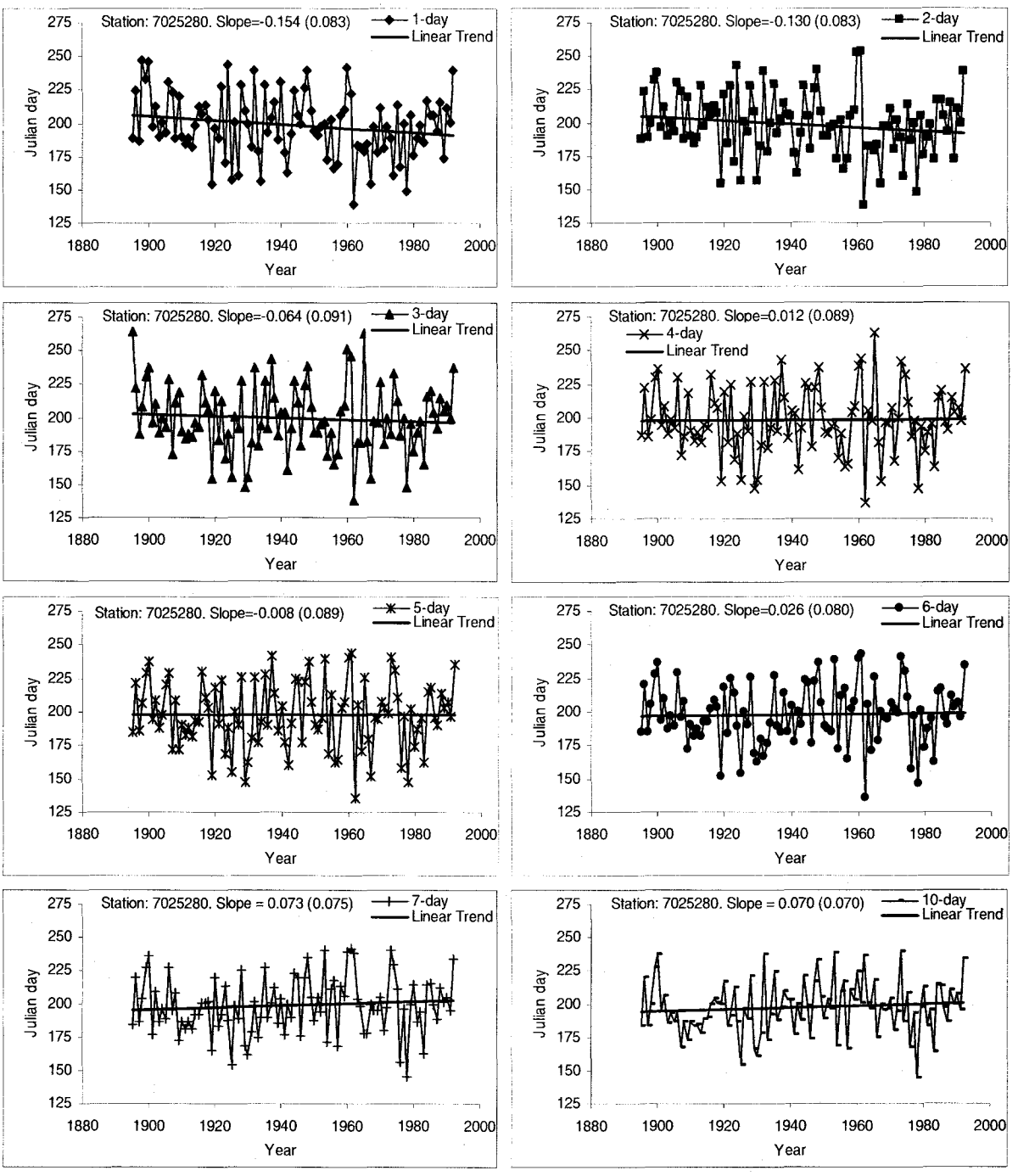


Figure B16. Contd.

