An Analysis of Peaks Over Threshold Flood Data

BY

Debra E. L. Boulet

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of

Master Of Science

Department of Civil and Geological Engineering
University of Manitoba
Winnipeg, Manitoba

0-612-32061-8

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

AN ANALYSIS OF PEAKS OVER THRESHOLD FLOOD DATA

BY

DEBRA E.L. BOULET

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University

of Manitoba in partial fulfillment of the requirements of the degree

of

MASTER OF SCIENCE

ABSTRACT

The purpose of this thesis is to analyze flood data for seasonality measures and trends in the flood peaks. The flood peak data are assembled using the Peaks Over Threshold approach to extract flood peak data from daily flow records. In order to develop the Peaks Over Threshold technique, criteria to set the threshold and identify the peaks must be determined. As there are several different sets of criteria, they must be carefully analyzed to determine the most appropriate criteria of the possible choices of the sets.

After the peaks have been extracted from the daily flow records, tests are run in order to determine the most appropriate average number of peaks per year. The results are classified by region, with regional average peaks per year being calculated. Using this data, a final extraction is run, and summarized.

The final data sets are examined for trends in flood peak magnitudes, the standard deviation of each year of the record, and the number of peaks per year. Any stations that exhibit trends are analyzed for commonalties. The final data sets are also analyzed for seasonality in the mean day of flood. These data are analyzed on a regional basis to determine if patterns exist.

The results indicate that sites in Newfoundland are more likely to experience trends. These trends are generally increasing trends in the daily flow values (standard deviation), peak flow values and the number of peak flow occurrences per year. The sites in the Prairies are less likely to experience trends, and those

that do occur are decreasing trends. The seasonality results indicate that flood events in the Prairie sites are strongly seasonal and occur during the spring. The results for the Newfoundland sites indicate that flood events occur throughout the year, although there is a higher tendency towards flooding during the winter months of January, February and March. This indicates that there are seasonality patterns associated with both geographical areas.

Acknowledgements

I would like to express my sincere thanks to the following people:

- My advisor, Don Burn, for helping me through this long process without whom, this thesis would not exist
- My parents for providing me with everything I needed in order to make this possible
- David Bonin, for all his assistance and for being my sounding board for the new programming ideas and for the Mann Kendall program he wrote
- And my husband, for having to put up with me through this entire process and having to listen to me figure out what kept going wrong.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

POT - Peaks Over Threshold
MEP - Mean Excess Plot
MPD - Mean Peak Day
SDMPD - Standard Deviation about Mean Peak Day
ADS - Annual Duration Series

CHAPTER 1: INTRODUCTION

A. Objectives

The objectives of this thesis are:

- to analyze flood data for seasonality measures and trends in flood magnitudes and the timing associated with the flood events
- to determine if climate variability is affecting flood magnitudes and the timing of flood events
- to compare two different geographical regions for seasonality measure and climate variability effects

The purpose of this thesis is to analyze flood data for seasonality measures and trends in the flood peaks. Two different geographical regions, Newfoundland and the Prairies, are analyzed to see how the results of seasonality measures and trends compare. A comparison between the results of the analysis for each geographical area will be made in order to determine the effects of climate variability on flood magnitudes and the timing of flood events. The effects of climate variability have been observed over the last few decades in many different areas of the world and in many different ecological systems. Due to these effects, it is important to discover if the changes in flood magnitudes and the timing of the flood peaks are truly effects of climate change, or simply a part of a larger pattern known as climate variability. Climate variability effects are best observed on a regional basis, where there is more consistency in weather patterns affecting the sites for analysis. Hence the need for distinct regions for analysis.

The effects of climate variability can have both positive and negative impacts on hydrologic variables. A positive trend indicates an increase in flood magnitudes or in the number of flood peaks occurring in a single year. This can negatively affect the region around the site in several ways. An increase in flooding can mean that farmland is not available for planting, crops can rot in the fields due to too much water or that the crops cannot be harvested. It can also have negative impacts on the lives of people living in the flooded area, as the 1997 Red River Flood has shown, wrecking houses and eroding land. A decrease in flooding can mean that drought conditions prevail, causing crops to fail due to lack of water. Also fishing can be severely curtailed due to a lower water level for fish to spawn and live in. These conditions do not necessarily apply when flooding decreases, but they are possibilities. A decrease in flooding can simply indicate that the land is able to absorb the inflow of water.

Flood data have been traditionally defined as a series consisting of the largest flow value in each year. The flow value is defined as the maximum daily flow or the maximum instantaneous value available for the day of record. The resulting series is known as the Annual Duration Series (ADS). This approach has several disadvantages however. The assumption that there is one flood peak per year can be erroneous depending on the region to be analyzed. In some regions, such as Newfoundland, there can be three to four occurrences of flood peaks in a single year. In regions such as the Prairies, however, there is typically one flood peak every year. The Peaks Over Threshold (POT) approach allows for the possibility of there being more or less than one peak per year. This has the advantage of ensuring that peaks of smaller magnitudes than the largest peak in the year can be included in the analysis, ensuring that all are considered "floods". A flood is defined as a flow event of a magnitude that will overflow the banks normally containing the water body.

The most common reason for defining flood data is to perform flood frequency analysis. Flood frequency analysis is used to determine the return period for floods of various magnitudes. Using the ADS approach requires that there be a single peak flow value for each year of the record being analyzed. The POT Extraction results in a series which may have several peak flow values in a single year, and no peak flow values in a different year. This allows smaller floods to be defined and included in flood frequency analyses. Floods with return periods of less than two or three years are not easily found in the ADS analysis. Using the POT Extraction technique, the smaller floods that can occur once or twice a year can be included in the flood frequency analysis. The POT Extraction technique does not allow for traditional time series analysis because there may be no flood peaks in one year and several peaks in another year. Adaptations to existing analysis techniques are required and can be made.

B. Scope

The literature search in Chapter 2 reviews previous research contributions. This review presents the different Peaks Over Threshold (POT) approaches available, and an overview of some of the different analysis approaches available. It continues with a review of literature available on climate variability effects. In addition, a summary of regional analysis techniques is presented.

In Chapter 3, the different analysis approaches used in this study are presented. Chapter 3 begins with an overview of the POT approach used in this study. Separate sections are allocated to the issues of selecting a suitable threshold value, regionalizing the sites, adapting the trend test for analyzing unequally spaced data sets and the seasonality analysis of the day of flood peak occurrence.

The development of the POT Extraction package, where the flood peak data are obtained, is presented in Chapter 4. The selection of the data sites is considered and an overview of the sensitivity of the POT Extraction technique is presented in Chapter 5. Chapter 5 also presents the various techniques used to analyze the flood peak data obtained by the POT Extraction technique and the application of these analysis techniques in the Prairies and Newfoundland.

A discussion of the results and a comparison of the results for the two regions analyzed are presented in Chapter 6 and conclusions and recommendations for further research are presented in Chapter 7.

CHAPTER 2: LITERATURE REVIEW

Flood data has been traditionally defined as a series consisting of the largest flow value in each year. The flow value is defined as the maximum daily flow or the maximum instantaneous value available for the day of record. The resulting series is known as the Annual Duration Series (ADS). This approach has several disadvantages however. The assumption that there is one flood peak per year can be erroneous depending on the region to be analyzed. In some regions, such as Newfoundland, there can be three to four occurrences of flood peaks in a single year. In regions such as the Prairies, however, there is typically one flood peak every year. The Peaks Over Threshold (POT) approach allows for the possibility of there being more or less than one peak per year. This has the advantage of ensuring that peaks of smaller magnitudes than the largest peak in the year can be included in the analysis, ensuring that all are considered "floods". A flood is defined as a flow event of a magnitude that will overflow the banks normally containing the water body.

A. Peaks Over Threshold Extraction

A literature review of several papers was used to determine what criteria to use in order to extract the peaks from a daily flow record. Each paper presented a different set of criteria for determining peak flows and the independence of adjacent peak flows.

Before beginning to select the peaks, a threshold must be selected. According to Bayliss and Jones (1993), the threshold is initially chosen to give more than five peaks per year (approximately 1.5 times the average daily flow value), and is then raised until an appropriate level of flood peaks is reached. Birikundavyi and Rouselle (1997a) disagree. Their research uses a fixed number of peaks per year for the base level selection that is dependent on the prevailing climatic conditions.

Once the threshold is selected, the data are examined and the peak flows are extracted. These peak flows are then subjected to independence testing in order to remove excess peaks. There are two tests that the peaks must pass in order for them to be considered independent. The first independence test considers the number of days separating the two events. According to Birikundavyi and Rouselle (1997a) flood peaks are considered independent if they are separated by 7 days. Madsen et al (1997) disagree, stating that peaks are to be considered independent if they are separated by $5 + \ln A$ days, where A is the catchment size in square miles. Bayliss and Jones (1993) use yet a third option, stating that independent peaks should be separated by 3 times the average time to rise. The time to rise is defined as the average time difference between the start of the rising limb and the peak calculated from five clean flood hydrographs whose peak is above the threshold.

The flow values between the adjacent flood peaks must also be tested in order for the peaks to be considered independent. According to Birikundavyi and Rouselle (1997a), the flow in the time between the adjacent peaks must drop below 50% of the smaller of the two peaks. Madsen et al (1993) disagree, using values of 75% less than the smaller of the two peaks. Bayliss and Jones (1993) use a value of 2/3 less than the value of the first of the two peaks.

B. Regionalization

The accurate estimation of flooding potential at a site is often required for effective development and management of water resources at that site. The hydrologic characteristics of a catchment must be understood in order to obtain a reliable estimate of the extreme flow quantiles and their associated recurrence intervals. In order to do this, a sufficiently long streamflow record is required. However, the streamflow record is often much shorter than desired. In these situations, a regional approach to flood frequency analysis can be used. Regionalization is used to define a catchment that contains sites sufficiently similar to compensate for a short data record. The main goal of regionalization is the identification of a group of sites that are sufficiently similar to warrant the transfer of extreme flow data. Information from similar sites can be used to supplement the available at site information or to assist in characterising an ungauged site (Burn et al, 1997).

There are several different ways of regionalizing catchments available. Regionalization can be performed on basis of geographical location and similarity between site characteristics (Sceviour, 1993), or regionalization can be performed using a series of test statistics that characterize the sites and compare them to one another (Hosking and Wallis, 1993). Regions can also be created using cluster analysis (Tasker, 1982) and regional regression models (Wandle, 1977). With such a variety of methods available for creating regionalized catchments, it is difficult to determine which method should be used. Each of the available methods has different applications and therefore, consideration should be given to the results required. In this case, the methods used are geographical regionalization and the use of the test statistics developed by Hosking and Wallis (1993).

The region created based on geographical location and site characteristics is very general and the sites can be very different within a region because of the generality of the requirements. This method should be used only when comparisons are being made between regions and there are no requirements that data be transferred to sites that have shorter flow records.

Regions created by the other methods are more labour intensive, but create regions based upon similarities in physical site characteristics such as drainage area, catchment slope, etc. and flood statistics such as L-moment ratios and other statistics calculated from the annual flood series. Difficulties occur when using physical characteristics as similarity variables because similarity in a specific physical characteristic does not necessarily imply similarity in catchment response. This difficulty does not arrive when using flood statistics as a similarity variable (Burn et al, 1997). This method defines a region more accurately than one defined by a simple geographical area.

C. Climate Variability Effects

Climate variability effects have been discussed for nearly two decades, and still scientists disagree on the causes and the effects on hydrological variables. Indeed, there are still disagreements as to if what is occurring in the world is part of the natural cycle of the weather patterns or if the weather events are affected by pollution and deforestation (Mannion and Bowlby, 1992). With so many disagreements, it is difficult to ascertain if the changes discovered in hydrologic events are natural. There are indications that the changes are not natural however (Burn, 1994b). Efforts have been made to monitor the effects of climate variability by streamflow (Pilon et al., 1991) levels of lakes (Churchill et al., 1978) and tree ring growth rates (Lawford, 1988). It is believed that hydrological and hydrometerological measurements could be used to detect any possible indication of climatic change (Westmacott, 1996).

Various researchers have hypothesized that global warming will lead to changes in the timing of hydrologic events. Detecting changes in the hydrologic regime that might result from climate variability is complicated due to the inherent variability and randomness in nearly all hydrologic variables. There are several possible climate variability impacts on hydrologic variables that can be identified based upon previous work (Burn, 1994b). Burn (1994a) has determined that there is evidence to suggest that the minimum monthly flows and winter flow conditions are increasing with time. There are also indications that the mean annual flow and the daily peak flow are sensitive to changes in temperature. These results are site specific to the Canadian Prairies, and should not be used as an assumption that all areas are experiencing the same conditions.

There are other studies that investigate the effect of climate variability on water resources systems. Tasker (1993) investigated drought risks in the three river basins draining approximately 19 000 square miles in Georgia, Alabama and Florida. He showed that lower flows occur more frequently if the regional climate warms and the monthly precipitation decreases during the critical summer months. There are also potential impacts on streamflow from climate variability. Nkemdirim and Purves (1994) developed a multiple regression equation relating the historical mean annual streamflow in the Oldman River Basin in Alberta to the mean annual temperature and precipitation. Their results indicate that streamflow will decrease if temperature rises and that temperature is six times as important as precipitation in determining changes in streamflow. This emphasizes the importance of temperature in the evaluation of climatic change impacts. Streamflow measurements are essential in order to provide an early indication of the impacts of climatic change on water resources.

The effect of climate variability on extreme flow values is also of interest. Burn (1997) indicates that it has become apparent that the magnitude and direction of changes for particular hydrologic variables depend on the geographic area. For the Prairies, it is demonstrated that the flood magnitudes are decreasing and flood events are occurring earlier in the year. The relationship between trends in the flood variables and trends in a temperature series implies that the sites and their extreme flow values are being affected by climate variability within the region.

CHAPTER 3: BACKGROUND

A. Introduction

There are several different techniques used for extracting and analyzing the flood peak data. The relationship of the various techniques is explored, as the final set of data used for the analysis requires several iterations of three of the techniques to be performed. The initial data set is extracted using the Peaks Over Threshold Extraction technique. These data are then analyzed using a Mean Excess Plot and the resulting data set is regionalized. Using the results from the regionalization, the Peaks Over Threshold Extraction technique is performed again to extract the final data set. The final data set is divided into three different categories for trend analysis and is then analyzed using the Mann Kendall test for trends. A seasonality analysis is also performed on the final data set.

B. Peaks Over Threshold Extraction Technique

Peaks Over Threshold Extraction (POT) is a technique that allows flood peaks to be extracted from a daily flow record. In essence, it is a method of defining flood peaks that differs from the standard Annual Maximum flood series. The Annual Maximum or Annual Duration Series (ADS) requires that there be a single peak value for each year of the record being analyzed. This value is the maximum flow value for the year. The POT Extraction technique allows there to be multiple peak values or no peak flow values for any year of the record according to the predefined threshold.

The POT Extraction allows extraneous data to be discarded, and leaves only the peak flow values above some predefined threshold. This allows for a greater number of flood peaks to be analyzed and will represent the actual events more accurately than a single peak value per year would. The assumption that only one flood event occurs in a single year is flawed as many catchments can have more (or less) than one flood event in a year.

When compared with the Annual Maximum flood series, which uses the largest daily (or instantaneous) value of each year of the flow record, the POT technique has several advantages. By extracting flood peaks using a specific set of predetermined criteria, a more appropriate number of flood peaks for the catchment can be analyzed. A traditional time series analysis is not possible with the flood data extracted by the POT Extraction technique, but other analyses are used, as the main concern is not a time series analysis of the flood data.