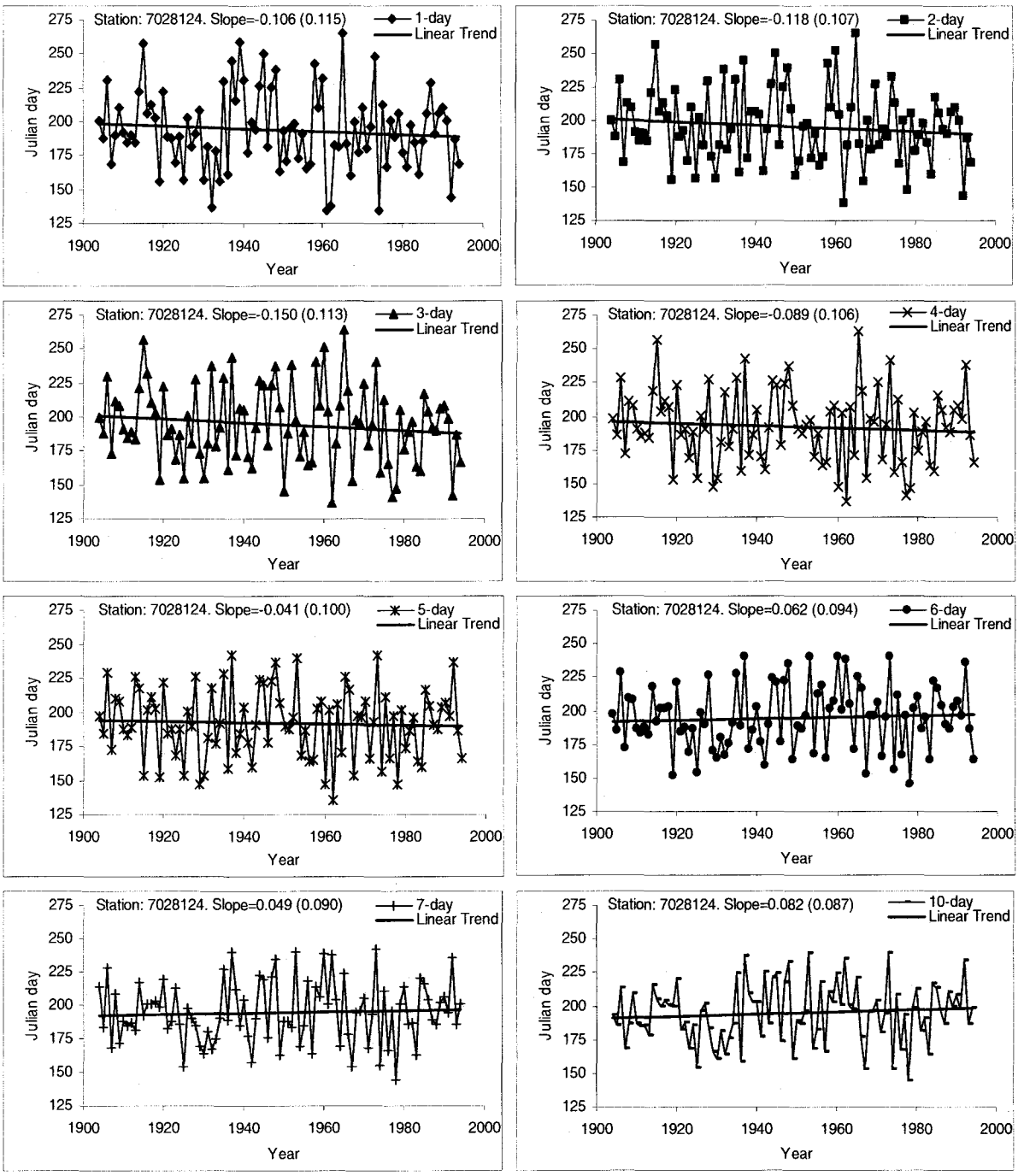


Figure B16. Contd.

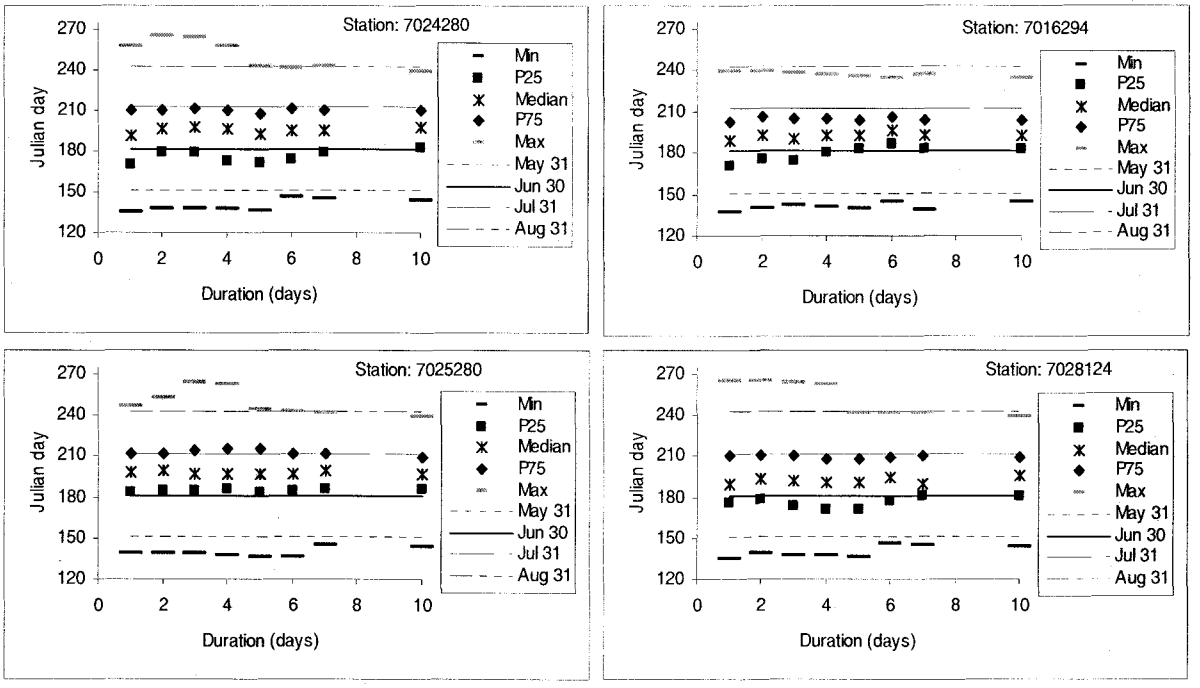


Figure B17. Quartile plots of dates of first occurrence of heatwaves of 1-10 days durations.

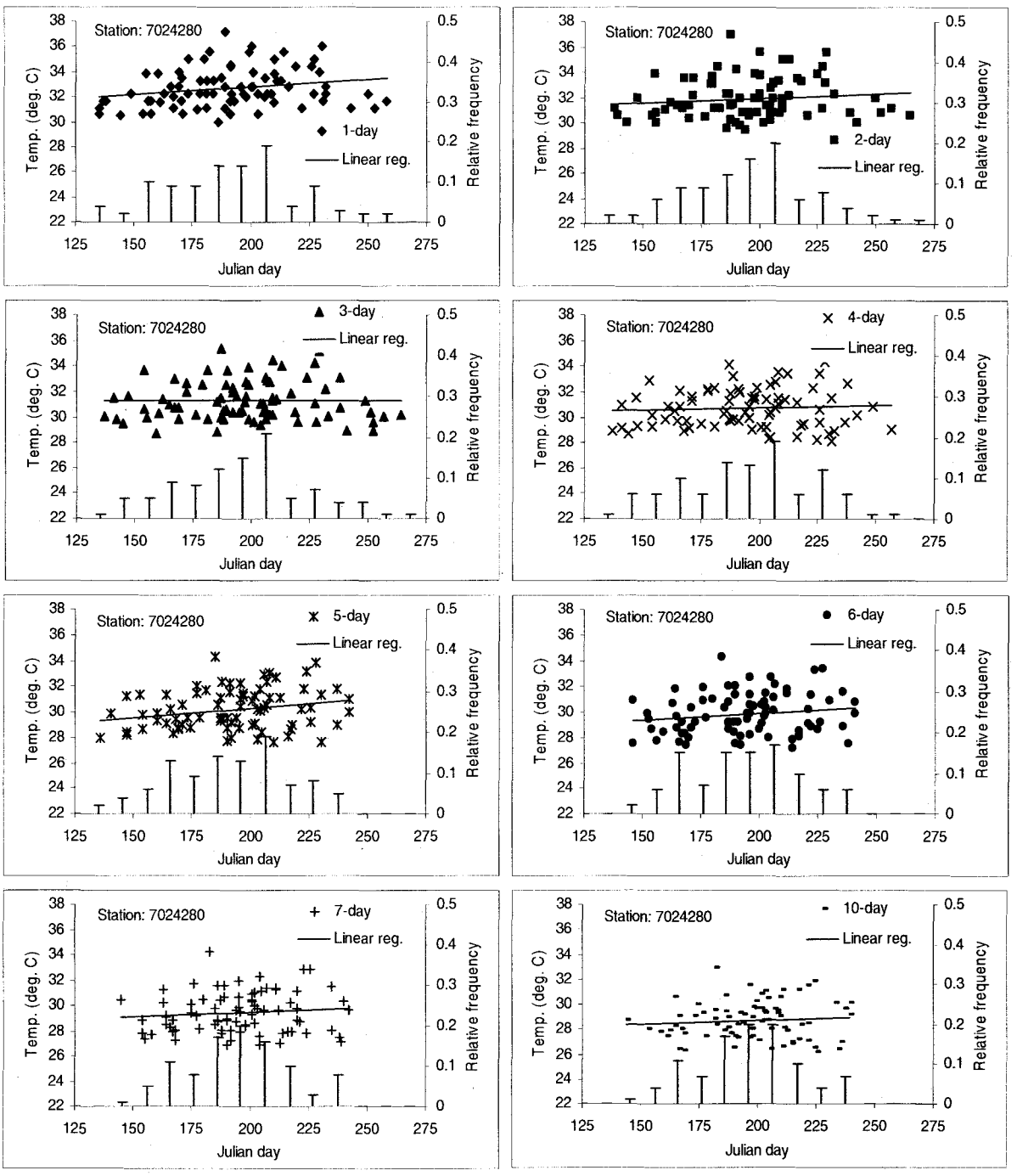


Figure B18. Relationship between heatwave magnitude and date of first occurrence for heatwave durations of 1-10 days. Relative frequencies are shown as stem plots (vertical lines).