C. Mean Excess Plot

The first analysis the initial data set is subjected to is the Mean Excess Plot. This technique allows the initial data set to be examined in order to determine if the initial number of peaks occurring per year is an accurate representation of the data and if not define an "appropriate" threshold. In order to perform this analysis, several sets of information are required. The first three pieces of information required are the initial threshold calculated by the POT Extraction program; the number of years of the record; and the initial number of peaks occurring per year. Finally, the peak flow values and their related counters are required.

Using this information, the initial threshold is increased in increments of ten percent. The threshold values are then subtracted from the peak flow values, and a mean residual value and a new value for the number of peaks occurring per year is calculated. The mean residual value is defined as the average of every peak flow value minus the threshold value yielding a positive result. A graph of the threshold values versus the mean residual values is plotted for further analysis.

The plot is then carefully examined in order to determine the threshold. A general guideline is to use the point at the beginning of a straight descending line (see Figure 1) as the new threshold (Naden, 1992). The point at the beginning of a straight descending line is chosen as the new threshold because this indicates the maximum threshold and maximum mean residual value. This guideline has its drawbacks, as often the plots have several sets of peaks in them, or no clearly defined straight descending line. This can lead to difficulty in selecting an appropriate threshold for a specific data set (see Figure 2).

As seen in Figure 1, a well behaved Mean Excess Plot can have several small sets of peaks, followed by a straight line descending from the preferred threshold. A decision is easily made, using the suggested guidelines. Little interpretation is required for clearly defined plots such as this one. The threshold chosen for this plot is 17.670 m^3/s as shown on the plot in Figure 1, which yields 0.88 peaks per year.

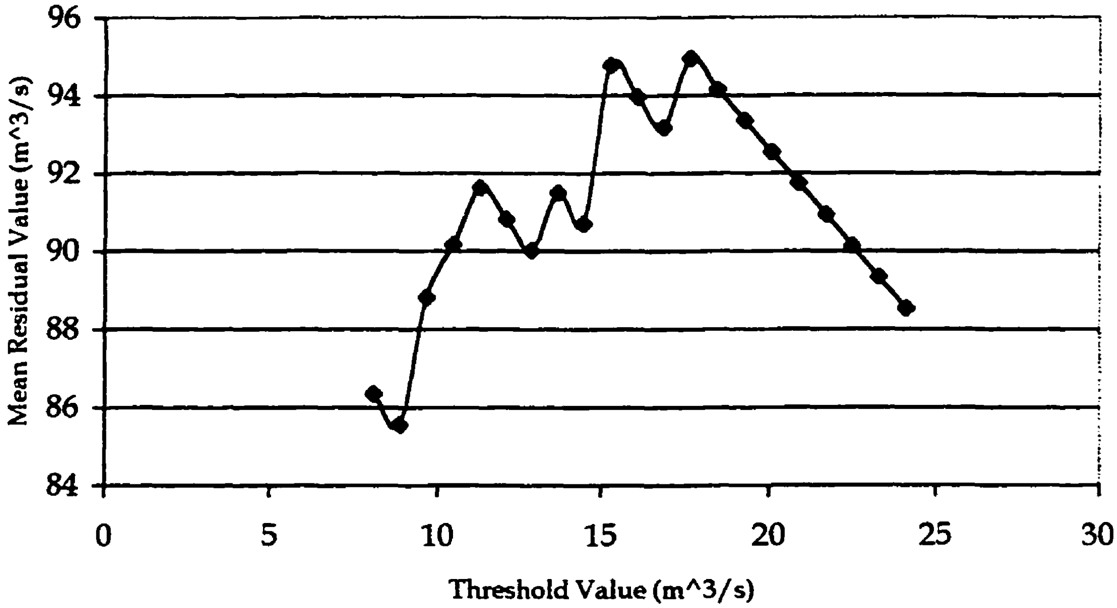

Figure 1: Carrot River at Armley Mean Excess Plot

Figure 2, however, demonstrates a poor example of a Mean Excess Plot. This plot has a line that meanders slightly, a series of many small peaks and no clear straight descending line. An initial decision must be made, so the new threshold is chosen as the first peak on the descending line. This choice yields a threshold that allows only a minimal number of peaks to pass, indicating that other criteria are required in order to determine the preferred threshold. In this plot, the chosen threshold is 72.601 m^3/s, which yields 0.18 peaks per year. This value is clearly unsuitable as there are very few places in the world which have less than one flood every six years. The original number of peaks per year for this site is 8.16 or a threshold of 13.445 m^3/s, a much greater number of peaks per year than the value selected by the Mean Excess Plot. Obviously further refinement is necessary to obtain a reasonable value.

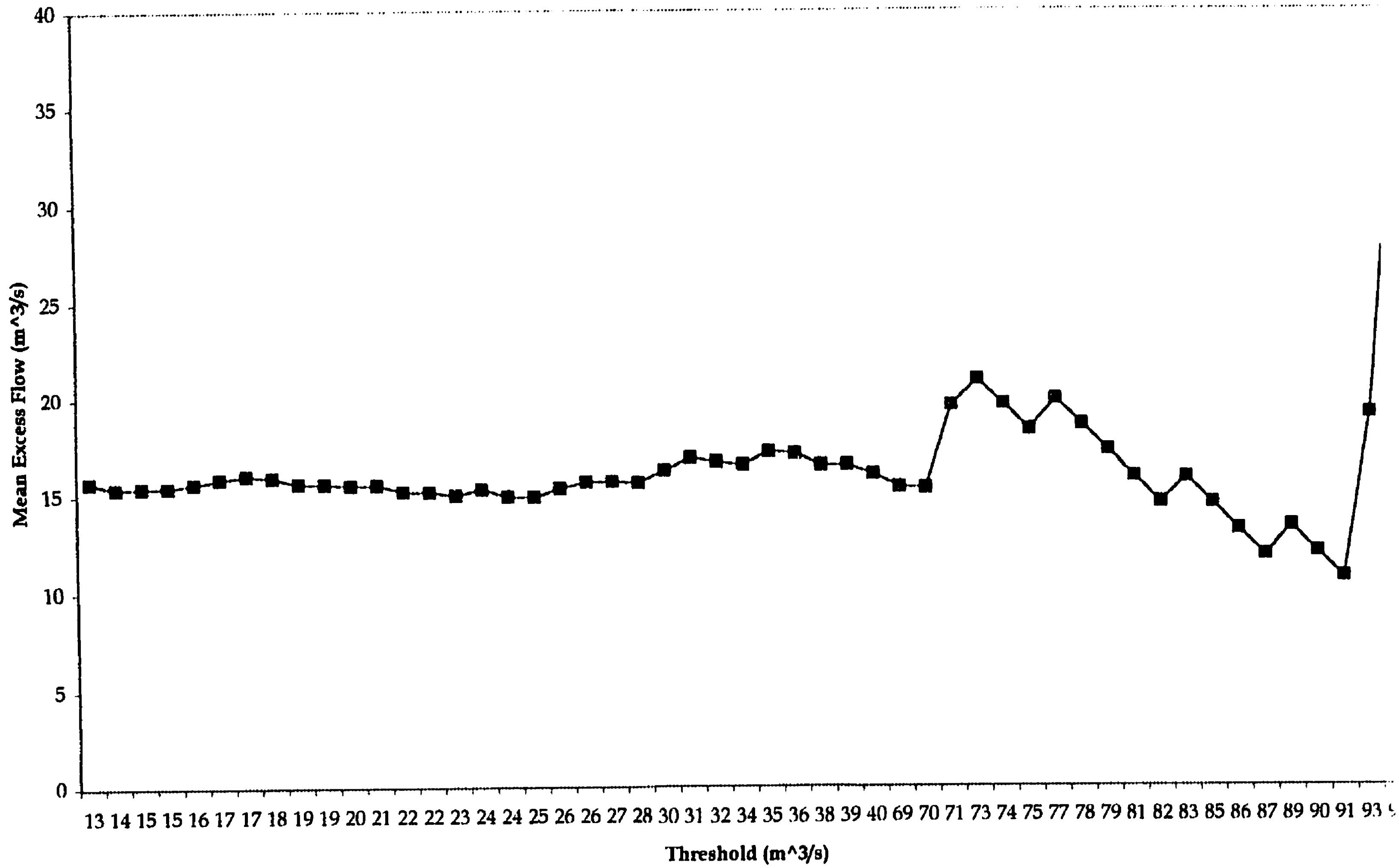

Figure 2: Garnish River at Garnish River Mean Excess Plot

B. Regionalizing The Data

The second analysis the initial data sets undergo is regionalization. By regionalizing the sites, the results of the Mean Excess Plot become clearer. It can be noted that various sites in a geographical region tend to exhibit commonalties in flood peak occurrence and timing. By regionalizing the sites, these commonalties are more evident.

The regionalization is performed two ways. The first regionalization method is to group various sites together based upon geographical location. The second method is to use seasonality data and the distance between the sites to form regions. The regions were then revised if they were not homogenous as determined by the Hosking and Wallis (1993) homogeneity test, which is based upon L-Moments. As the sites for the regions were drawn from two previous works, Burn et al (1997) and Sceviour (1993), the regions were formed using their methods. Some regions taken from Burn et al (1997) did not have more than one site per region, necessitating that two or more regions be combined based upon geographical location. The regions used in Burn et al (1997) had more than one site, but the subset selected resulted in a limited number of sites in some regions. Using these two methods, the sites are regionalized, and an average value of number of peak occurrences per year (regionalized peak occurrences value) is calculated for each region based upon the initial Mean Excess Plot results. Using this information, a more appropriate value for the number of peak occurrences per year is calculated.

In order to determine the "final" value for the number of peak occurrences per year, a chart was created containing by region: the initial peaks per year value; the new peaks per year value from the Mean Excess Plot; and the regionalized peaks per year value. The regionalized peak occurrences value is calculated by averaging the values for the number of peak occurrences in a year for all of the sites in the region. In many cases, the new peaks per year value was obviously inappropriate, as demonstrated in a previous example. Using the regionalized peak occurrence value, the final value for the number of peak occurrences per year was adjusted to be within plus or minus 0.50 peaks per year of the regionalized peak occurrence value. This allowed the peak occurrences to be more consistent within a region.

C. Mann Kendall Tests

The final data sets are analyzed using a trend test. The trend test selected was the Mann Kendall non-parametric test for trend (Mann, 1945, Kendall, 1975). This trend test has been used by other researchers in similar applications and has been found to be an effective tool for identifying trends in hydrologic variables (Burn, 1994a). The test statistic for the Mann Kendall test is given as

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}(x_j - x_k) \tag{1}$$

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

$$\operatorname{sgn}(\theta) = \begin{cases} 1 & if \theta > 0 \\ 0 & if \theta = 0 \\ -1 & if \theta < 0 \end{cases} \tag{2}$$

The theoretical mean and variance of the test statistic, under the null hypothesis of no trend in the series, are given as

$$E[S] = 0 \tag{3}$$

and

$$Var[S] = \frac{n(n-1)(2n+5) - \sum t(t-1)(2t+5)}{18} \tag{4}$$

where t is the extent of any tie (i.e. the number of data points involved in a tie) and the summation is over all ties. For sample sizes larger than ten, the statistic is very nearly normally distributed if a continuity correction is applied, giving

$$S' = S - \operatorname{sgn}(S) \tag{5}$$

where S' is the corrected test statistic value. Assuming the corrected tests statistic follows the normal distribution, a Z value associated with the trend statistic can be calculated as

$$Z = \frac{S'}{(Var[S])^{\frac{1}{2}}} \tag{6}$$

where Z is a standard normal value. It is also possible to obtain a non-parametric estimate for the magnitude of the slope following Hirsch et al. (1982)

$$\beta = \mathrm{Median}\left(\frac{x_j - x_k}{j - k}\right) \forall k < j \tag{7}$$

where β is a robust estimate of the slope.

D. Seasonality Analysis

The seasonality analysis is performed on the final set of peak values for each site. The analysis requires the year of occurrence and the day of occurrence of each peak in the data set. Using the seasonality test developed by the Institute of Hydrology in Wallingford, England (Bayliss and Jones, 1993); the following calculations are performed on the peak flow data. The year of occurrence is used to determine what value should be used for the length of the year in the formulas. The value for the length of the year is denoted as LENYR, and will be 365 or 366 days depending on if the year in question is a leap year. The day of occurrence in the year is converted into an angle to ensure that day 1 and day 365 are adjacent to one another. The theta value is calculated using

$$\theta = \left(day \times \frac{2\pi}{LENYR} \right) - ADJUST \tag{8}$$

where *day* is the day of occurrence of the peak flow value and LENYR is the length of the year (either 365 or 366) depending on if the year in question is a leap year. The adjustment, ADJUST, is a correction required because the flood peaks have been logged at various times during the day. The adjustment is made to each value to show the POT day represented at its midpoint.

$$ADJUST = \frac{1}{2} \times \left(\frac{2\pi}{365} \right) \tag{9}$$

Once the day of the year has been converted a set of x and y coordinates are calculated in order to locate the occurrence on a circular graph. The angles move in a counter clockwise direction from the x-axis.

The x and y coordinates for the angle are calculated by taking the cosine and sine values of theta respectively. There is one set of coordinates for each peak flow value and the corresponding day of occurrence. The mean of these coordinates is taken in order to determine the mean day of occurrence. The following formula is used to determine this value,

$$\bar{\theta} = \tan^{-1}\left(\frac{\bar{y}}{\bar{x}} \right) \tag{10}$$

where $\bar{x}$ and $\bar{y}$ are the mean of all the pairs of coordinates for each peak flow value. The mean direction is converted back to a day value using the following formula. A half is added to the MPD (Mean Peak Day) to compensate for the subtraction of half a day during the conversion of day numbers to angles. The value is then rounded to the nearest day number to be expressed as a calendar day. This indicates the actual day of occurrence. The following formula is used to calculate the MPD,

$$MPD = \left(\bar{\theta} \times \frac{365}{2\pi} \right) + 0.5 \tag{11}$$

where $\bar{\theta}$ is the mean theta value of all of the data points.

Using the information the above equations provide, the following type of graph is produced. The POT days are represented by points falling on the unit circle, while the MPD is the data point by itself in the plot (see Figure 3). It should be noted that this particular plot has a data point in the fourth quadrant, representing a fall flood (October to December). The quadrants move in a counter clockwise direction, with the x - positive, y - positive quadrant considered the first quadrant. Data points in the other three quadrants represent floods in January to March, April to June and July to September respectively. The plot shown in the following figure is of a site located in the Prairies. Most floods in the Prairies tend to occur in the early spring, as can be seen by the greater concentration of data points in the second quadrant. Floods in Newfoundland tend to occur throughout the year.