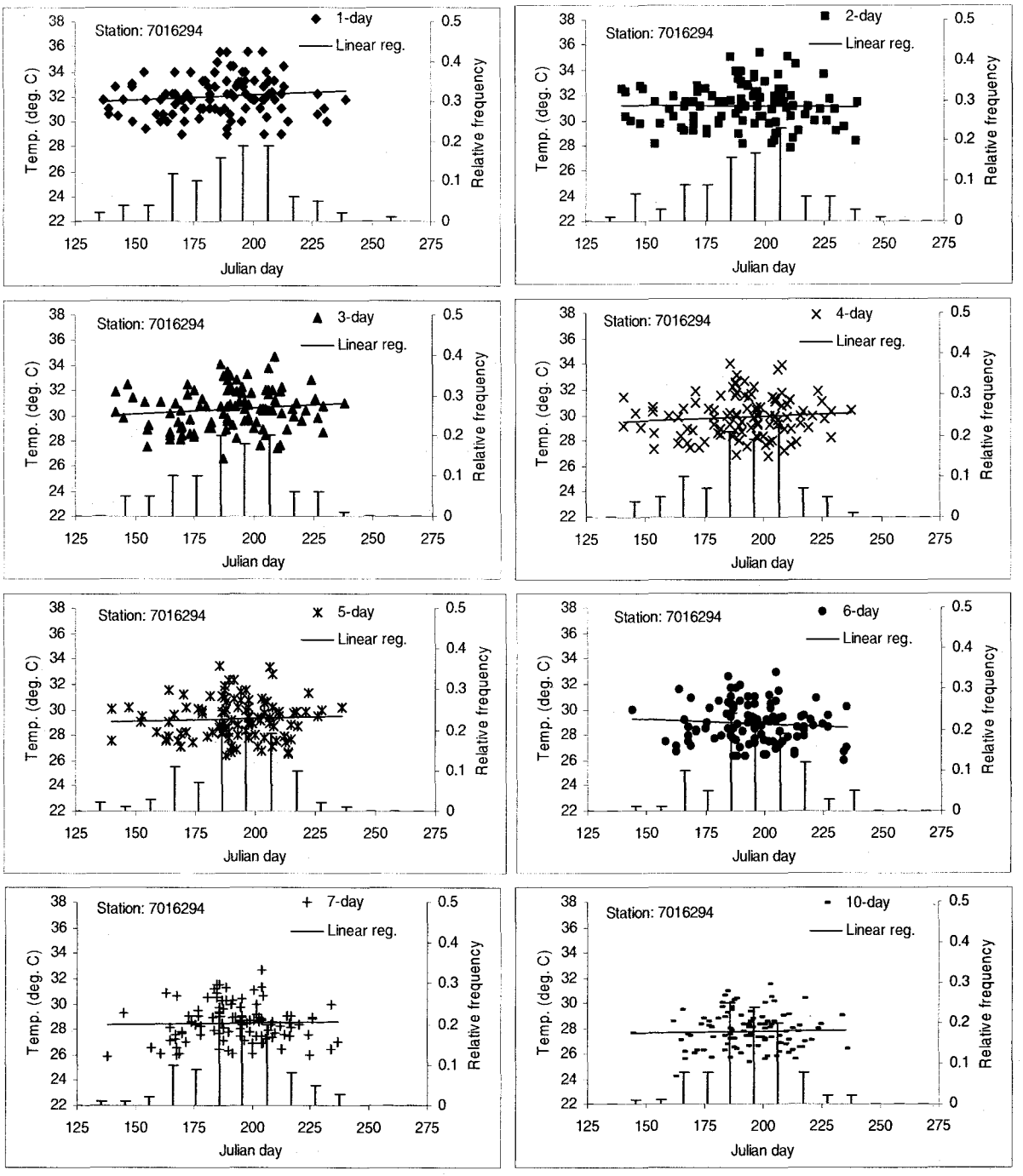


Figure B18. Contd.

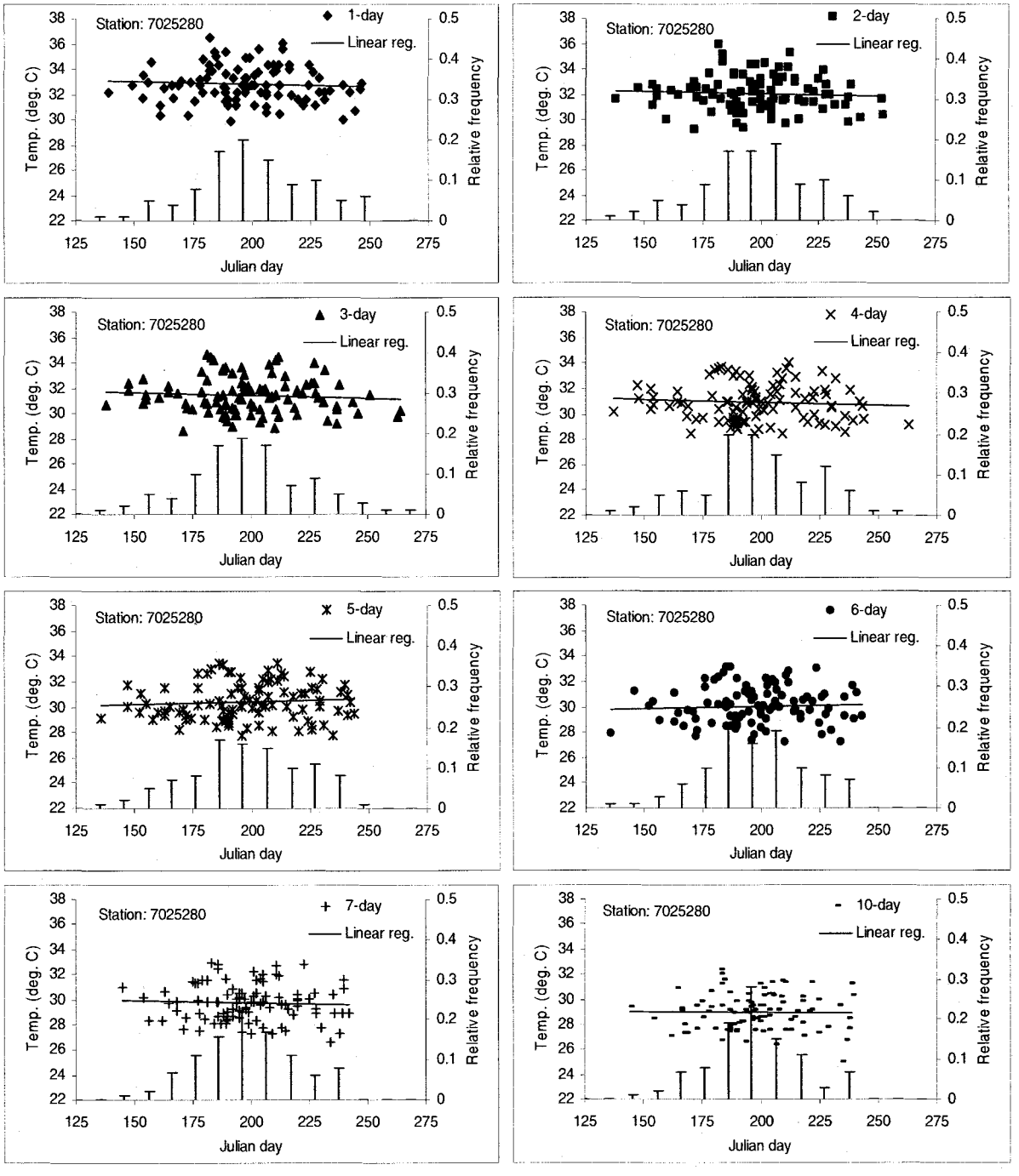


Figure B18. Contd.