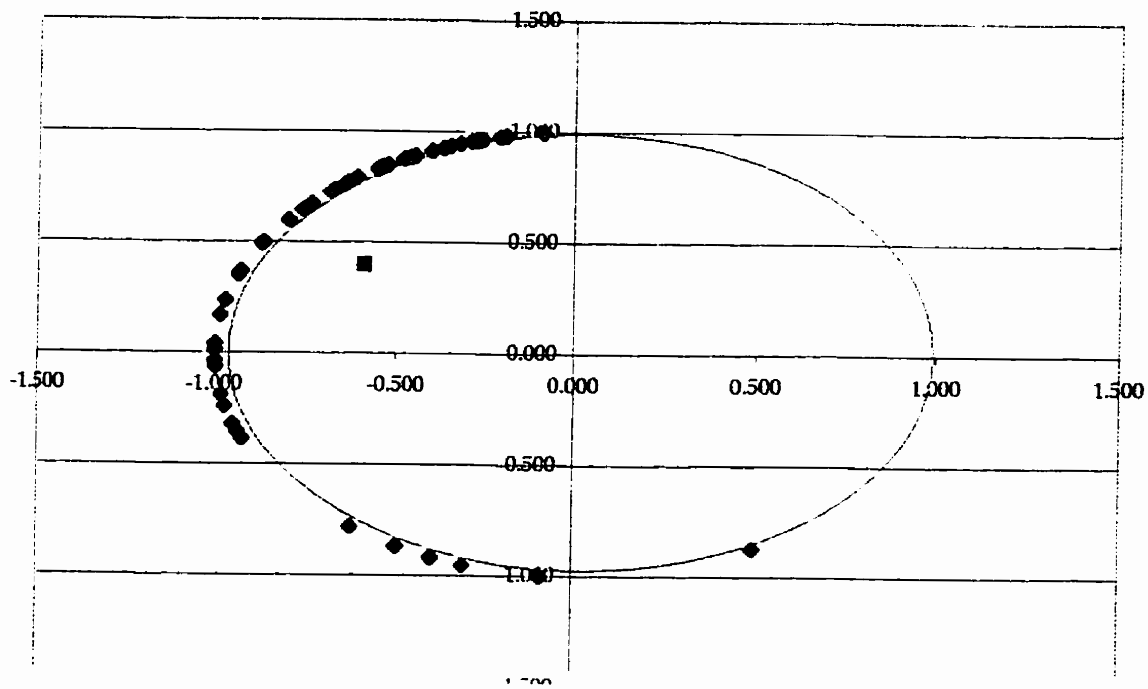

Figure 3: Overflowing River at Overflowing River Mean Day of Flood Plot

The MPD is the average of all POT dates for the record. While most floods often occur at approximately the same time of the year, there is likely to be flooding at other times during the year which will influence the position of the MPD in the year. For this reason, the standard deviation and the mean resultant are calculated to determine if there is strong seasonality in the data; or if there are floods occurring during other times in the year that are affecting the seasonality of the MPD. The mean resultant is calculated using,

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2} \tag{12}$$

where $\bar{x}$ and $\bar{y}$ are the mean of the x and y coordinates.

If the value of the mean resultant is close to one, then there is a strong seasonal component in the day of flood. If, however, the value of the mean resultant is close to zero, then the data are not strongly seasonal and the MPD value is less meaningful. A standard deviation of the circular data is calculated using the mean resultant value and

$$S_o = \sqrt{-2\ln \bar{r}} \tag{13}$$

where S_o is the standard deviation.

This provides a standard deviation in radians that can be converted to a standard deviation in days about the mean POT day, SDMPD, by

$$SDMPD = S_o - \left(\frac{365}{2\pi}\right) \tag{14}$$

The value is rounded to the nearest whole number day value. If the flooding is confined to a particular time of the year, then the SDMPD value is small and the MPD value indicates when flooding is most likely to occur. In sites where the SDMPD value is quite high, the MPD value is less likely to represent a period when flooding will typically occur. For the previous example, the mean resultant value is 0.714 and the SDMPD is 48 days. These values indicate that the seasonality of the data is not strong, but can be used to predict flood periods with a reasonable amount of accuracy.

The choice of the threshold can affect the results of seasonality. In the following example, a comparison of the peak flow values for Garnish River near Garnish in Newfoundland is made following the selection of the peak flow values by the initial threshold and the final threshold.

In Figure 4, the Mean Day of Flood is calculated using the initial results from the POT Extraction, comprising some 300 peaks over the course of 38 years of record. It can be seen that there are occurrences of flooding at all times during the year. The Mean Peak Day (MPD) for this plot is 28 (January 28 of the year) with a standard deviation of 89 days and a mean resultant value of 0.309. This indicates that the flooding is not strongly seasonal.

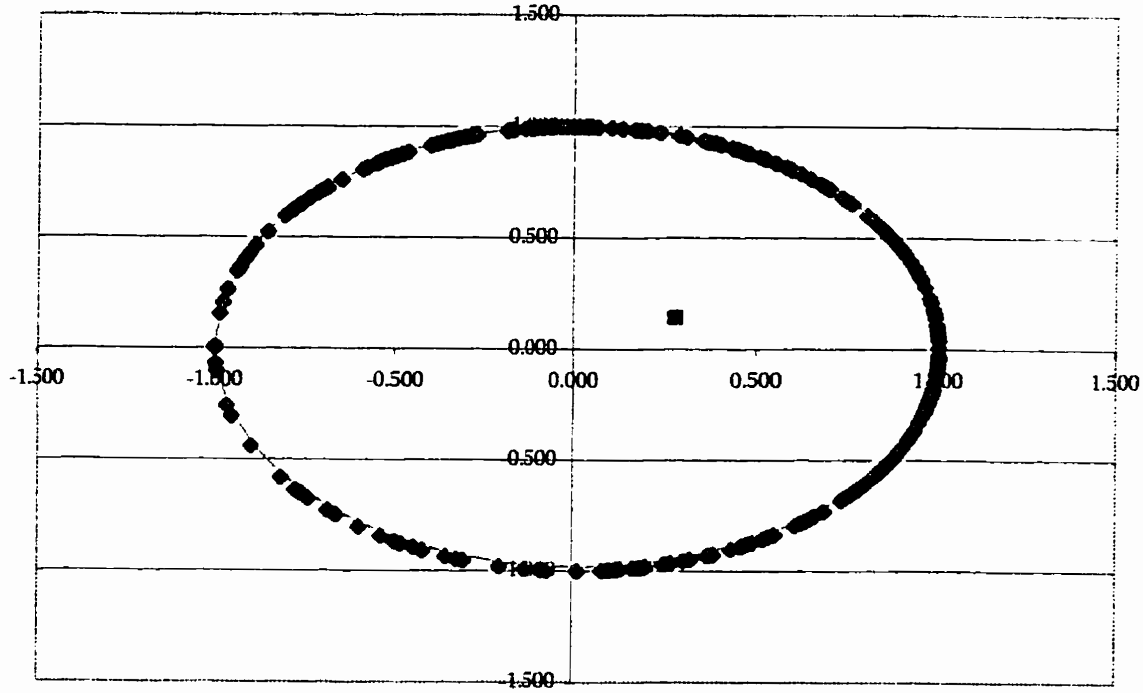

Figure 4: Garnish River near Garnish Initial Mean Day of Flood

By comparison, Figure 5 shows the Mean Day of Flood calculated using the final results from the POT Extraction, comprising 92 peaks in 38 years of record. It can be seen that there are occurrences of flooding at all times during the year. The Mean Peak Day (MPD) for this plot is 49 (February 18 of the year) with a standard deviation of 80 days and a mean resultant value of 0.389. This also indicates that the flooding is not strongly seasonal, but it can be seen that most of the floods occur in the winter months with more sporadic occurrences of flooding during the remainder of the year.

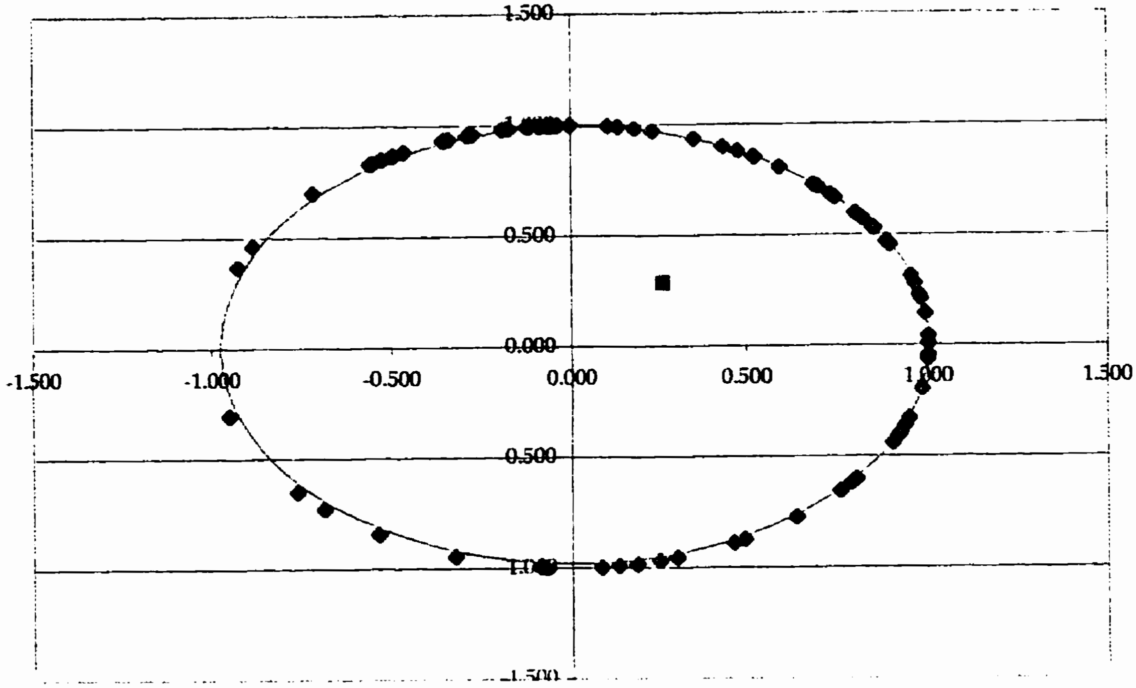

Figure 5: Garnish River near Garnish Final Mean Day of Flood

Comparing the two plots, both indicate that the majority of the floods occur in the winter months. The plots differ at this point. The initial results indicate flooding can occur at any point in time during the year whilst the final results indicate sporadic flooding during most of the year. It is obvious that the choice of the threshold has a significant effect on the seasonality results.

CHAPTER 4: DEVELOPING THE PEAKS OVER THRESHOLD EXTRACTION PROGRAM

A. Developing The Criteria

In order to develop the extraction technique, the criteria to be used in the analysis had to be determined. Using the information found in the literature search, decisions were made about which criteria were suitable. The initial threshold is generally assumed to be between 1.5 and 1.75 times the average daily flow of the record. After some consideration, it was decided that an initial threshold of 1.5 times the average daily flow value of the record would be used (Bayliss and Jones, 1993). By experimentation it was determined that a threshold of 1.75 times the average daily flow would eliminate too many possible peak flow events for sites with limited records. The program allows the user to review the results of the search and decide if the threshold needs to be lower or higher based upon the initial results.

In order to extract the peak flow values, a set of criteria had to be met. The criteria determined if the peak flow values were the result of independent flow events or if the peak flow values were dependant on another flow event. In order to perform this independence testing, several criteria were used. To determine the required number of days between peaks for independence, a value for the formula $5 + \ln A$ was calculated (Madsen et al., 1997), where A is the drainage area of the basin. The area was calculated by converting the gross drainage into square miles. If there was a net drainage area, then the net drainage area was used. A comparison was made between the formula value and seven days (Madsen et al., 1997; Bayliss and Jones, 1993). The smaller of the two values was used as the minimum number of days between peak flow values in order for the peak flows to be considered independent. Therefore, the separation in time between two peak flow values must be greater than the minimum number of days previously calculated. Then a comparison is made between each set of two peaks in order to determine the minimum flow values. In order to be considered independent of each other, the flow between each subsequent set of peaks must drop below 50% of the smaller of the two peaks. If either of these two criteria is not satisfied, the smaller of the two peaks is discarded, and the larger of the two is kept for the next comparison (Birikundavyi and Rouselle, 1997a). In Figure 6, an example of the process is shown.

All of the values below the threshold were discarded, as were the non-peak values above the threshold. Peaks kept for further analysis are indicated by a check mark in Figure 6. The first peak of the year is kept automatically unless there is another peak immediately following it that is larger and does not meet the minimum flow criteria. In this case, the next peak is smaller and the minimum flow drops below 50% of the smaller of the two peaks. The next two peaks are too close together and the flow does not drop below 50% of the smaller of the two peaks. The remaining peaks above the threshold that are kept meet the criteria of being separated by seven days and the flow drops below 50% of the smaller of the two peaks.

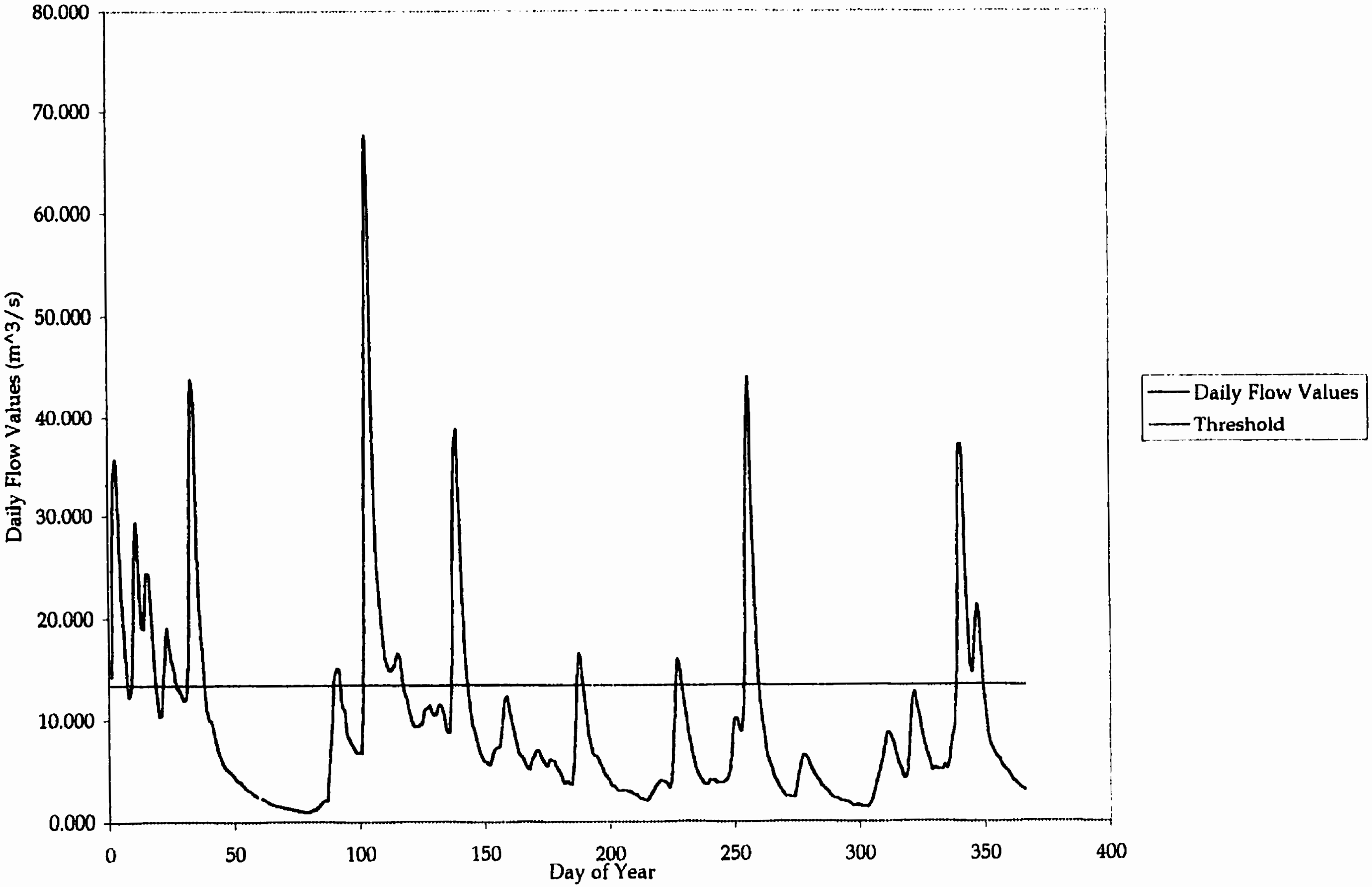

Figure 6: Peak Flow Values Independence Testing

The values in the flow record are compared to the threshold and below threshold values are discarded. The peak values are extracted from the remaining data and subjected to the independence tests. The number of days between each pair of peaks is calculated and compared to the previously determined value. If the value is less than the predetermined value, the smaller of the two peaks is discarded. If the separation is greater, both peaks are kept for the next round of testing. Each pair of peaks is tested for the minimum flow value. If the flow does not drop below half the value of the smaller of the two values between a pair of peaks, the lesser of the two values is dropped (See Figure 6).