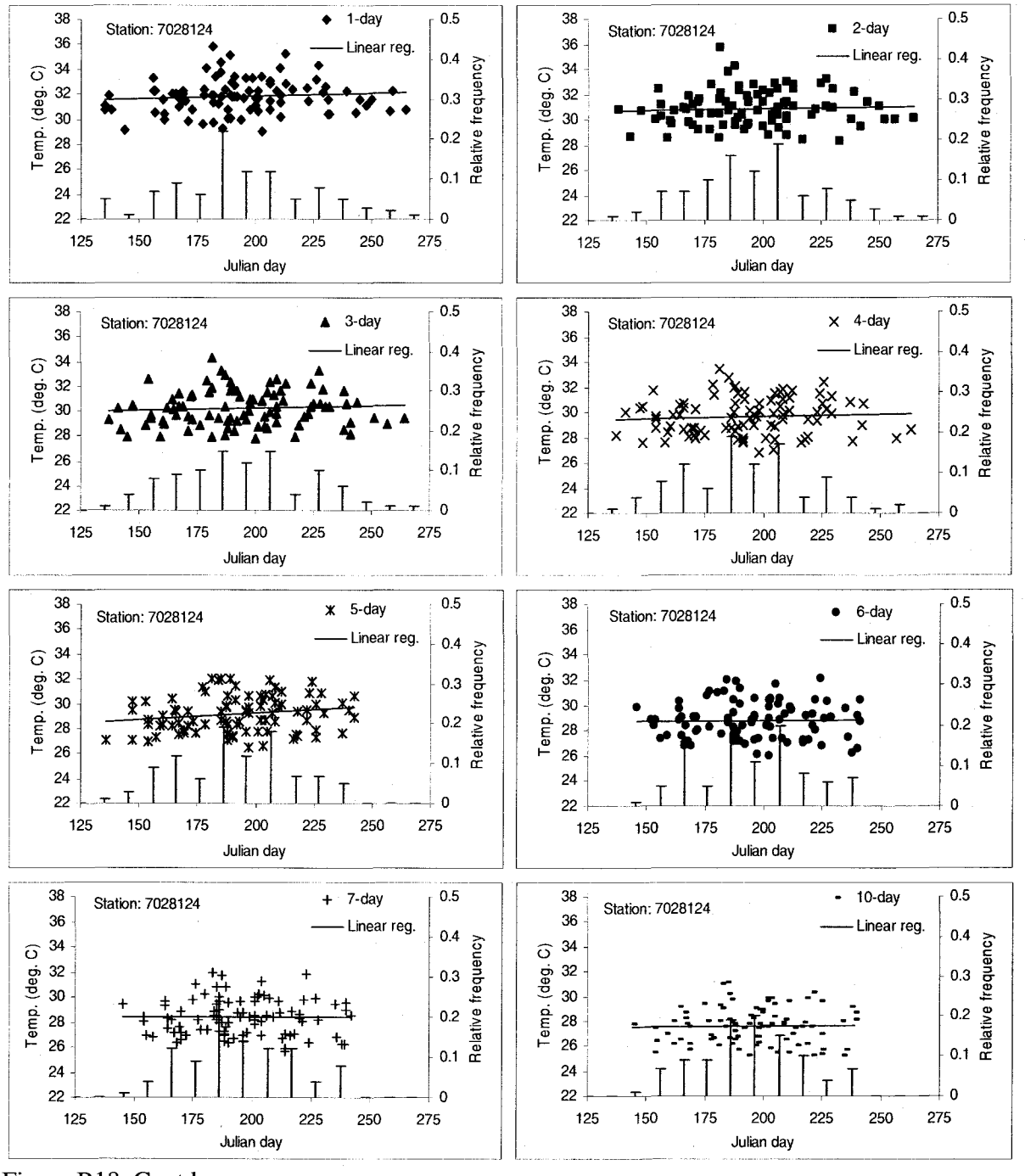


Figure B18. Contd.

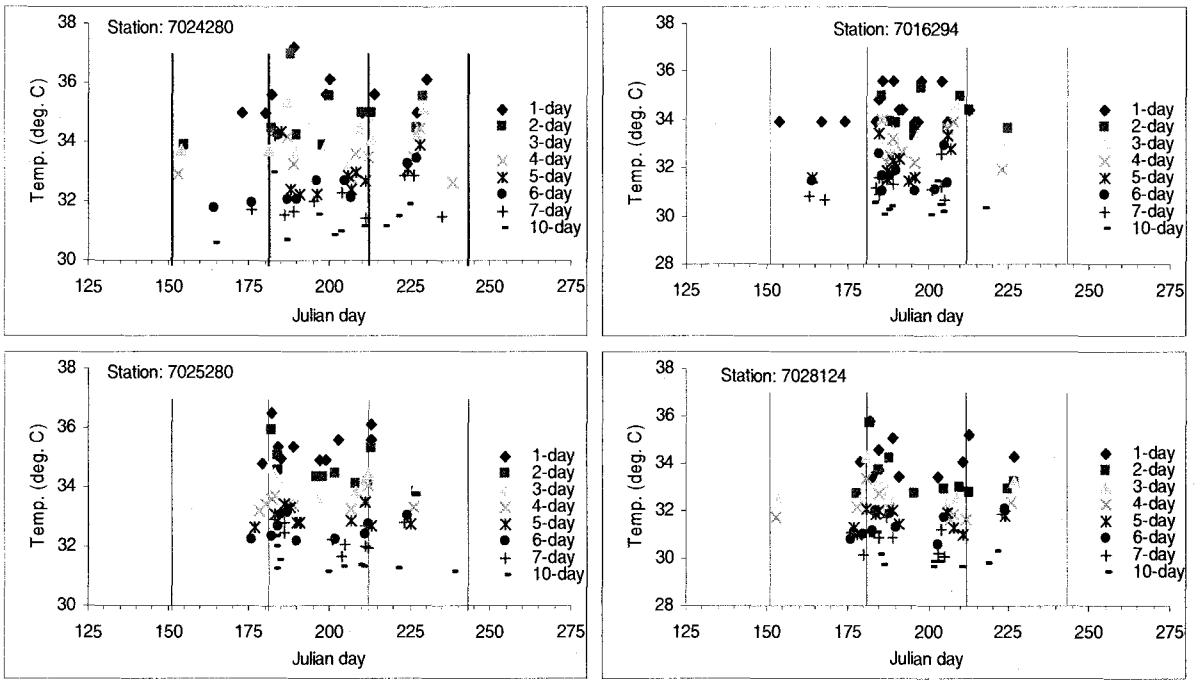


Figure B19. Ten largest heatwaves and their corresponding dates of occurrence for heatwave durations of 1-10 days. Vertical lines (from left to right) represent end of May, June, July and August, respectively.

9.0 APPENDIX C

This appendix contains a summary of non-parametric tests used in the study for trend analysis. The regional homogeneity tests proposed by Hosking and Wallis (1993) are also included.

Spearman's Rank-Correlation Test:
The test is based on the Spearman rank-correlation coefficient, R_{sp}, which is defined as (Dahmen and Hall, 1990):

$$R_{sp} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{C.1}$$

where n is the total number of data, d is difference, and i is the chronological order number. The difference between rankings is computed with $d_i = RX_i - RY_i$, RX_i is the rank of the variable, X, which is the chronological order number of the observations. The series of observations, Y_i, is transformed to its rank equivalent, RY_i, by assigning the chronological order number in the ranked series, Y. If there are ties, i.e. two or more ranked observations, Y, with the same value, the convention is to take RX as the average rank. One can test the null hypothesis, $H_0 : R_{sp} = 0$ (there is no trend), against the alternate hypothesis, $H_0 : R_{sp} <> 0$ (there is a trend), with the test statistic:

$$t_t = R_{sp}\left[\frac{n-2}{1-R_{sp}^2}\right]^{0.5} \tag{C.2}$$

where t_t has Student's t-distribution with $v = n-2$ degrees of freedom. At a significance level of 5% (two-tailed), the time series has no trend if: $t_{v,2.5\%} < t_t < t_{v,97.5\%}$ or the p-value of the R_{sp} statistic of the observed sample is > *0.05*. If the time series does have a trend, the data cannot be used for frequency analyses and recourse has to be made to other approaches to model such behavior.

Mann-Kendall Test
The Mann-Kendall test is based on the test statistic S defined as:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sgn}(X_j - X_i) \tag{C.3}$$

where X_i and X_j are the sequential data values, n is the length of the data set, and

$$\mathrm{sgn}(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases} \tag{C.4}$$

At certain probability level (i.e. α level of significance), H_0 is rejected in favor of H_1 if the absolute value of S equals or exceeds a specified value $S_{\alpha/2}$, where $S_{\alpha/2}$ is the smallest S which has the probability less than $\alpha/2$ to appear in case of no trend. A positive (negative) value of S

indicates an upward (downward) trend. For $n \geq 8$, Mann (1945) and Kendall (1975) have documented that the statistic S is approximately normally distributed with the mean and variance as follows:

$$E[S] = 0, \text{ and} \tag{C.5}$$

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{n} t_i i(i-1)(2i+5)}{18} \tag{C.6}$$

where t_i is the number of ties of extent i (i.e. how many X's are involved in a tie). The standardized test statistic Z is computed as:

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{\text{Var}(S)}} & S > 0 \\ 0 & S = 0 \\ \dfrac{S+1}{\sqrt{\text{Var}(S)}} & S < 0 \end{cases} \tag{C.7}$$

The standardized test statistic Z follows the standard normal distribution with mean of zero and variance of 1. The presence of a statistically significant trend is evaluated using the Z value. A positive (negative) value of Z indicates an upward (downward) trend. At α level of significance, H_0 is rejected in favor of H_1 if the absolute value of Z is greater than $Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is obtained from the standard normal distribution.

Sen Slope

The magnitude of the slope of trend is estimated non-parametrically using the approach of Sen (1968),

$$\text{slope} = \text{Median}\left(\frac{X_i - X_j}{i-j}\right) \quad \forall j < i \tag{C.8}$$

where X_i and X_j are data values at time points i and j, respectively. The slope determined this way is a robust estimate of the magnitude of monotonic trend (i.e. linear trend). If there are n values in the time series then one can get as many as $N = n(n-1)/2$ slope estimates. If these N estimates are ranked from the smallest to the largest then the Sen's estimator is taken as the central value if N is odd and the average of two central values if N is even. A 100(1-α)% , two sided confidence interval about the slope estimate is obtained in a non-parametric manner using the variance of Mann-Kendall test statistic (i.e. $\text{Var}(S)$) and the normal distribution i.e. by computing $B_\alpha = Z_{1-\alpha/2}\sqrt{\text{Var}(S)}$, $I_L = (N - B_\alpha)/2$, and $I_U = (N + B_\alpha)/2$. The lower and upper limits of the confidence interval are the values located at I_L th and I_U th indices in N ordered

slope estimates. If the indices are not whole numbers then the limits are obtained by interpolating the slope estimates at two nearest whole number indices.

Test of Serial Correlation Coefficients

The serial correlation coefficients can help to verify the independence of a time series or the absence of persistence in a time series. If a time series is completely random, the population auto-correlation function will be zero for all lags other than zero where its value is unity because all data sets are perfectly correlated with themselves. The sample serial correlation coefficients will deviate from zero only because of sampling effects. According to Dahmen and Hall (1990), for checking the independence of a time series, it is usually sufficient to compute the lag-1 serial correlation coefficient, i.e., the correlation between adjacent observations in a time series. The lag-*k* serial correlation coefficient, r_k, is given by (Box and Jenkins, 1970)

$$r_k = \{\sum_{i=1}^{n-k}(x_i - \bar{x}) \times (x_{i+k} - \bar{x})\} / \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{(C.9)}$$