B. Programming

Using Visual Basic for Applications (Excel), the POT Extraction analysis begins by importing the daily flow record taken from the HYDAT CD in Lotus format into the Excel worksheet and arranging the data into the proper format for analysis. Preliminary calculations are performed in order to determine the average daily flow value, the threshold which is 1.5 times the average daily flow value and the number of days required between peaks.

Once the initial calculations are complete, the daily flow values are compared to the threshold value and the program determines which flow values are larger than the threshold and which flow values should be dropped. The program then determines what the peak flow values are and removes all non-peak flow values.

The peak flow values are then subjected to the independence criteria, and the dependent peak flow values are removed. The peak flow values are finally compiled into a data set identifying the year, day of occurrence and magnitude of the peak. An average number of peaks per year is calculated based upon this data. The user can access the Mean Excess Plot program in order to determine the suitability of the threshold.

If the threshold is found to be unsuitable, there are several more steps available in the program that the user may access. The program will organize the peaks in ascending order, allowing the user to determine where the new threshold should be set. Once the user has decided upon a new threshold, a final program can be run that will distill all of the peaks from the previous data set. If the average number of peaks per year is still unsatisfactory, the program will allow the user to set a new threshold and extract all the peaks using that threshold.

CHAPTER 5: ANALYZING THE DATA

A. Selecting The Data Sets

The sites for the final analysis were selected based upon previous work by Burn et al. (1997) and Sceviour (1993). There were two hundred and seventeen available sites for analysis in the work performed by Burn et al. In order to limit the number of sites analyzed for this thesis, a subset of the two hundred and seventeen available sites were chosen. This subset of sites consisted of forty-two sites in different regions analyzed in Burn et al. As some of the regions consisted of one or two sites, the regions were combined into larger regions based upon geographical location. The sites in the Prairies used for the final analysis are shown in Figure 7 by region. The 23 sites in Newfoundland were grouped geographically (see Figure 8). The following table (Table 1) lists the sites, their location and the characteristics of the site. The characteristics include the drainage area, the number of years of record and the years of record.

The sites selected for the analysis were all "natural" data sets, meaning the sites were unregulated and showed no evidence of land use changes. The sites selected were in rural areas in order to eliminate urban runoff effects and negative impacts from urbanization. The regions in question are located in two different parts of Canada, and experience very different climate conditions. The sites located in the Prairies are land locked in the midst of Central Canada, although a few sites are located close to Lake Winnipeg and the Lake of the Woods. The climate in the Prairies is quite warm and dry in the summer with an annual precipitation ranging from 350 mm in Saskatchewan to 625 in Northwestern Ontario, with an annual snowfall ranging from 120 cm in Saskatchewan to 200 cm in Northwestern Ontario. Flooding is most likely to occur in the spring during the snowmelt runoff period, especially if heavy rains occur during the melt period. The sites located in Newfoundland are surrounded by the ocean and rainfall is quite heavy, especially during hurricane season in the winter months. The annual precipitation ranges from 860 mm to 1500 mm in Newfoundland, with snowfall ranging from 200 cm to 450 cm.

There are other differences between the two regions. The topography of Newfoundland is quite hilly and rocky with numerous fjords and inlets on the island. The land is used mainly for mining and the main industry is fishing. The Prairies are broken down into three main climatic zones: the Prairies, the Northeastern Forest and the Northwestern Forest. These zones can be broken down into the plains and the Canadian Shield. The plains are flat, with a few rolling hills and small stands of trees. The main land uses are farming and urban areas. The Canadian Shield is rocky and heavily forested. The main land use is forestry.

The Newfoundland sites have slightly shorter daily flow records available, ranging from 12 to 68 years in length, with most sites having 25 to 40 years of record available. They also have smaller catchment areas, ranging in size from 4 km^2 to 4400 km^2, with most sites having catchment areas of 300 to 800 km^2. The sites in the Prairies have longer daily flow records available, ranging from 35 to 82 years in length, with most sites having between 35 and 50 years of record available. They also have larger catchment areas, ranging from 88 km^2 to 13 900 km^2.

Figure 7: Sites for Analysis in Prairie Provinces

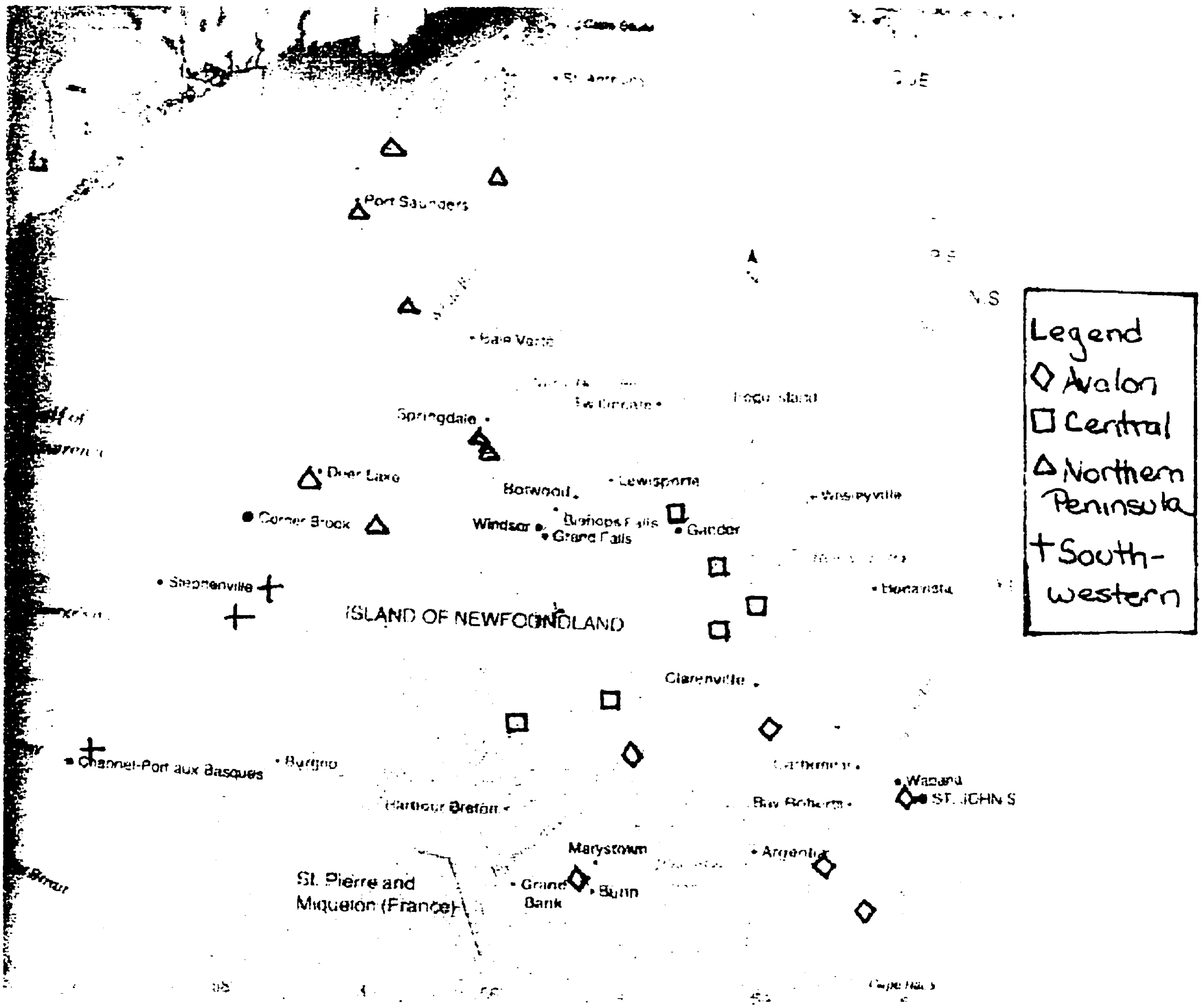

Figure 8: Sites for Analysis in Newfoundland

Table 1: Site Characteristics and Locations

Name	*Station Code*	*Latitude*	*Longitude*	*File Length*	*Years Of Record*	*Gross Drainage Area*	*Effective Drainage Area*
Newfoundland Sites							
Bay du Nord River at Big Falls	02ZF001	47:44:48N	055:26:30W	46	1950 - 1995	1170	N/A
Beaver Brook near Roddickton	02YD001	50:54:51N	056:09:26W	20	1959 - 1978	237	N/A
Cat Arm River above Great Cat Arm	02YF001	50:04:33N	056:55:22W	15	1968 - 1982	611	N/A
Come by Chance near Goobies	02ZH002	47:55:07N	053:56:59W	28	1961, 1969 - 1995	43	N/A
Gander River at Big Chute	02YQ001	49:00:55N	054:51:13W	47	1949 - 1995	4400	N/A
Garnish River near Garnish	02ZG001	47:12:50N	055:19:45W	38	1958 - 1995	205	N/A
Harry's River below Highway Bridge	02YJ001	48:34:31N	058:21:48W	28	1968 - 1995	640	N/A
Hinds Brook near Grand Lake	02YK004	49:04:21N	057:10:46W	24	1956 - 1979	529	N/A
Isle aux Morts River below Highway Bridge	02ZB001	47:36:50N	059:00:33W	34	1962 - 1995	205	N/A
Lewaseechjeech Brook at Little Grand Lake	02YK002	48:37:20N	057:56:00W	40	1952 - 1967, 1972 - 1995	470	N/A
Middle Brook near Gambo	02YR001	48:48:28N	054:13:28W	37	1959 - 1995	275	N/A

Northeast Pond River at Northeast Pond	02ZM006	47:38:06N	052:50:14W	43	1953 - 1995	4	N/A
Northwest Brook at Northwest Pond	02ZN001	46:51:08N	053:18:11W	30	1966 - 1995	53	N/A
Pipers Hole River at Mother's Brook	02ZH001	47:56:49N	054:17:08W	44	1952 - 1995	764	N/A
Rocky River near Collnet	02ZK001	47:13:29N	053:34:06W	48	1948 - 1995	301	N/A
Salmon River at Long Pond	02ZE001	47:56:40N	055:54:50W	22	1944 - 1965	2640	N/A
Sheffield Brook near Trans Canada Highway	02YK005	49:20:11N	056:39:56W	24	1972 - 1995	391	N/A
Sheffield River at Sheffield Lake	02YK003	49:20:10N	056:37:57W	12	1955 - 1966	362	N/A
Southwest Brook at Terra Nova National Park	02YS003	48:36:25N	053:58:50W	29	1967 - 1995	37	N/A
Ste Genvieve near Forrestor's Point	02YA001	51:08:18N	056:47:32W	27	1969 - 1995	306	N/A
Terra Nova River at Eight Mile Bridge	02YS001	48:26:30N	054:22:21W	34	1951 - 1984	1290	N/A
Torrent River at Bristol's Pool	02YC001	50:36:27N	057:09:04W	37	1959 - 1995	624	N/A
Upper Humber River near Reidville	02YL001	49:14:26N	057:21:45W	68	1928 - 1995	2110	N/A
Prairies Sites							
Name	***Station Code***	***Latitude***	***Longitude***	***File Length***	***Years Of Record***	***Gross Drainage Area***	***Effective Drainage Area***

Arrow River near Arrow River	05MG001	50:05:10N	100:57:30W	37	1958 - 1995	670	N/A
Assiniboine River at Sturgis	05MC001	51:56:23N	102:32:49W	51	1944 - 1995	1930	1140
Carrot River near Armley	05KB003	53:08:11N	104:01:15W	41	1955 - 1995	4400	2450
Carrot River near Smoky Burn	05KC001	53:25:00N	103:08:30W	41	1955 - 1995	9250	7120
Conjuring Creek near Russell	05ME005	50:47:30N	101:17:55W	37	1959 - 1995	88	N/A
Cooks Creek at Cooks Creek	05OJ006	50:01:12N	096:46:23W	35	1960 - 1994	527	N/A
Crystal Creek near Crystal City	05OB006	49:07:28N	098:56:28W	36	1959 - 1962, 1964 -1995	135	N/A
Cypress Creek near Clearwater	05OB010	49:05:21N	098:58:54W	37	1959 - 1995	398	N/A
English River near Sioux Lookout	05QA002	50:04:15N	091:56:40W	61	1921 - 1981	13900	N/A
English River at Umfreville	05QA001	49:52:30N	091:27:30W	75	1921 - 1995	6230	N/A
Fisher River near Dallas	05SD003	51:21:20N	097:30:50W	35	1961 - 1995	1720	N/A
Indianhead Creek near Indian Head	05JL002	50:38:42N	103:36:14W	55	1941 - 1995	327	188
Little Saskatchewan River near Minnedosa	05MF001	50:21:28N	099:54:00W	54	1914 - 1930, 1959 - 1995	2630	N/A
Little Souris River near Brandon	05MH006	49:43:56N	099:50:52W	35	1961 - 1995	461	N/A
Main Drain near	05OD028	49:08:01N	097:08:42W	36	1960 - 1995	240	N/A

Dominion City							
Manning Canal near Ile des Chenes	05OE006	49:37:50N	096:54:20W	35	1961 - 1995	496	N/A
Mink Creek near Ethelbert	05LJ019	51:25:00N	100:21:10W	42	1954 - 1995	131	N/A
Netley Creek near Petersfield	05OJ008	50:19:38N	097:02:39W	36	1960 - 1995	645	N/A
North Pine River near Pine River	05LG001	51:49:00N	100:32:00W	41	1954 - 1994	210	N/A
North Shannon Creek near Myrtle	05OF015	49:20:34N	097:50:33W	35	1960 - 1994	165	N/A
Oak Creek near Stockton	05NG010	49:33:44N	099:30:49W	35	1961 - 1995	1030	N/A
Ochre River at Ochre River	05LJ005	51:03:04N	099:47:15W	61	1912 - 1930, 1948 - 1995	341	N/A
Overflowing River at Overflowing River	05LD001	53:09:15N	101:06:30W	42	1954 - 1995	3350	N/A
Pinewood River near Pinewood	05PC041	48:45:10N	094:14:00W	44	1952 - 1995	461	N/A
Rat River near Sundown	05OE004	49:12:35N	096:17:00W	36	1960 - 1995	398	N/A
Roseau River at Gardenton	05OD004	49:05:16N	096:41:15W	42	1915 - 1922, 1962 - 1995	4390	N/A
Roseau River near Caribou	05OD030	48:58:54N	096:27:46W	77	1917 - 1995	4080	N/A
Roseau River near Dominion City	05OD001	49:11:33N	096:59:04W	82	1913 - 1995	5260	N/A
Shannon Creek near	05OF014	49:21:17N	097:25:18W	36	1960 - 1995	649	N/A

Morris							
Shell River near Inglis	05MD005	50:57:40N	101:19:05W	48	1948 - 1995	2000	N/A
Sprague Creek near Sprague	05OD031	48:59:33N	095:39:43W	54	1928 - 1981	455	N/A
Stony Creek near Neepawa	05LL009	50:13:58N	099:34:00W	35	1959 - 1993	163	N/A
Stony Creek near Stenen	05MC002	51:52:58N	102:19:48W	36	1957 - 1992	254	153
Sturgeon River at McDougall Mills	05QA004	50:10:20N	091:32:30W	35	1961 - 1995	4450	N/A
Sturgeon River at Outlet of Salvesen Lake	05QE009	50:21:20N	094:27:30W	36	1960 - 1995	1530	N/A
Swan River near Minitonas	05LE006	52:11:47N	101:03:37W	36	1960 - 1995	4230	N/A
Turtle River near Laurier	05LJ007	50:56:40N	099:31:15W	57	1922 - 1929, 1947 - 1995	938	N/A
Turtle River near Mine Centre	05PB014	48:51:00N	092:43:30W	82	1914 - 1995	4870	N/A
Whitemouth River near Whitemouth	05PH003	49:56:20N	095:57:20W	54	1942 - 1995	3750	N/A
Willow Creek near Gimli	05SB002	50:34:50N	097:02:40W	31	1960 - 1991	253	N/A
Wilson River near Dauphin	05LJ011	51:12:30N	100:10:45W	39	1924 - 1928, 1948 - 1982	925	N/A
Woody River near Bowsman	05LE004	52:15:30N	101:08:15W	42	1954 - 1995	2110	N/A

The sites were initially analyzed using the POT Extraction program, and the results were tabulated (see Table 2). Using the initial site results, a random sampling was selected for sensitivity analysis.

Table 2: Initial Site Results

Name	*Initial Threshold Value*	*Initial Average Peaks Per Year*
Bay du Nord River at Big Falls	59.469	4.35
Beaver Brook near Roddickton	13.102	5.20
Cat Arm River above Great Cat Arm	40.264	7.40
Come by Chance near Goobies	2.878	13.50
Gander River at Big Chute	178.877	3.40
Garnish River near Garnish	13.445	8.16
Harry's River below Highway Bridge	39.841	8.79
Hinds Brook near Grand Lake	24.373	3.33
Isle aux Morts River below Highway Bridge	20.389	15.59
Lewaseechjeech Brook at Little Grand Lake	25.924	5.65
Middle Brook near Gambo	10.046	3.27
Northeast Pond River at Northeast Pond	0.204	16.95
Northwest Brook at Northwest Pond	4.719	13.67
Pipers Hole River at Mother's Brook	38.239	8.55
Rocky River near Colinet	16.871	14.21
Salmon River at Long Pond	128.267	2.59
Sheffield Brook near Trans Canada Highway	15.779	3.21
Sheffield River at Sheffield Lake	14.559	2.42
Southwest Brook at Terra Nova National Park	1.572	13.38
Ste. Genvieve near Forrestor's Point	13.099	2.26
Terra Nova River at Eight Mile Bridge	56.819	4.18
Torrent River at Bristol's Pool	37.441	4.38
Upper Humber River near Reidville	129.956	6.65
Arrow River near Arrow River	0.573	1.97
Assiniboine River at Sturgis	4.747	1.57
Carrot River near Armley	8.032	1.10
Carrot River near Smoky Burn	21.024	1.78

Conjuring Creek near Russell	0.178	2.03
Cooks Creek at Cooks Creek	**1.962**	**2.69**
Crystal Creek near Crystal City	0.263	1.97
Cypress Creek near Clearwater	**0.799**	**2.08**
English River near Sioux Lookout	180.804	0.98
English River at Umfreville	**86.001**	**1.11**
Fisher River near Dallas	206.593	2.31
Indianhead Creek near Indian Head	**0.264**	**2.02**
Little Saskatchewan River near Minnedosa	6.266	1.70
Little Souris River near Brandon	**0.649**	**1.83**
Main Drain near Dominion City	0.786	2.69
Manning Canal near Ile des Chenes	**1.548**	**2.46**
Mink Creek near Ethelbert	0.643	2.93
Netley Creek near Petersfield	**1.565**	**2.56**
North Pine River near Pine River	2.002	3.51
North Shannon Creek near Myrtle	**0.537**	**2.37**
Oak Creek near Stockton	0.920	1.91
Ochre River at Ochre River	**3.484**	**3.78**
Overflowing River at Overflowing River	19.129	1.50
Pinewood River near Pinewood	**5.922**	**4.02**
Rat River near Sundown	1.837	4.36
Roseau River at Gardenton	**16.534**	**1.86**
Roseau River near Caribou	15.495	1.88
Roseau River near Dominion City	**15.414**	**2.00**
Shannon Creek near Morris	1.726	2.11
Shell River near Inglis	**5.335**	**2.04**
Sprague Creek near Sprague	2.334	3.33
Stony Creek near Neepawa	**0.662**	**2.60**
Stony Creek near Stenen	0.417	1.06
Sturgeon River at McDougall Mills	**54.727**	**1.31**
Sturgeon River at Outlet of Salvesen Lake	15.464	1.28
Swan River near Minitonas	**9.373**	**2.11**
Turtle River near Laurier	5.246	3.54
Turtle River near Mine Centre	**55.627**	**1.37**
Whitemouth River near Whitemouth	20.501	2.80
Willow Creek near Gimli	**1.134**	**1.55**
Wilson River near Dauphin	5.317	2.51
Woody River near Bowsman	**10.427**	**2.45**

B. Mean Excess Plots

In order to achieve reasonable results from the Mean Excess Plots, it was necessary to prepare several sets of preliminary plots for each data set before making an initial choice for the new threshold. Using the initial extraction results, a series of new thresholds is calculated using increments of ten percent above the initial threshold. The residual value of the peak is then calculated, and an average number of peaks per year is determined. The mean of the residuals is plotted for each threshold, and an attempt is made to determine the "best" threshold based upon this data. The preliminary results are compiled into one graph (see Figure 9). Using this graph, a new threshold is selected based upon the position of the straight-line section of the graph. In this example, the new threshold value is 20.086 m^3/s or 1.43 peaks per year, as indicated on the graph. The initial threshold value is the initial data point, with a value of 19.129 m^3/s or 1.48 peaks per year.

This new threshold was used in the POT Extraction program to remove excess peak values, and a new value was calculated for the average number of peaks per year. The new threshold is entered into the program and peak flow values below this threshold were removed from the data set. This allows a "better" number of peak occurrences per year to be determined in order to reduce the number of extraneous smaller peak flow values. After all the sites were completed, the results were tabulated for further analysis by regionalization.

Once all of the results have been tabulated and analyzed on a regional basis, final decisions must be made in order to determine the final number of peaks per year for each site. Using the initial results from the regionalization, the regional peak occurrences value is determined. The regional peak occurrences value is defined as the average value of the number of peak occurrences for each site in the region. The number of peak occurrences for each site is then assumed to be within plus or minus 0.50 peaks per year of this value. Using this regional peak occurrences value each site in the region is then given an approximate final value for the number of peaks per year. Each plot is then examined to find the threshold that will yield a value close to the assumed value.

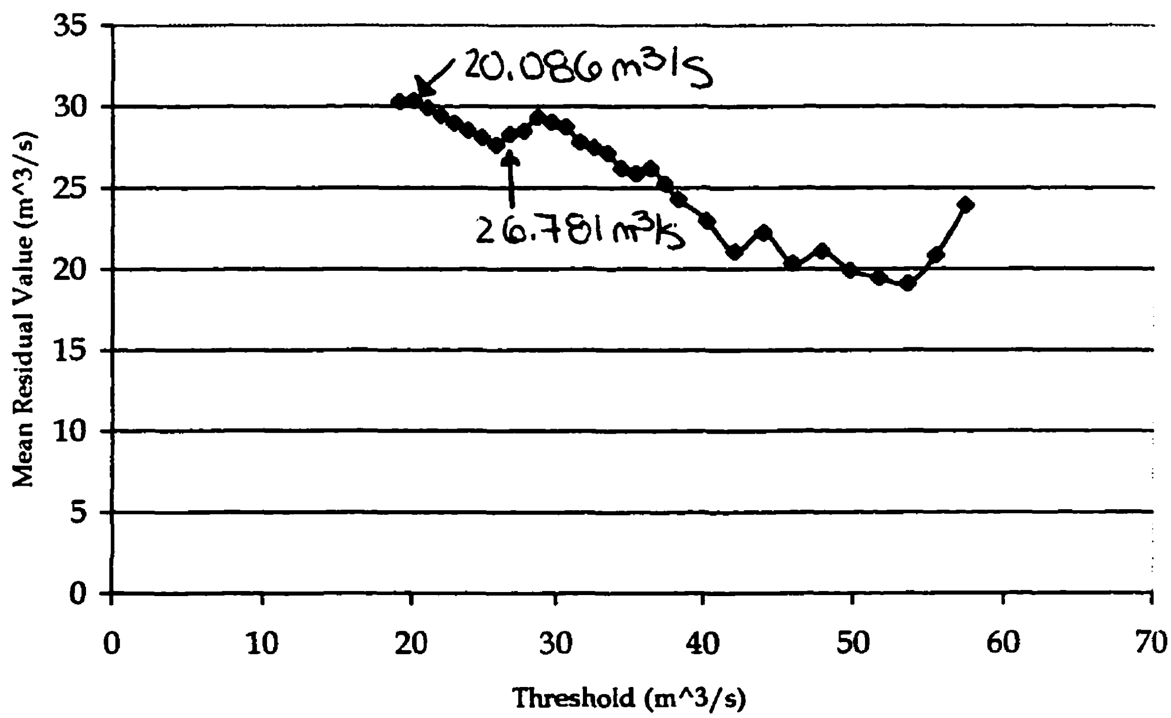

Figure 9: Overflowing River at Overflowing River Mean Excess Plot

A concentrated effort is made to follow the rules for analyzing a MEP. Therefore, whenever possible, the new threshold is chosen at the start of a descending straight line, as can be seen in the above figure. The initial selection for a new threshold was 1.43 peaks per year or 20.086 m^3/s (shown in Figure 9) and the final threshold was selected was 1.21 peaks per year or 26.781 m^3/s (also shown in Figure 9).

C. Regionalization

The sites were grouped into regions, based upon the work by Burn et al (1997) and Sceviour (1993). Using the Mean Excess Plot results, a regionalized value for the average number of peaks per year was calculated. As the Mean Excess Plot results were inconsistent at best, the regionalized average peaks per year were used to make the results more consistent.

The regionalization assisted in clearing up difficulties encountered during the initial assignment of new threshold values in the Mean Excess Plots. It was previously noted that some sites exhibited unusual variations in the plots, making it difficult to determine the best threshold value. Using the regionalized average peaks per year removed those difficulties.

It should be noted that the Newfoundland sites were more difficult to work with, as they had a higher average number of peaks per year, and the Mean Excess Plot results yielded thresholds that were extremely high. The thresholds chosen from the Mean Excess Plot were predicting less than one peak every two or three years, which is inconsistent with Newfoundland's tendency towards two or three flood events peaks per year. The Prairie sites were much more consistent in terms of the number of peaks per year selected, with the Mean Excess Plot results predicting approximately one peak per year. This is consistent with the Prairies tendency to have a single flood event in a year.

The new thresholds were chosen from the Mean Excess Plot tests, using the new values of number of peak occurrences per year to select the threshold. The results were tabulated for analysis. Again, regionalized peak occurrences per year values were calculated, and the results were carefully examined. This time the results were more consistent and the thresholds yielded acceptable values for the number of peak occurrences per year.

Using the values of number of peak occurrences per year obtained by analyzing each plot, final regionalized results were obtained. It should be noted that in the Newfoundland sites, there is a great deal of difference between the initially selected MEP values and the final selection (see Table 3). In addition, the number of peak occurrences in a year in each region has a greater range than does the number of peak occurrences in a year for the Prairie regions. This phenomenon could be attributed to the more varied weather patterns experienced by different areas in Newfoundland. The Prairie regions tend to experience similar weather patterns, and there is a smaller range in the number of peak occurrences in a year because of this (see Table 3).

Table 3: Regionalized Peak Occurrences per Year

Region	*Station Name*	*Number Of Initial Peaks*	*Number Of Selected Peaks (MEP)*	*Regional Peak Value*	*Final Selected Peaks Per Year*	*Regional Peak Value*
Avalon	Garnish River near Garnish	8.16	0.18	1.87	2.45	2.62
	Pipers Hole River at Mother's Brook	8.55	0.89		2.75	
	Come by Chance near Goobies	13.50	1.04		2.79	
	Rocky River near Colinet	14.21	0.65		2.48	
	Northeast Pond River at Northeast Pond	16.95	2.53		2.53	
	Northwest Brook at Northwest Pond	13.67	5.93		2.73	
Central	Gander River at Big Chute	3.40	1.34	2.61	2.77	2.79
	Middle Brook	3.27	3.27		2.68	
	Terra Nova River at Eight Mile Bridge	4.18	4.18		2.91	
	Southwest Brook at Terra Nova National Park	13.38	0.66		2.79	
	Salmon River at Long Pond	2.59	2.59		2.59	
	Bay du Nord River at Big Falls	4.35	3.63		2.98	

Northern	Ste Genvieve near Forrestor's Point	2.26	0.41	1.92	2.26	2.51
Peninsula	Torrent River at Bristol's Pool	4.38	2.62		2.62	
	Beaver Brook near Roddickton	5.20	2.45		2.45	
	Cat Arm River above Great Cat Arm	7.40	2.27		2.47	
	Sheffield River at Sheffield Lake	2.42	0.83		2.42	
	Hinds Brook near Grand Lake	3.33	1.83		2.63	
	Sheffield Brook near Trans Canada Highway	3.21	1.21		2.63	
	Upper Humber River near Reidville	6.65	3.71		2.62	
Southwestern	Harry's River below Highway Bridge	8.79	3.29	2.74	3.29	3.18
	Lewaseechjeech Brook at Little Grand Lake	5.65	2.03		3.18	
	Isle aux Morts River below Highway Bridge	15.59	2.91		3.06	

Region	*Station Name*	*Number Of Initial Peaks*	*Number Of Selected Peaks (MEP)*	*Regional Peak Value*	*Final Selected Peaks Per Year*	*Regional Peak Value*
1	Roseau River near Dominion City	2.00	1.63	1.23	1.10	1.18
	Roseau River at	1.86	1.43		1.19	

	Gardenton					
	Roseau River near Caribou	1.88	1.88		1.03	
	Rat River near Sundown	4.36	1.17		1.17	
	Turtle River near Mine Centre	1.37	1.37		1.11	
	English River near Sioux Lookout	0.98	0.44		0.98	
	English River at Umfreville	1.11	0.73		0.92	
	Sturgeon River at McDougall Mills	1.31	0.54		1.14	
	Sturgeon River at Outlet of Salvesen Lake	1.28	1.28		1.00	
	Pinewood River near Pinewood	4.02	0.64		0.95	
2	Indianhead Creek near Indian Head	2.02	1.20	0.87	1.09	0.86
	Assiniboine River at Sturgis	1.57	0.92		0.92	
	Stony Creek near Stenen	1.06	0.75		0.75	
	Carrot River near Smoky Burn	1.78	0.93		0.93	
	Carrot River near Armley	1.10	0.88		0.88	
	Conjuring Creek near Russell	2.03	0.73		0.73	
	Arrow River near Arrow River	1.97	0.70		0.70	

3	Overflowing River at Overflowing River	1.50	1.43	1.15	1.21	1.03
	Woody River near Bowsman	2.45	1.57		1.19	
	Swan River near Minitonas	2.11	0.94		0.94	
	Shell River near Inglis	2.04	0.90		0.90	
	Sprague Creek near Sprague	3.33	0.89		0.89	
4	Cypress Creek near Clearwater	2.08	1.24	1.08	1.24	1.09
	Stony Creek near Neepawa	2.60	1.11		1.11	
	Little Souris River near Brandon	1.83	0.89		0.94	
	Oak Creek near Stockton	1.91	1.14		1.14	
	Crystal Creek near Crystal City	1.97	1.00		1.00	
5	Whitemouth River near Whitemouth	2.80	2.59	1.85	1.43	1.47
	Cooks Creek at Cooks Creek	2.69	1.60		1.60	
	Manning Canal near Ile des Chenes	2.46	1.37		1.37	
6	Shannon Creek near Morris	2.11	1.64	1.42	1.44	1.44
	North Shannon Creek near Myrtle	2.37	1.03		1.31	

	Main Drain near Dominion City	2.69	1.58		1.58	
7	Willow Creek near Gimli	1.55	0.77	1.09	1.19	1.23
	Fisher River near Dallas	2.31	1.09		1.09	
	Netley Creek near Petersfield	2.56	1.42		1.42	
8	Ochre River at Ochre River	3.78	0.84	0.96	0.84	0.92
	Turtle River near Laurier	3.54	0.84		0.84	
	Little Saskatchewan River near Minnedosa	1.70	1.06		1.06	
	Wilson River near Dauphin	2.51	1.33		1.10	
	Mink Creek near Ethelbert	2.93	0.86		0.86	
	North Pine River near Pine River	3.51	0.80		0.80	

D. Mann Kendall Tests

Mann Kendall tests are performed on three sets of data for each site. Once the final extraction of the Peaks Over Threshold data is complete, test files for each site are created. The data sets tested for trends are the standard deviation of the daily flow for each year, the peak flow values and the number of peak flow occurrences in each year of the record. The standard deviation is defined as the standard deviation of the daily flow values over a period of a year. A value is calculated for each year in the daily flow record. The peak flow values are extracted using the predetermined criteria of the POT Extraction technique and gathered into a data set for each site. The criteria defining the POT Extraction data cause the peak flow values to be spaced at unequal time intervals which means that the standard time series analysis where the data are assumed to be spaced an equal time period apart, cannot be used. To overcome the difficulty in analyzing these unequal intervals, the series for peak flow values are assigned a numerical counter value for each occurrence. A separate series is created to analyze the number of peak occurrences in each year of the flow record.