where x_i is an observation, x_{i+k} is the kth following observation, $\bar{x}$ is the mean of the time series, and n is the number of data values. After computing r_k, one can test the null hypothesis $H_0 : r_k = 0$ (that there is no correlation between two consecutive observations) against the alternate hypothesis $H_1 : r_k <> 0$. According to Salas et al. (1980), who referred to Anderson (1942), the confidence limits on the correlogram of an independent series are given by

$$r_k(\alpha\%) = \{-1 \pm z\sqrt{n-k-1}\}/(n-k), \qquad \text{(C.10)}$$

where α is the level of significance. The values of z vary according to the chosen level of significance, e.g., $z = 1.645$ at $\alpha = 10\%$, 1.96 at $\alpha = 5\%$ and 2.575 at $\alpha = 1\%$. A computed value of r_k becomes significant if it lies outside the confidence interval. For a random series, lagged values of the series are uncorrelated and we expect $r_k \cong 0$. However, any given r_k has $\alpha\%$ chance of being outside the confidence limits. For example, if $\alpha = 5\%$ then one value outside the limits might be expected in a correlogram plotted to lag 20 even for a purely random series. Whether a given r_k outside the confidence band really indicates "significant autocorrelation" or not is open to question. However, a large r_k is less likely to occur by chance than an r_k barely outside the confidence bands. Two additional factors that can be considered are: (1) the magnitude and (2) the position (lag k) of the significant r_k to decide about the presence of persistence and/or periodic signals.

The F-test for stability of variance

The statistics for testing the stability of variance of a time series is the ratio of the variances of two sub-samples of the time series. The distribution of the variance ratio of the sub-samples from a normal distribution is known as the F, or Fisher, distribution. The test statistic is

$$F_t = \sigma_1^2 / \sigma_2^2 = S_1^2 / S_2^2 \qquad \text{(C.11)}$$

where S^2 is the variance. Thus it is irrelevant whether one uses the sample variance or population variance. The variance of the time series is stable if $F(\nu_1,\nu_2,2.5\%) < F_t < F(\nu_1,\nu_2,97.5\%)$. Alternatively, the p-value of the statistic F_t can be obtained from the F-distribution. A small p-value (e.g., ≤ 0.05) indicates that the variance is probably instable (at 5% level of significance). Detailed description of this test can be found in most of the graduate level statistical books, e.g., Walpole and Myers (1989).

The t-test for stability of mean

The t-test for stability of mean involves computing and then comparing the means of two (or three) non-overlapping sub-samples of the time series. The null hypothesis, $H_0 : \overline{X}_1 = \overline{X}_2$, can be tested against the alternate hypothesis, $H_1 : \overline{X}_1 <> \overline{X}_2$, by calculating test statistic

$$t_t = \frac{\overline{X}_1 - \overline{X}_2}{[\{(n_1-1)S_1^2 + (n_2-1)S_2^2\} \times (n_1^{-1} + n_2^{-1})/(n_1+n_2-1)]^{0.5}} \qquad \text{(C.12)}$$

where n is the number of data in the sub-sample, $\overline{X}$ the mean of the sub-sample and S^2 its variance. In samples from a normal distribution, t_t has a Student t-distribution. The mean of the time series is considered stable if $t(\nu,2.5\%) < t_t < t(\nu,97.5\%)$, where $\nu = n_1 - 1 + n_2 - 1$ degrees of freedom. Alternatively, the p-value of the statistic t_t can be obtained from the Student t-distribution. A small p-value indicates that the mean is probably instable. Detailed description of this test can be found in most of the graduate level statistical books, e.g., Walpole and Myers (1989).

Hosking and Wallis (1993)'s 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

$$V_1 = \sum_{i=1}^{N} n_i (t_{2(i)} - \bar{t}_2)^2 / \sum_{i=1}^{N} n_i \qquad \text{(C.13)}$$

2. A measure of dispersion for both the L-CV and the L-skewness coefficients in the L-CV-L-skewness space

$$V_2 = \sum_{i=1}^{N} n_i [(t_{2(i)} - \bar{t}_2)^2 + (t_{3(i)} - \bar{t}_3)^2]^{1/2} / \sum_{i=1}^{N} n_i \qquad \text{(C.14)}$$

3. A measure of dispersion for both the L-skewness and the L-kurtosis coefficients in the L-skewness-L-kurtosis space

$$V_3 = \sum_{i=1}^{N} n_i [(t_{3(i)} - \bar{t}_3)^2 + (t_{4(i)} - \bar{t}_4)^2]^{1/2} / \sum_{i=1}^{N} n_i \qquad \text{(C.15)}$$

where $\bar{t}_2$, $\bar{t}_3$, and $\bar{t}_4$ are the group mean 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*; 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 to the variation that would be expected in a homogeneous group. The heterogeneity measures are evaluated using the following expression:

$$W_k = (V_k - \mu_{V_k}) / \sigma_{V_k}, \qquad \text{for } k = 1,2,3 \qquad \text{(C.16)}$$

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 $1 \le W_k < 2$, and "definitely heterogeneous" if $W_k \ge 2$. Note some of the original notation has been changed to suit present study.