These series were then analyzed using the Mann Kendall test to detect trends in the record. The test indicates if there is a trend in the series. If a trend is discovered, it must be determined if it is an upwards or downwards trend (see Figure 10).

The plot in Figure 10 shows the graph for the Overflowing River peak values. A trend is quite evident in this figure, as indicated by the grey line. The purple line is a regression line calculated after the data was tested in order to clearly indicate the direction of the trend. The Mann Kendall Test checks the input data, determines if a trend exists, and indicates the direction of the trend.

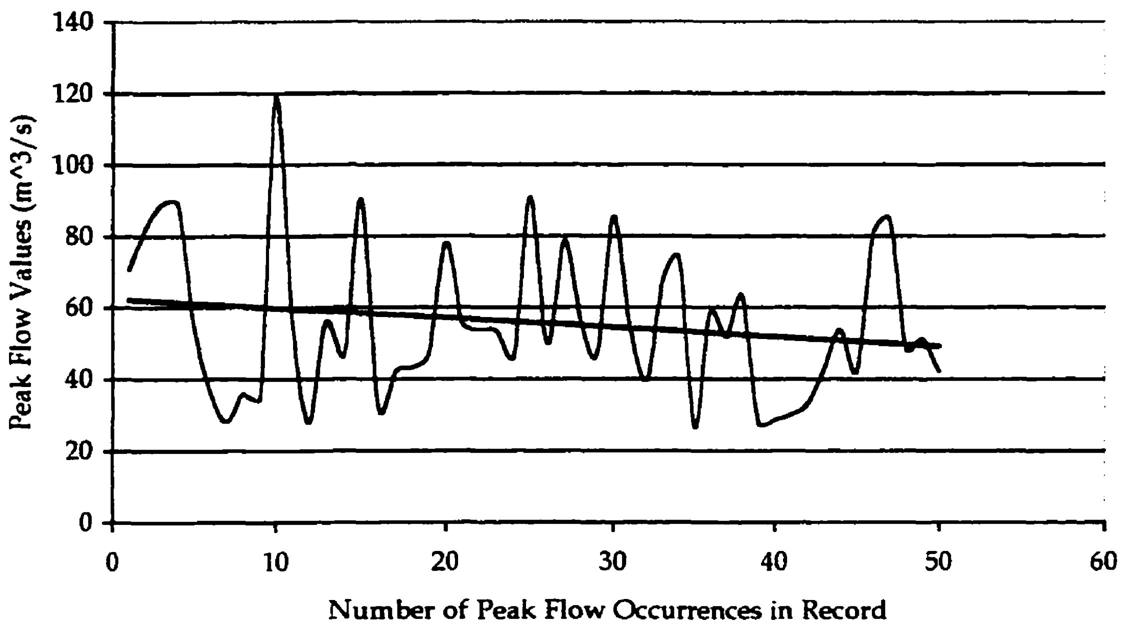

Figure 10: Overflowing River Peaks Mann Kendall Test

Once the Mann Kendall Tests were completed, the results were tabulated for further analysis. The analysis attempted to determine if specific trends occurred together, and if trends were more likely to occur in specific regions. The sites were first analyzed as a group to determine if there were any distinctive patterns in the trend analysis. The significance level of the trend is 10 percent. With a 90% confidence limit, the number of trends expected to occur by chance is 6.5 trends in 65 sites. Breaking this down by category, a possible 19.5 trends are expected to occur by chance in a possible 195 cases. The following table summarizes the results for the entire group of sites.

Table 4: Mann Kendall Test Results

Type of Trend	*Number Of Sites*	*Percentage Of Sites*
No Trend Detected in Site	37	56.9
Trend Detected in Site	28	43.1
Total	65	100
Standard Deviation Trend in all Sites	18	27.7
Peak Values Trend in all Sites	6	9.2
Peaks Per Year Trend in all Sites	16	24.6
Total Trends Detected	40 of 195	
Standard Deviation & Peak Values Trends	3	
Standard Deviation & Peaks Per Year Trends	9	
Peak Values & Peaks Per Year Trends	0	
Standard Deviation, Peak Values & Peaks Per Year	0	
Total Sites Exhibiting more than one Trend	12	

Based upon these results, it can be determined that the sites are only slightly less likely to exhibit a trend as they are to exhibit no trend in any of the data sets examined for each site. In the following table, the trends are broken down into positive and negative trends in order to determine if patterns exist. It can be seen that of the sites exhibiting trends, the majority of them exhibit positive trends. A combination trend is defined as a site that exhibits a trend in more than one category.

Table 5: Directional Analysis Results

Direction of Trend	*Number Of Sites*	*Percentage Of Sites*
Sites with no trend	37	56.9
Sites with Positive trend	15	23.1
Sites with Negative trend	13	20
Total Sites	65	100
Sites with Positive Combination trends	6	
Sites with Negative Combination trends	6	
Sites with Positive/Negative Combination trends	0	
Total Combination Trends	12	

A regional analysis is performed to determine if results that are more conclusive can be obtained. To perform the analysis, the data is first broken down into the two major regions, the Prairies and Newfoundland. The results can be found in Tables 6 and 7.

Table 6: Newfoundland Mann Kendall Test Results

Type of Trend	*Number Of Sites*	*Percentage Of Sites*
No Trend Detected in Sites	10	43.5
Trend Detected in Sites	13	56.5
Total	23	100
Standard Deviation Trend in all Sites	10	43.5
Peak Values Trend in all Sites	3	13
Peaks Per Year Trend in all Sites	7	30.4
Total Trends Detected	20 of 69	
Standard Deviation & Peak Values Trends	2	
Standard Deviation & Peaks Per Year Trends	5	
Peak Values & Peaks Per Year Trends	0	
Standard Deviation, Peak Values & Peaks Per Year	0	
Total Sites Exhibiting more than one Trend	7	

Table 7: Prairies Mann Kendall Test Results

Type of Trend	*Number Of Sites*	*Percentage Of Sites*
No Trend Detected in Sites	27	64.2
Trend Detected in Sites	15	35.7
Total	42	100
Standard Deviation Trend in all Sites	8	19
Peak Values Trend in all Sites	3	7.1
Peaks Per Year Trend in all Sites	9	21.4
Total Trends Detected	20 of 126	
Standard Deviation & Peak Values Trends	1	
Standard Deviation & Peaks Per Year Trends	4	
Peak Values & Peaks Per Year Trends	0	
Standard Deviation, Peak Values & Peaks Per Year	0	
Total Sites Exhibiting more than one Trend	5	

By examining the data for Newfoundland in Table 6, it is obvious that sites in Newfoundland are more likely to exhibit a trend in the data sets tested. This implies that the daily flow values and the associated flood events are being affected by climate variability. By breaking the Newfoundland data down into the four component regions, more conclusions can be drawn. The data for the Prairies indicate that the sites are more likely to exhibit no trend in the data sets tested, than is the case for Newfoundland.

Analyzing the component regions of each of the major regions yields interesting results. In the Avalon region of Newfoundland, the sites are equally split between exhibiting a trend and not exhibiting a trend. In the Central region, however, of six possible sites, only one (Salmon River at Long Pond) exhibits no trends (see Table 8). It should also be noted that only two sites (Bay du Nord River at Big Falls and Gander River at Big Chute) exhibit upwards or positive trends. All of the other sites exhibit downward trends. In the Northern Peninsula, of a possible eight sites, only three exhibit trends. Two of the sites exhibit trends in the Standard Deviation and Peaks per Year categories, while the third exhibits a trend in the Peak Values category. In the Southwestern region, one site exhibits no trend and the other two sites exhibit trends (see Table 8). All of the trends exhibited in this region are positive.

Table 8: Newfoundland Mann Kendall Test Z Results

Region	*Station Name*	*Standard Deviation (Z)*	*Peak Flow Values (Z)*	*Peaks Per Year (Z)*
Central	**Gander River at Big Chute**	**2.146**	**1.765**	0.247
	Middle Brook at Gambo	**-1.766**	0.252	1.164
	Terra Nova River at Eight Mile Bridge	**-3.083**	1.031	**-2.223**
	Southwest Brook at Terra Nova National Park	0.506	0.798	**-1.938**
	Salmon River at Long Pond	0.000	0.028	0.493
	Bay du Nord River at Big Falls	**3.276**	1.547	**1.742**
Southwestern	**Harry's River below Highway Bridge**	1.284	0.134	0.208
	Lewaseechjeech Brook at Little Grand Lake	**5.278**	**4.500**	1.602
	Isle aux Morts River below Highway Bridge	0.296	0.445	**2.125**

The analysis of the Prairies regions shows that the sites are less likely to exhibit trends in comparison to regions in Newfoundland. In region 1, of a possible ten sites, six exhibit no trends in the data sets. Two of the sites that exhibit trends exhibit trends in the categories of Standard Deviation and Peaks per Year. All of these trends are negative. In region 2, four sites of a possible seven exhibit no trend. Of the remaining sites, Indianhead Creek near Indian Head exhibits an upwards trend in the Peak Values category (see Table 9). All of the other trends exhibited are downwards. In region 3, two sites of a possible five exhibit trends and in region 4 none of the sites exhibit trends.

Region 5 of the Prairies is unique because all of the sites in the region exhibit a trend and all of the trends are negative (See Table 9). Region 6 of the Prairies has only a single site of three exhibiting trends. This site, however, exhibits trends in two categories, the Standard Deviation and Peak Values. Both trends are downwards. The remaining two regions, 7 and 8, exhibit the same set of trend characteristics. In both of these regions, only one site exhibits a trend. This trend occurs in the Peaks per Year category and all of the trends are downwards.

Table 9: Prairie Provinces Mann Kendall Test Z Results

Region	*Station Name*	*Standard Deviation (Z)*	*Peak Flow Values (Z)*	*Peaks Per Year (Z)*
2	Indianhead Creek near Indian Head	0.227	1.896	1.241
	Assiniboine River at Sturgis	0.994	0.195	0.769
	Stony Creek near Stenen	0.804	0.838	1.330
	Carrot River near Smoky Burn	-2.190	0.680	1.740
	Carrot River near Armley	-1.674	0.057	-2.610
	Conjuring Creek near Russell	1.020	0.639	1.092
	Arrow River near Arrow River	0.140	0.301	0.118
5	Whitemouth River near Whitemouth	-1.776	0.238	1.283
	Cooks Creek at Cooks Creek	-1.818	0.755	-2.315
	Manning Canal near Ile des Chenes	1.576	0.119	-2.396

E. Seasonality Analysis

The Mean Day of Flood for each record is also determined using the final data set. Using the day of occurrence of each peak, a set of radian coordinates is calculated. The radian coordinates for each data point are plotted as shown in Figure 11.

An average is taken of the coordinates in this set, and plotted on a graph with the radian coordinates for each peak (see Figure 11). Using the average coordinates, a Mean Day of Flood is calculated. The Mean Day of Flood is plotted for further regional analysis. This information is used to determine if patterns are present in the mean day of occurrence of the peak flows on a regional basis. Other useful information such as the standard deviation and the spread of the data can be calculated using the mean coordinates. The information obtained from a Mean Day of Flood plot allows one to determine if there is a strong seasonality component in the data, and to predict with tolerable accuracy during which time of the year flooding is most likely to occur.

In the following figure, a site from the Prairies is shown. The plot obtained from the Mean Day of Flood analysis shows a very strong seasonal tendency in the data. The majority of the peak flow occurrences are grouped tightly together with a few aberrations in the set. The Mean Peak Day for this set is 112 (April 22), which occurs in the main grouping. The standard deviation is 21 days, a value that indicates that there is not likely to be much variation in the date of the peak occurrences and the mean resultant, r^2, is 0.937, indicating that the data is very strongly seasonal.

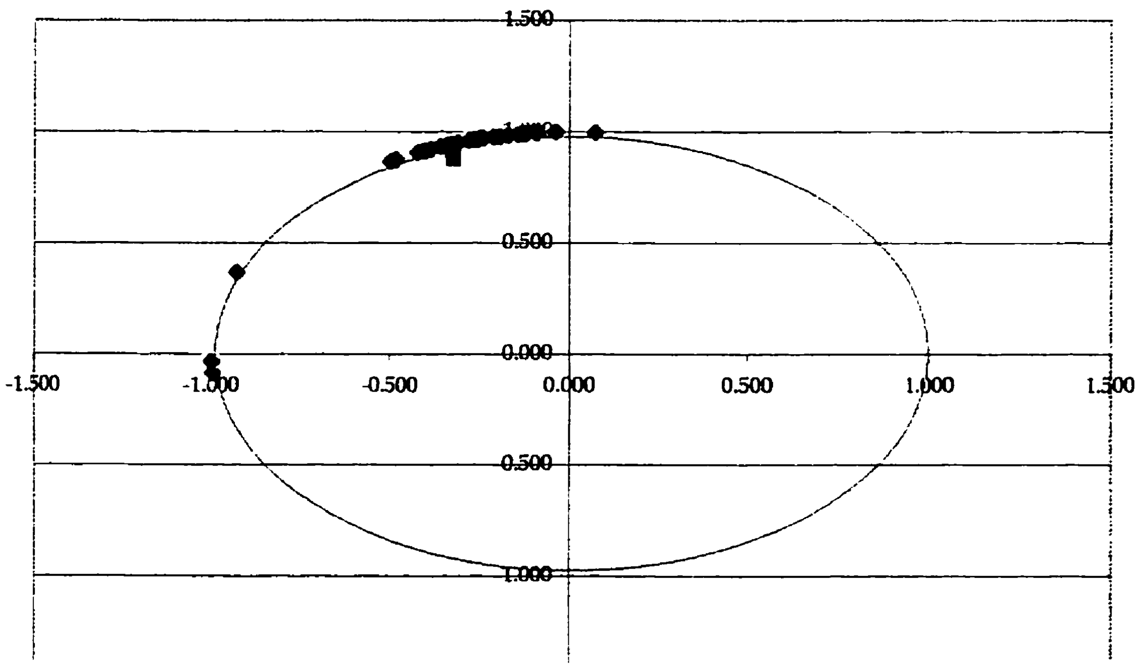

Figure 11: Carrot River At Armley Mean Day of Flood

In order to analyze the results of the Seasonality tests, the Mean Day of Flood coordinates were plotted on a single graph each for Newfoundland and the Prairies (See Figures 12 & 13). Upon close examination of the Newfoundland plot, it is observed that there does not seem to be any consistency for the occurrence of the Mean Day of Flood. This indicates that a regional analysis is required to determine if there are commonalties for the Mean Day of Flood in each region.

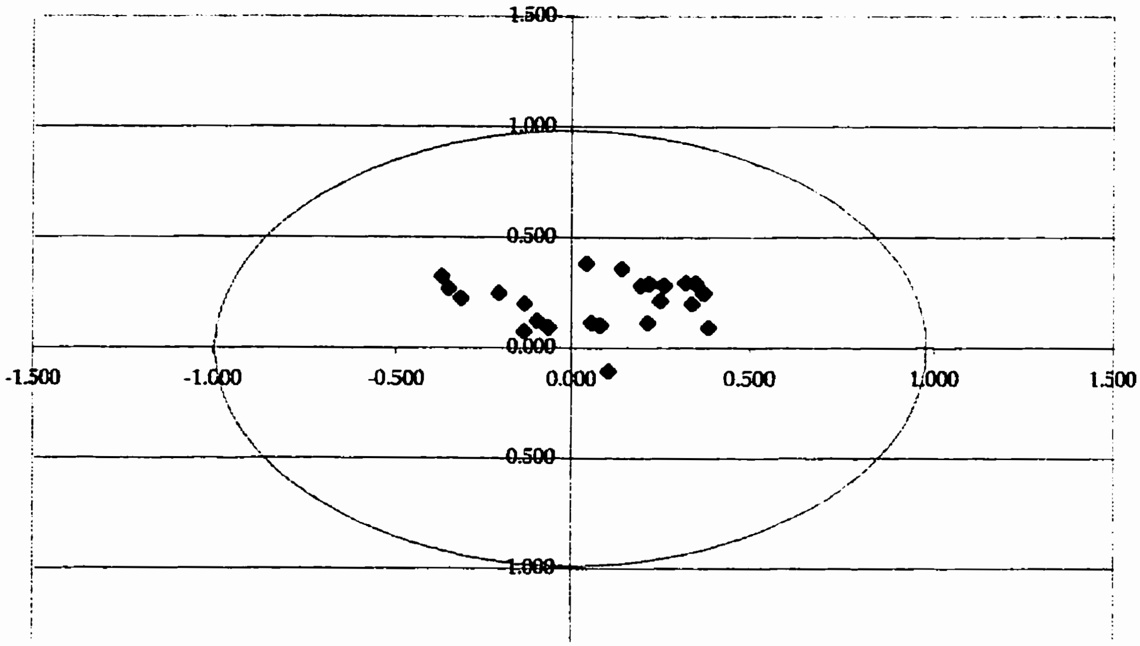

Figure 12: Newfoundland Mean Day of Flood

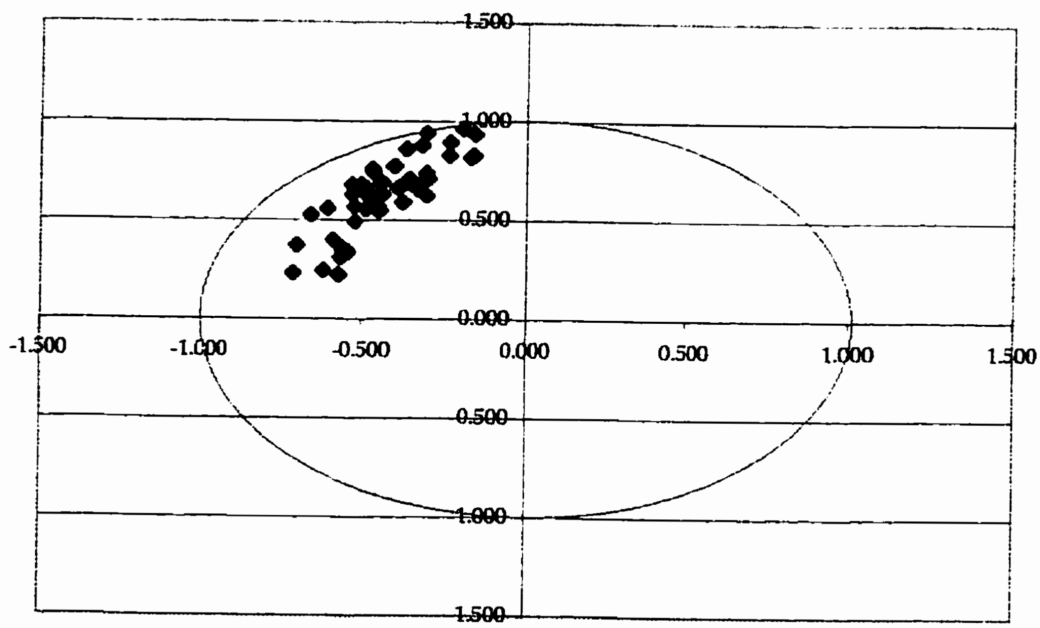

Figure 13: Prairies Mean Day of Flood

Close examination of the plot for the Prairies, however, indicates that there is a consistent seasonality pattern occurring. The Mean Day of Flood occurs within the same three month period for every site, and there is very little scatter in the data points. A regional analysis is performed to confirm that this observation is correct.

A regional analysis of the Newfoundland Mean Day of Flood data indicates that within each region, the Mean Day of Flood is more homogenous for each separate region than the initial plot indicates. The graph in Figure 14 shows regional groupings with some scatter in the data points. The Newfoundland sites generally tend to have two or three distinct times of the year when flooding occurs. This means that the seasonality component of the analysis is more difficult to apply properly. The mean resultants for the Newfoundland Mean Day of Flood vary from 0.115 to 0.495, indicating that there is not a strong seasonality component in the data. The Standard Deviation about the Mean Day of Flood also varies between 69 days and 121 days around the Mean Day of Flood. In comparison, the Prairie regions have much lower variations.

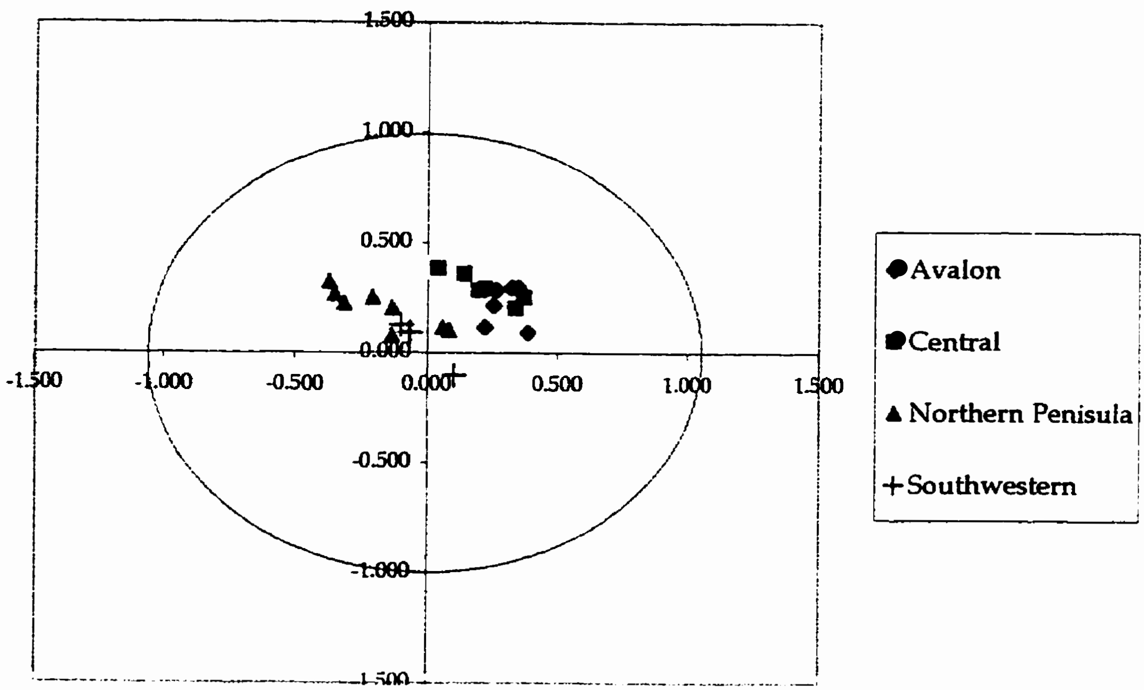

Figure 14: Regionalized Mean Day of Flood for Newfoundland

An examination of the regionalized results for the Prairies Mean Day of Flood confirms the observation that the Mean Day of Flood is homogeneous for each region. The graph in Figure 15 shows tight groupings of the Mean Day of Flood by region, with little scatter in the data. This can be attributed to the manner in which the regions for the Prairies are formed. This indicates a strong seasonality component in flood peaks on the Prairies. The mean resultants for the Prairie Mean Day of Flood vary from 0.616 to 0.992, indicating that there is a strong seasonality component in the data. The Standard Deviation about the Mean Day of Flood varies between 7 days and 57 days around the Mean Day of Flood.

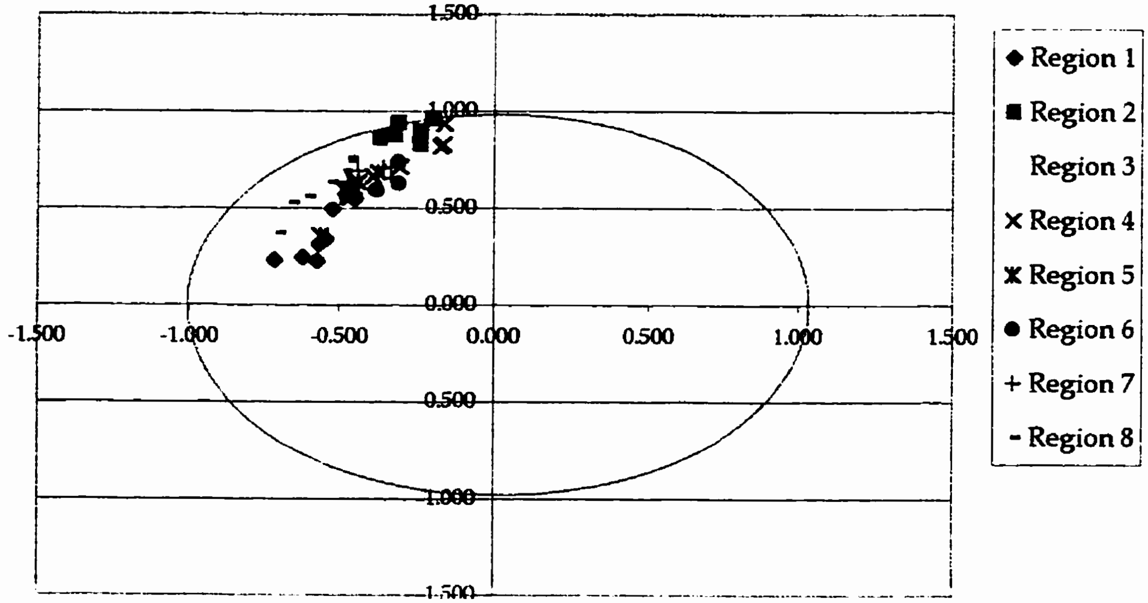

Figure 15: Regionalized Mean Day of Flood for the Prairies

F. Sensitivity of Initial Threshold

Once the final POT Extraction is complete, sites are randomly selected from the final set, in order to test the sensitivity of the results to an increase in the initial threshold level. The initial threshold is increased from 1.5 times the average flow value to 1.75 time the initial average flow value.

Upon careful observation of Table 10, it can be noted that in 2 of the 16 cases the threshold set by the sensitivity analysis was too high when compared with the final selected threshold. The sensitivity analysis indicates that the initial choice of 1.5 times the average daily flow for the record is the better choice of the two for conducting a preliminary analysis. A choice of 1.75 times the average daily flow for the record will be too high for some data sets. The sites selected for the sensitivity analysis contain two sites out of a possible four in the entire data set that have the same initial and final threshold values. For these two sites, the new threshold is too high. In the event that the final chosen threshold is smaller than the initial threshold, there is a great deal of work to be done to extract the peaks, as the program must be run from the beginning. If the final chosen threshold is larger than the initial threshold, much less work is required to extract the peaks. This indicates that a low initial threshold value should be chosen, hence the selection of 1.5 times the average daily flow value.

Table 10: Sensitivity Results

Name	*Initial Threshold Value*	*Threshold Value (Sensitivity Test)*	*Final Threshold Value*
Lewaseechjeech Brook at Little Grand Lake	25.924	30.245	44.071
Northeast Pond River at Northeast Pond	0.204	0.238	1.124
Northwest Brook at Northwest Pond	4.719	5.505	18.403
Salmon River at Long Pond	128.267	149.645	128.267
Sheffield Brook near Trans Canada Highway	15.779	18.409	19.724
Terra Nova River at Eight Mile Bridge	56.819	66.289	79.547
Upper Humber River near Reidville	129.956	161.615	311.894
Carrot River near Armley	8.032	9.371	17.670
Conjuring Creek near Russell	0.178	0.207	1.173
Cooks Creek at Cooks Creek	1.962	2.289	6.474
English River near Sioux Lookout	180.804	210.938	180.804
Little Saskatchewan River near Minnedosa	6.266	7.310	11.905
Main Drain near Dominion City	0.786	0.917	3.694
North Pine River near Pine River	2.002	2.336	9.811
Overflowing River at Overflowing River	19.129	22.317	26.781
Willow Creek near Gimli	1.134	1.323	2.268

Chapter 6: Discussion

In order to determine the effectiveness of the previous analysis, some consideration should be given to the regions analyzed. The regions in question are located in two different parts of Canada, and experience very different climate conditions. The sites located in the Prairies are land locked in the midst of Central Canada, although a few sites are located close to Lake Winnipeg and the Lake of the Woods. The climate in the Prairies is quite warm and dry in the summer with an annual precipitation ranging from 350 mm in Saskatchewan to 625 in Northwestern Ontario, with an annual snowfall ranging from 120 cm in Saskatchewan to 200 cm in Northwestern Ontario. Flooding is most likely to occur in the spring during the snowmelt runoff period, especially if heavy rains occur during the melt period. The sites located in Newfoundland are surrounded by the ocean and rainfall is quite heavy, especially during hurricane season in the winter months. The annual precipitation ranges from 860 mm to 1500 mm in Newfoundland, with snowfall ranging from 200 cm to 450 cm.

There are other differences between the two regions. The topography of Newfoundland is quite hilly and rocky with numerous fjords and inlets on the island. The land is used mainly for mining and the main industry is fishing. The

Prairies are broken down into three main climatic zones: the Prairies, the Northeastern Forest and the Northwestern Forest. These zones can be broken down into the plains and the Canadian Shield. The plains are flat, with a few rolling hills and small stands of trees. The main land uses are farming and urban areas. The Canadian Shield is rocky and heavily forested. The main land use is forestry.

The Newfoundland sites have slightly shorter daily flow records available, ranging from 12 to 68 years in length, with most sites having 25 to 40 years of record available. They also have smaller catchment areas, ranging in size from 4 km^2 to 4400 km^2, with most sites having catchment areas of 300 to 800 km^2. The sites in the Prairies have longer daily flow records available, ranging from 35 to 82 years in length, with most sites having between 35 and 50 years of record available. They also have larger catchment areas, ranging from 88 km^2 to 13 900 km^2.

The use of the POT Extraction technique to gather flood data is a good choice. By allowing more than one peak flow value per year, a more comprehensive picture of flood occurrences than observed by the Annual Duration Series can be collected. Some consideration should be given to the choice of the initial threshold value however. The use of the predetermined 1.5 times the average daily flow value for the entire record for the initial threshold is a good starting estimate for sites in the Prairies where generally only one flood event occurs in a year. This value is not as good an estimate for sites in

Newfoundland, where there are generally two to three flood events in a year. The flow values are more varied, swinging from extremely low values to extremely high values. This could be attributed to the severe storms experienced by the region, where the rainfall is either quite heavy or is almost non-existent. Because of this phenomenon, the average daily flow tends to be somewhat lower than the majority of the high flow values and initial estimates of between five and sixteen flood events a year occur. This makes it more difficult to determine an appropriate number of flood peak occurrences per year.

The use of Mean Excess Plots to choose a new threshold is difficult. Much interpretation is required to apply this theory, and it is difficult to determine what constitutes a straight descending line. The sites in the Prairies presented plots that were well behaved, making the choice of a new threshold simple. The Newfoundland sites were more difficult to analyze, presenting plots that shifted up and down without any discernable pattern. This is caused by the large variations in the size of the peak occurrence. With a larger number of very high flood peak values, the removal of smaller flood peaks, the mean residual value fluctuates very rapidly with the increase in the threshold value. The use of regionalization made the choice of the new threshold simpler. By determining a regional average, the choice of a new threshold was dependent on the number of peak flow values per year produced by the MEP analysis, and not exclusively on the predetermined criteria of a point at the beginning of a straight descending line.

The use of regionalization to determine the new threshold worked very well for both sites in Newfoundland and the Prairies. By regionalizing the sites in Newfoundland, it became simple to determine an average number of peak flow occurrences per year, and use this average value to determine what value could be used for the sites that appeared to have no discernable pattern. The regionalization of the sites in the Prairies confirmed the results determined by the use of MEP in most cases and allowed for a better choice to be made in a few difficult cases where the MEP results indicated less than 0.5 flood peaks per year in areas in Newfoundland experiencing 2 or 3 flood peaks per year.

The trend analysis was easily applied to both regions although the results were quite different. When a trend was detected in the sites in Newfoundland, over forty percent of the sites analyzed showed an increase in flood peak data in at least one of the categories tested. Less than fifteen percent of the sites in Newfoundland showed a decreasing trend. This indicates that flood peak occurrences are increasing and that the flood peak values are becoming larger, and more severe. In contrast, only twelve percent of the sites analyzed in the Prairies showed an increase in flood peak data, and twenty four percent of the sites showed a negative trend. This indicates that the climate is more stable in the Prairies, and there are less likely to be negative impacts on the region. Overall, the Newfoundland sites were more likely to exhibit a trend than the Prairies sites were.

Each sites was analyzed for trends in three variables. In Table 11, the variables are analyzed for directional results and the number of trends occurring for each variable. It is obvious that the sites in Newfoundland are being affected by something. As the sites were chosen in rural areas where there is no evidence of changes in the land use patterns, these trends imply that climate variability is impacting these sites. The number of trends expected to occur by chance at the 90% significance level for the sites in Newfoundland is two trends in any category or a total of six trends. Obviously, with a total of 20 trends exhibited of a possible 69 or 29%, there is evidence that climate variability is affecting the Newfoundland sites. The Prairies sites exhibit less trends for the same categories. These sites were also chosen to be stationary and there is no evidence of changes in the land use patterns. The number of trends expected to occur by chance for the Prairies at the 90% significance level is four trends per category or a total of twelve trends. With the Prairie sites exhibiting 20 trends of a possible 126, there is evidence that climate variability is not impacting the Prairie sites as strongly as the Newfoundland sites.

Another consideration is the direction of the trend in both regions. Although the Newfoundland sites and the Prairie sites both exhibit trends in the upwards and downwards directions, there is a tendency to one direction over the other. It was previously noted that with a decrease in precipitation, the daily streamflow values would decrease, therefore implying that the flood peak values would also decrease. There is ample evidence to suggest that this theory holds true for the sites in the Prairies where both the peaks per year and the daily flow values (standard deviation) are decreasing. Conversely, the number of peaks per year, the peak flow values and the standard deviation are generally increasing in Newfoundland. This implies that there is an increase in precipitation in the area.

Table 11: Mann Kendall Trend Test Results

	Standard Deviation	*% Of Sites*	*Peak Flow Values*	*% Of Sites*	*Peaks Per Year*	*% Of Sites*
Newfoundland Sites						
Sites with no Trend	15	65.2	18	78.3	16	69.6
Sites with Positive Trends	8	34.8	3	13.0	5	21.7
Sites with Negative Trends	0	0.0	2	8.7	2	8.7
Total	23	100.0	23	100.0	23	100.0
Prairies Sites						
Sites with no Trend	34	81.0	39	92.9	33	78.6
Sites with Positive Trends	1	2.4	2	4.8	2	4.8
Sites with Negative Trends	7	16.7	1	2.4	7	16.7
Total	42	100.0	42	100.0	42	100.0

The implications of these increasing and decreasing trends are serious. Traditional flood frequency analysis will become less reliable as an analysis tool if these trends continue. With an increase or decrease in both the number of flood events in a year and the size of those flood events, the ADS will be less accurate. Other analysis techniques may need to be developed in order to deal with these trends. Water resources may become less available, and new management techniques will need to be developed in order to cope with the decreasing water supplies. With an increase in water supply new management techniques must also be developed to determine how to use the excess water effectively.

The seasonality analysis was a useful diagnostic tool. Seasonality effects were easily discernable in sites in the Prairies, but were more difficult to discern in the sites in Newfoundland. Due to the year round flooding ability in Newfoundland, strong seasonal effects are difficult to observe. By breaking the sites down into regions, however, seasonality effects were more noticeable. Each of the four sub-regions in Newfoundland exhibited seasonality effects that were not evident in the original plot of all of the sites. The MPD for each of the sites in the sub-regions tend to be located in the same quadrant of the plot, with some scatter in the data. This is indicative of seasonality effects experienced by the sites in the sub-region. The MPD for the sites in the Prairies are all gathered in one quadrant. By breaking the sites into sub-regions, it can be observed that there are very strong seasonality effects experienced by the sites in the Prairies. The MPD for the sites in each of the sub-regions are tightly grouped with little scatter, indicating very strong seasonal effects. This indicates that flooding is likely to occur during only a single season, and that the sites in each region are very strongly tied together.

Using the analysis in Chapter 5 as a basis, it is obvious that there are considerably more trends demonstrated in the two geographical areas than could be expected to occur by chance. This indicates that the areas are being affected by some external force. Given that the sites were chosen to be unaffected by urbanization and that they were all rural sites on uncontrolled rivers, it can be concluded that the changes were not directly associated with humans. The patterns of the trends tend to be consistent within the geographical areas. Within each region, similar trends are demonstrated at the different sites. The sites in Newfoundland tend to demonstrate increasing trends, whilst the sites in the Prairies tend to demonstrate decreasing trends. This can be linked with the changes in temperature noticed over the last several decades. The Atlantic area has experienced slightly decreasing temperatures and the Prairies have tended to experience increasing temperatures. A decrease in temperature does not necessarily mean an increase in flooding but the correlation may be correct. It has been observed that an increase in temperature is correlated with a decrease in hydrologic variables (Burn, 1994b). This indicates that climate change as opposed to climate variability is a plausible explanation for the trends identified.

CHAPTER 7: CONCLUSIONS

The major contribution of this thesis is to determine if climate variability effects are causing discernible changes to flood frequency patterns and the seasonality of flooding occurrences. In order to perform this analysis, the Peaks Over Threshold extraction technique was used to obtain flood peak data for sites in Newfoundland and the Prairies. This technique differs from the Annual Duration Series analysis often used, requiring adaptations to be made to existing trend tests. A special seasonality test is also used to analyze the resulting data.

With the completion of the tests on the flood peak data, the results are analyzed for trends in the flood data. Trends are quite evident in forty five percent of the sites analyzed. Once divided into regions, fifty seven percent of the sites in Newfoundland exhibit trends in at least one category. Of these trends, an increasing trend is present in 43.5% of the sites examined. In contrast, a mere 11.9% of sites examined in the Prairies demonstrate a positive trend, and sixty four percent demonstrate no trend at all. It can be concluded that there are changes occurring in the flooding patterns in Newfoundland. The changes to flooding patterns in Newfoundland indicate that flooding occurrences are becoming more frequent and the flood peak values are becoming larger. This will cause more damage to property and could cause economic problems. The increase in flooding occurrences and the flood peak values implies that flood frequency analysis will become more imprecise if the more conventional means of analysis such as the Annual Duration series continue to be used. The Prairies have fewer sites displaying trends but 23.8% of the sites examined demonstrate negative trends, indicating a decrease in flooding occurrences and in flood peak values. This indicates that the Prairies are able to absorb the water input with little difficulty.

The seasonality analysis indicates strongly seasonal tendencies for flooding in the Prairies, and very weak seasonal tendencies for flooding in Newfoundland. When analyzed as one complete region, Newfoundland flooding patterns display a large amount of scatter in the data with some small groupings. The Prairies flooding patterns are tightly grouped with some scatter. A sub-regional analysis of the two distinct geographic areas indicates that the sub-regions of each area have discernible flooding patterns. The small groupings observed in the Newfoundland data consist of sites from one of the sub-regions, indicating patterns in the flood peaks experienced by sites in that region. The sub-regions of the Prairies also indicate that sites in the sub-regions tend to experience flood peaks during the same time of the year.

It can be concluded that the POT Extraction is a useful analysis technique for examining flood peak data for trends and patterns. The inclusion of smaller flood peaks and the possibility of more than one flood peak per year allows a better analysis for examining climate variability effects.

Comparisons of the two regions, the Prairies and Newfoundland, indicate that there are few similarities between them. The differences however are great. With the different climate conditions experienced by each region, it is not unexpected that such differences exist in their flooding patterns. The analysis of these two regions does indicate that each region is undergoing an increase in severe weather conditions. Newfoundland is experiencing more severe storms and flooding conditions, while the Prairies are experiencing more typical weather conditions.

REFERENCES

Bayliss, Adrian C. and Richard C. Jones, March 1993. Peaks Over Threshold Flood database: Summary Statistics and Seasonality. Crowmarsh Gifford, Wallingford: Institute of Hydrology Report No. 121.

Birikundavyi, S. and J. Rouselle, 1997a. A Technique for Selecting the Threshold in Applications of the Poisson-Exponential Flood Model. Canadian Journal of Civil Engineering: Vol. 24, p. 12 – 19.

Birikundavyi, S. and J. Rouselle, 1997b. Use of Partial Duration Series for Single-Station and Regional Analysis of Floods. Journal of Hydrologic Engineering, April 1997, p. 69 –75.

Burn, D. H., 1994a. Hydrologic Effects of Climatic Change in West-Central Canada. Journal of Hydrology, 160 (1994), p. 53 – 70.

Burn, D. H., 1994b. The Impact of Climatic Change on Hydrologic Variables. Proceedings of Canadian Water Resources Association: 47th Annual Conference, p. 485 – 498.

Burn, D. H., 1997. The Impact of Climate Change on Flood Magnitudes. Proceedings of the 1997 CSCE Annual Conference: 13th Canadian Hydrotechnical Conference, Vol. 3, p. 179 – 188.

Burn, D., Z. Zrinji and M. Kowalchuk, 1997. Regionalization of Catchments for Regional Flood Frequency Analysis. Journal of Hydrologic Engineering, April 1997, p. 76 -82.

Changnon, S. A. and K. E. Kunkel, 1995. Climate Related Fluctuations in Midwestern Floods during 1921 - 1985. Journal of Water Resources Planning and Management, Vol. 121, No. 4, Jul/Aug.

Churchill, D. M., R. W. Galloway and G. Singh, 1978. Closed Lakes and the Paleoclimatic Record, in *Climatic Change Variability*, Ed. by A. B. Pittock, L. A. Frakes, D. Jenssen, J. A. Peterson and J. W. Zillman, Cambridge University Press, Cambridge, United Kingdom.

Cunnane, C., 1978. A Note on the Poisson Assumption in Partial Duration Series Models. Water Resources Research: Vol. 15, No. 2, p. 489 - 494.

Davison, A. C., 1984. Modelling Excesses over High Thresholds, with an application. In Statistical Extremes and Applications, ed. J. Tiago de Oliveira, p. 461 – 482. D. Reidel Publishing Company.

Hirsch, R. M., J. R. Slack, and R. A. Smith, 1982. Techniques of Trend Analysis for Monthly Water Quality Data. Water Resources Research, 18(1), p. 107 – 121.

Hosking, J. R. M. and J. R. Wallis, 1993. Some Statistics Useful in Regional Frequency Analysis. Water Resources Research, 29(2), p. 271 – 281.

Kendall, M. G., 1975. *Rank Correlation Measures*. Charles Griffin, London, 202 pp.

Lawford, R. G., 1994. The Potential Role for Hydrological and Hydrometerological Parameters in a Climatic Change Detection Network. Canadian Water Resources. Vol. 19, No. 3.

Madsen, Henrik, Charles P. Pearson and Dan Rosbjerg, 1997. Comparison of Annual Maximum Series and Partial Duration Series Methods for Extreme Modeling Events, Part 2. Water Resources Research: Vol. 33, No. 4, p. 759 – 769.

Mann, H. B., 1945. Non-parametric Tests Against Trend. Econometrica, 13: 245 – 259.

Mannion, A. M. and S. R. Bowlby, 1992. *Environmental Issues in the 1990s*. John Wiley & Sons, London, 349 pp.

Naden, P. S. 1992. Analysis and use of Peaks Over Threshold data in Flood Estimation. In BHR Group Conference on Floods and Flood Management, ed. A. J. Saul, p. 131 – 143. Kluwer Dordrecht.

Nkemdirim, L. C. and H. Purves, 1994. Estimating the Potential Impact of Climatic Change on Streamflow in the Oldman River Basin, Alberta: An Analogue Approach. Canadian Water Resources Journal, Vol. 19, No. 2.

Pilon, P. J., T. Winkler, K. D. Harvey and D. R. Kimmet, 1991. Hydrometric data in Support of Climate Change Studies in Canada. Presented at NATO Advanced Research Workshop on Opportunities for Hydrologic Data in Support of Climate Change Studies, Lahnstein Germany, 26 – 20 August, 1991.

Sceviour, Edward, 1993. Flood Frequency Analysis for the Island of Newfoundland and the Test for Homogeneity.

Steel Structures Education Foundation. NBCC 1990 Structural Design Education Aid. National Research Council of Canada. 220 pp.

Stone, J. M. R. 1992. The Canadian Response to the Climate Change Issue. Canadian Water Resources Journal. Vol. 17, No. 1.

Tasker, G. D., 1993. Effects of Climate Change on Drought Risks in the Apalacicola, Chattahoochee and Flint River Basin. Proceedings for the Symposium in Engineering Hydrology. U. S. Geological Survey, ASCE, p. 49 - 54.

Westmacott, J. R., 1996. The Effects of Climate Change on Hydrologic Variables. M.Sc. Thesis, University of Manitoba, Winnipeg, Manitoba, Canada.

IMAGE EVALUATION TEST TARGET (QA-3)

150mm

6"

APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989